CIPPM Centre for Intellectual Property Policy & Management

Jean Monnet Centre of Excellence 2018-2021 for European Intellectual Property and Information Rights

CIPPM / Jean Monnet Woking Papers

No. 02-2020

Virtues and Perils of Algorithmic Enforcement and Content Regulation in the EU – A Toolkit for a Balanced Algorithmic Copyright Enforcement

Maria Lillà Montagnani

April 2020

Bournemouth University

Co-funded by the Erasmus+ Programme of the European Union

The Centre for Intellectual Property Policy and Management of Bournemouth University is a Jean Monnet Centre of Excellence for European Intellectual Property and Information Rights (2018-2021), co-funded by the Erasmus+ Programme of the European Union.

This Working Paper Series is peer reviewed by an Editorial Board led by prof. Ruth Towse and prof. Roger Brownsword.

VIRTUES AND PERILS OF ALGORITHMIC ENFORCEMENT AND CONTENT REGULATION IN THE EU – A TOOLKIT FOR A BALANCED ALGORITHMIC COPYRIGHT ENFORCEMENT[*]

Maria Lillà Montagnani [**]

Within the recent European policies and actions on illegal content, a trend towards algorithmic enforcement of content regulation has emerged. Regardless of the nature of the content, hard and soft law provisions more or less explicitly require online platforms to resort to technological systems to comply with the law. The use of technology to enforce the law is certainly not new, especially in the realm of copyright law. Copyright law is indeed the perfect example of how the adoption of technology by online intermediaries has over time altered the contours of the law itself. The last step in this process is the employment of algorithmic systems to filter content uploaded by third parties and the use of autonomous decision-making to select the content that can appear online. This controversial legislative move towards algorithmically enforcing legal rules and content regulation raises concerns not only as to its consistency with the current legal framework but also–and more worrisomely–as to its impact on individual rights and societal development in general. For this reason, this paper proposes a regulatory toolkit for a more balanced algorithmic copyright enforcement that could, hopefully, also provide insights for a better algorithmic society overall

[*] Pre-print. To be published in the *Journal of Law, Technology & the Internet*, Vol. 11, No. 1, 2019-2020.

[**] Associate Professor of Commercial Law, Bocconi University of Milan, and Visiting Research Fellow of CIPPM / Jean Monnet Centre of Excellence for European Intellectual Property & Information Rights, Bournemouth University.

CONTENTS

I. INTRODUCTION

The European Digital Single Market Strategy ('DSM Strategy')[2] introduced a strong trend towards algorithmic enforcement of the rules aiming at preventing illegal content online.[3] This is part of a broader movement towards algorithmic content regulation.[4] The trend surfaces from the analysis of the various hard and soft law provisions touching upon online intermediaries' liability that have been adopted by the European institutions to fight the upload and spread online of illegal content.[5] Although the safe harbor regime for online intermediaries set by Directive 2000/31/EC ('e-Commerce Directive') remains untouched, the several instruments adopted point toward a system in which intermediaries are required to implement technological measures not only to take down, but also to prevent the (re)appearance of allegedly illegal content online.[6] This determines a shift from a regime in which the law is enforced after a violation of law has taken place (ex post) to a system where technology ensures that violations do not even occur in the first place (ex ante). In this way, the technologies implemented to comply with the law get ahead of the threshold of protection and in doing this they change the law itself.

The case of copyright law is probably the most explicit example of algorithmic enforcement and, more specifically, of the shift from an ex post to an ex ante system of technological enforcement of the law. Forms of algorithmic enforcement of copyright law were present well before the adoption of the controversial directive on copyright in the DSM ('DSM Directive').[7] This instrument represents only the last step of a process targeting piracy online that started several years ago. The resulting shortcomings of this piece of legislation are widely known and discussed by many scholars who highlight how a lack of

---

[2] *See generally Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on A Digital Single Market Strategy for Europe*, COM (2015), 192 final (June 5, 2015).

[3] *See* Thomas Riis & Sebastian Felix Schemer, *Leaving the European Safe Harbor, Sailing Towards Algorithmic Content Regulation*, 1 J. INTERNET L. 1 (2019).

[4] *Id*.

[5] In this work, the terms 'intermediaries', 'online intermediaries' and 'online platforms' are used interchangeably. While the focus is on illegal content hosted by online platforms, the umbrella term 'intermediaries' will still be used. *See* Teresa Rodriguez de las Heras Ballell, *The Legal Autonomy of Electronic Platforms: A Prior Study to Assess the Need of a Law of Platforms in the EU*, 3 IT. L. J. 149, 153–154, 156–160 (2017), for a discussion on the difference between intermediaries and online platforms.

[6] *See* Directive 2000/31, of the European Council of June 8 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market, 2000 O.J. (L 178) 6, 12.

[7] *See* Directive 2019/790, of the European Council of April 17 2019 on Copyright and Related Rights in the Digital Single Market. 2019 O.J. (L 130) 92.

accountability challenges many of the legal principles such as freedom of expression, due process and the right to self-determination.

In this article, I start from the proposition that the algorithmic society is here to stay to formulate a proposal for a balanced algorithmic copyright enforcement. Given the technological developments we are witnessing in terms of autonomous systems and self-learning algorithms, it is likely that technology will continue to provide an increasingly sophisticated compliance tool. If this is the case, we need to be well-equipped to govern the developments to come.

The European Union has already offered a variety of hard and soft law provisions that can add more balance to algorithmic enforcement. However, the many principles, recommendations, suggestions and tools are not coherently organized—rather, they confirm the piecemeal approach that European institutions often adopt.

This article contributes to the existing debate on algorithmic copyright enforcement by gathering the acquis communautaire, organising it in the most consistent manner possible, and trying to fill the emerging gaps. As there is no need to reinvent the wheel, this article considers the rules already available, adapts them to the current scenario, and adds the missing links. As a result, a regulatory toolkit is proposed with the objective of introducing more balance to the algorithmic copyright enforcement. Ultimately, the goal of this toolkit is to provide useful insights for a better algorithmic society.

In the following pages, Section I illustrates the development of European policies targeting illegal content online and depicts how this interacts with the current liability regime for online platforms. Section II turns to the hard and soft law provisions that emerged from these EU policies, which drive towards the adoption of technology as a main tool of compliance. Section III focuses on copyright law as the recent debate surrounding Article 17 of the DSM Directive provides the clearest example of the shift from ex ante to ex post algorithmic enforcement. Section IV summarizes the concerns of algorithmic copyright enforcement—more generally, technologies of compliance—and illustrates the solutions so far envisaged. Section V introduces the main features of the regulatory toolkit for a balanced algorithmic copyright enforcement. Since algorithms are here to stay, there is the pertinent need to make them comply with the legal framework. Finally, in Section VI I conclude by stressing the need to continue the conversation on the unintended consequences of algorithmic regulation and enforcement.

II. TACKLING ILLEGAL CONTENT ONLINE IN THE DSM STRATEGY

Tackling illegal content is one of the priorities of the DSM Strategy as European institutions intend to ensure what is illegal offline is also illegal online. This represents one of the many regulatory challenges in the process of

creating the Digital Single Market.[8] Naturally, an analysis of the documents articulating the DSM Strategy reveals that tackling illegal content requires the assessment of online intermediaries' role and activity.

The DSM Strategy places special emphasis on three key pillars: (*i*) access for consumers and businesses to online goods and services across Europe; (*ii*) creating the right conditions for digital networks and services to flourish; and (*iii*) maximizing the growth potential of the digital economy.[9] The first two pillars touch upon the issue of illegal content and more precisely, the intermediaries' liability for illegal content. In this respect, the DSM Strategy clearly states that the "rules on the activities of intermediaries in relation to copyright-protected content" are to be clarified.[10] Similarly, in the pursuit of the second pillar's goal to optimize digital networks and services, the Commission stresses that online platforms' market power can potentially impact other participants in the marketplace.[11] While the level playing field conditions form one concern, the need to guarantee that minors are protected from harmful content and that internet users are protected from hate speech and misleading content form equally relevant considerations.[12]

This section illustrates the European institutions' policy on tackling illegal content and its impact on platforms' liability. The main argument centers on the proposition that these two fit poorly within the current conditional liability regime as set out by the e-Commerce Directive.

A. An Enhanced Liability Regime for Online Platforms

The proposition that tacking illegal content requires a more active role of online intermediaries emerges very clearly in the EU Commission's 2017 Communication on tackling illegal content online.[13] Here, the Commission expressly calls for an enhanced "liability regime" for intermediaries due to the strategic role that they perform in mediating content access to internet users.[14] Moreover, whatever content platforms would tackle, from incitement to terrorism, hate speech, child sexual abuse material to infringements of IPRs, the Commission underlines that online platforms should "adopt effective proactive

---

[8] *See supra* note 1, at 3.
[9] *Id.* at 3-4.
[10] *Id*.
[11] *Id.* at 9.
[12] *Id.* at 11.
[13] *See Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Tackling Illegal Content Online Towards an Enhanced Responsibility of Online Platforms*, at 2, COM (2017) 555 final (Sept. 28, 2017).
[14] *Id.* at 5.

measures to detect and remove illegal online content."[15] While encouraging the use and development of automatic technologies to prevent the re-appearance of illegal content online,[16] the Communication also recalls the need to comply with the respective fundamental rights and the necessity to implement the safeguards to limit the risk of removing legal online content.[17]

As a follow-up to the 2017 Communication, in March 2018 the Commission issued a recommendation, which, this time, encompasses more concrete measures to effectively tackle illegal content online.[18] Even though in this document, the Commission has elaborated in some more detail the notice and action procedure and at the same time called for the adoption of proactive measures for all types of illegal content, its emphasis on the need to implement such a procedure in a diligent and proportionate manner remains rather hazy.[19] In particular, the Commission once more focuses on the necessity to adopt automatic filtering systems and encourages online platforms to invest in automatic detection technologies.[20]

The 2017 Communication and the 2018 Recommendation are only the last initiatives of a series of soft law instruments issued by the European institutions. In 2015 the Commission started the discourse on illegal content with a public consultation on the role of online platforms.[21] Among several other topics, it also covered how best to tackle illegal content on the Internet and the need to reform online intermediaries' liability regime.[22] The responses to the consultation did not however reflect one uniform view, due to the vast differences between the type of respondents and their respective interests. For example, on the question of fitness of the liability regime under the e-Commerce Directive—most individual users, content uploaders, and intermediaries considered it fit-for-purpose—while right holders, their associations, and notice-providers identified gaps and were unsatisfied with its effectiveness.[23]

Despite the highly inconclusive results of the public consultation, in one 2016 Communication, the Commission set out its official position on the issue

---

[15] *Id.* at 10.

[16] *Id*. at 18.

[17] *Id*. at 14, 16.

[18] *See Commission Recommendation on Measures to Effectively Tackle Illegal Content Online*, at 9, COM (2018) 1177 final (Mar. 1, 2018).

[19] *Id* at 5.

[20] *Id.* at 6.

[21] *See generally Synopsis Report on the Public Consultation on the Regulatory Environment for Platforms, Online Intermediaries, Data and Cloud Computing and the Collaborative Economy* (Jan. 26, 2016), http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=15877.

[22] *Id.* at 15-21.

[23] *Id.* at 15.

of tackling illegal content and online platforms.[24] Specifically, the Commission proposed a problem-driven approach to content regulation, whereby "any future regulatory measures proposed at EU level only address clearly identified problems relating to a specific type or activity of online platforms in line with better regulation principles".[25] Having mentioned that online platforms "come in various shapes and sizes and continue to evolve at a pace not seen in any other sector of the economy,"[26] the Commission stresses their rising importance as well as the need to have them operate in a balanced regulatory framework—a framework in which, for their role in providing access to information and content, they bear more responsibility.[27]

The same principles mentioned above are further confirmed in the following mid-term review assessment of the progress towards the implementation of the DSM, where the issue of illegal content online is tackled from a more practical point of view—namely in relation to the mechanisms and technical solutions necessary for its removal, which in turn must be effective and, at the same time, fully respectful of fundamental rights.[28]

B. A Conditional Liability Regime for Online Intermediaries

The proposition that online intermediaries should take more responsibility for the content on their platforms—to the extent of adopting ad hoc technologies to enforce the law—is not fully in line with the conditional liability regime introduced by the e-Commerce Directive. This stems mainly from two aspects of the Directive. Firstly, Article 14 introduces a horizontal safe harbor exemption for "information society service providers" performing mere conduit, caching and hosting. In particular, intermediaries are exempted from liability for the illegal content uploaded by third parties as long as intermediaries are in no way involved with the information transmitted. Alternatively, in the case of hosting services, if they do not have knowledge or awareness of any illegal activities and, if such knowledge is acquired, they should act promptly to remove the illegal content. Secondly, Article 15

---

[24] *See Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Online Platforms and the Digital Single Market: Opportunities and Challenges for Europe*, at 9, COM (2016) 288 final (May 25, 2016).

[25] *Id.* at 5.

[26] *Id*. at 2.

[27] *Id.* at 4.

[28] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the Mid-Term Review on the Implementation of the Digital Single Market Strategy*, at 9 (May 10, 2017), https://eur-lex.europa.eu/resource.html?uri=cellar:a4215207-362b-11e7-a08e-01aa75ed71a1.0001.02/DOC_1&format=PDF.

introduces a prohibition for Member States to impose on online intermediaries monitoring obligations of a general nature.

The element of knowledge or awareness is the key requirement in this conditional liability regime.[29] If intermediaries acquire knowledge or awareness of illegal content on their platforms and fail to react promptly, they no longer fall within the safe harbor provided by the exemption. However, not only does the e-Commerce Directive leave vague the question of when knowledge and awareness are actually acquired, but it also does not specify the process of removing the illegal content in question. Unlike the US system, in which the legislature has regulated in detail the procedure to be followed by online intermediaries,[30] even though limiting itself to the violation of copyright under section 512 of the DMCA,[31] the EU rules leave the process specifics to the discretion of market operators—or rather to the Member States.[32] In the absence of a precise provision at European level on how intermediaries should react and remove infringing content, the US model of notice and take down (N&TD) has become the standard practice for the majority of online intermediaries. In any event, in both jurisdictions, under the safe harbor regimes, online intermediaries are required to take measures strictly of reactive nature.

This regime of conditional liability was envisaged as one of the necessary means for the development of online services and the flourishing of the information society.[33] At the same time it also resulted from the gatekeeping function of intermediaries.[34] In recent years, the evolving nature of the intermediaries and the multiplicity of services and functions they provide has positioned the conditional liability regime at the heart of the debate on intermediaries' fitness to regulate the increasingly complex phenomenon of illegal content online.[35] Despite the profoundly changed market, European institutions have nevertheless declared the safe harbors fit for purpose in the

---

[29] *See generally* Miquel Peguera, *The DMCA Safe Harbors and Their European Counterparts: A Comparative Analysis of Some Common Problems*, 32 COLUM. J. L. & ARTS 481 (2009).

[30] *See* 17 U.S.C. § 512 (1998).

[31] Jennifer Bretan, *Harboring Doubts About the Efficacy of 512 Immunity under the DMCA*, 18 BERKELEY TECHNOL. LAW J. 27, 43 (2003).

[32] *See supra* note 6, where recital 46 states that the removal and disabling of access should be dealt with at national level.

[33] *See* Carsten Ullrich, *Standards for Duty of Care? Debating Intermediary Liability from a Sectoral Perspective*, 8 J. INTELL. PROP. INFO. TECH. & ELEC. COM. L. 111 (2017).

[34] *See* Jonathan L Zittrain, *A History of Online Gatekeeping*, 19 HARV. J.L. & TECH. 253 (2006).

[35] *See* Niva Elkin-Koren, *After Twenty Years: Revisiting Copyright Liability of Online Intermediaries, in* THE EVOLUTION AND EQUILIBRIUM OF COPYRIGHT IN THE DIGITAL AGE 29 (S. Frankel & D. Gervais eds., Cambridge Univ. Press, 2017); *see also* Peggy Valcke, Alexandra Kuczerawy & Pieter-Jan Ombelet, *Did the Romans Get It Right? What Delfi, Google, eBay and UPC TeleKabel Wien Have in Common*, *in* THE RESPONSABILITIES OF ONLINE SERVICE PROVIDERS 101 (M. Taddeo & L. Floridi eds., 2017).

many documents adopted to tackle illegal content online.[36] As a consequence of their so declared fitness, the safe harbors stay intact. Yet, the doubt arises as to whether such fitness is more apparent than real. In fact, several recently adopted provisions that touch upon the issue of platforms liability for illegal content—namely the provisions pointing in the direction of algorithmic enforcement—erode the shield offered to online intermediaries under the e-Commerce Directive,[37] to the extent that we can say that although within the DSM strategy safe harbors have not been directly revised, their *indirect* revision has certainly taken place.

III. ALGORITHMIC ENFORCEMENT IN THE DSM STRATEGY

The problem-driven approach to content regulation consistently promoted by the EU is what prompted the adoption of several instruments, each tackling a content of specific nature. The most debated of these instruments is surely the DSM Directive targeting copyright infringing content.[38] However, provisions on content harmful to minors, terrorist content, IPRs infringing content and misleading content are encompassed in, respectively, the Amending Directive Audio Visual Media Services ('AVMSD'),[39] the Regulation on preventing the dissemination of terrorist content online ('TERREG'),[40] the Guidance on certain aspects of the Directive on enforcement of intellectual property rights ('IPRs Enforcement Guidance'),[41] and the Guidance on unfair commercial practice ('UCPD Guidance').[42] In line with the proposition that platforms should bear more responsibility for their role in providing access to

---

[36] *See supra* note 24, at 9.

[37] *See* Giancarlo Frosio, *To Filter or Not to Filter? That Is the Question in EU Copyright Reform*, 32 CARDOZO ARTS ENTERTAIN. L. J. 331, 348-349 (2018). *See also* Maria Lilla Montagnani & Alina Trapova, *Safe Harbours in Deep Waters: A New Emerging Liability Regime for Internet Intermediaries in the Digital Single Market*, 26 INT. J. L. INF. TECHNOL. 284, 306-307 (2018); Maria Lillà Montagnani & Alina Trapova, *New Obligations for Internet Intermediaries in the Digital Single Market—Safe Harbors in Turmoil?*, 22 J. INTERNET L. 3, 8 (2019).

[38] *See* discussion *infra* Section III.

[39] *See generally* Council Directive 2018/1808 of the European Parliament and of the Council of 14 November 2018 Amending Directive 2010/13/EU on the Coordination of Certain Provisions Laid Down by Law, Regulation or Administrative Action in Member States Concerning the Provision of Audiovisual Media Services, 2018 O.J. (L 303/699).

[40] *See generally* European Parliament Legislative Resolution P8_TA(2019)0421 of 17 April 2019 on The Proposal for a Regulation of the European Parliament and of the Council on Preventing the Dissemination of Terrorist Content Online, COD (2018) 331.

[41] *See generally Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on Guidance on Certain Aspects of Directive 2004/48/EC of the European Parliament and of the Council on the Enforcement of Intellectual Property Rights*, COM (2017) 708 final (Nov. 29, 2017).

[42] COMMISSION STAFF WORKING DOCUMENT, GUIDANCE ON THE IMPLEMENTATION/APPLICATION OF DIRECTIVE 2005/29/EC ON UNFAIR COMMERCIAL PRACTICES, (COM 320) (2016).

information and content, all of these rules point towards algorithmic enforcement, namely enforcement through technological measures and autonomous decision systems.[43] The use of technology to comply with the law is certainly not new to market operators,[44] nor is the use of technological measures providing some degree of automation in the enforcement of copyright protection.[45] The novelty here lies in the legislative drive and legitimization of the adoption of such systems to enforce content regulation.

The push for the use of automated measures to filter and prevent illegal content from appearing online was very clear in the initial Commission proposals of the DSM Directive,[46] the AVMSD,[47] and the TERREG.[48] The proposed texts raised significant concerns. These concerns provoked years of debate on several aspects of the texts. One such debate centered on the use of technology with regard to the enforcement of content regulation and tackling illegal content online and its consistency with the conditional liability regime.[49] The provisions that were eventually adopted reflect the debate and considerably soften the push towards the use of automated systems to detect or remove illegal content.[50] Although not as clearly expressed as in the initial versions, algorithmic enforcement still remains the anonymous protagonist of content regulation in the DSM Strategy.

---

[43] *Id.*

[44] Kenneth A Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 73 669, 729 (2009).

[45] *See infra* Section III A.

[46] *See Proposal for a Directive of the European Parliament and of the Council On Copyright in the Digital Single Market*, COM (2016) 593 final (Oct. 12, 2016); *see also* Open letter to the European Commission by 40 academics [https://perma/cc/DR76-GZFV] (2017); *see also* Open Letter to Members of the European Parliament and the Council of the European Union, *Copyright Reform: Open Letter #2 from European Research Centers* [https://perma.cc/4686-MNQS]; *see also* Axel Metzger & Mathias Leistner, *The EU Copyright Package: A Way Out of the Dilemma in Two Stages*, 48 IIC 381 (2017).

[47] *See Proposal for a Directive of the European Parliament and of the Council Amending Directive 2010/13/EU on the Coordination of Certain Provisions Laid Down by Law, Regulation or Administrative Action in Member States Concerning the Provision of Audiovisual Media Services in View of Changing Market Realities* COM (2016) 0287 final; *see also* Indrek Ibrus & Ulrike Rohn, *Sharing Killed the AVMSD Star: The Impossibility of European Audio-Visual Media Regulation in the Era of the Sharing Economy*, 5 Internet Policy Review 11 (2016).

[48] *Proposal for a Regulation of the European Parliament and of the Council on Preventing the Dissemination of Terrorist Content Online*, COM (2018) 640 final; *see also* Joris van Hoboken, *The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications* (2019), https://www.ivir.nl/publicaties/download/TERREG_FoE-ANALYSIS.pdf

[49] *See supra* note 46-47.

[50] DSM Directive, *supra* note 7, at Art. 17; AVMSD, *see supra* note 39, at Art. 28(b); TERREG, *see supra* note 40, at Art. 6.

I will illustrate that the above mentioned provisions, despite the amendments in their wording, still allude to algorithmic enforcement.[52] My discussion then turns to the soft law instruments on misleading and IPR infringing content. While soft law is not binding, [53]the principles these rules enshrine also deserve attention as they contribute extensively to the EU's content regulation framework and further confirm the European institutions' push towards algorithmic enforcement.[54]

A. Misleading Content in the DSM Strategy

In the pursuit of the European institution's goal to stimulate and promote consumer confidence in the digital market, the UCPD Guidance introduces new duties for online intermediaries hosting content.[55] Given the growing body of principles derived from the decisions of the Court of Justice of the EU ('CJEU') and several national courts, the Commission adopted the Guidance to clarify some concepts and provisions of the Directive on unfair commercial practices ('UCPD').[56]

First, the UCPD Guidance extends its discipline to online platforms.[57] The Guidance declares that the unfair commercial practices discipline also applies to the new business models that develop in a digital environment—that is, to an online platform that qualifies as a "trader" according to article 2(b) of the UCPD. Thus, a platform that, while acting for purposes relating to its business, charges a commission on the transactions between suppliers and users, provides additional paid services, or draws revenues from targeted advertising, can definitely be deemed to be a trader, falling within the scope of the UCPD.[59]

Second, online platforms qualifying as traders must comply with professional diligence and transparency requirements, which require them to refrain from misleading actions and omissions while intermediating the promotion, sale, or supply of a product to consumers.[60]

To this end, online platforms must take appropriate measures to enable relevant third-party traders to comply with EU consumer and marketing law requirements and help users to clearly understand with whom they are

---

[52] *See Id.*

[53] K. C. Wellens & G. M. Borchardt, *Soft Law in European Community Law*, 14 EUR. L. REV. 267 (1989).

[54] *See Id.*

[55] UCPD Guidance, *supra* note 42.

[56] *See generally* Council Directive 2005/29/EC of May 11 2005 on Unfair Business-to-Consumer Commercial Practices in the Internal Market, 2005 O.J. (L 149/22).

[57] *See* UCPD Guidance, *supra* note 42, at 110.

[59] UCPD, *supra* note 56, at Art. 5.

[60] *Id.* Art. 6 and Art. 7.

concluding contracts.[61] Appropriate measures in this context imply, among other things, designing the structure of the website in a way that allows professional third parties to present information to users of the platform in compliance with Union rules on commercial law and consumers.[62] More particularly, according to Article 7(4) of the UCPD, this relates to information on the invitations to purchase.[63]

It is this particular obligation that triggers the need for platforms to resort to technologies of compliance. In fact, pursuant to the UCPD Guidance, online platforms may be considered liable if they do not take appropriate measures to avoid the upload of misleading content.

This obligation raises concern as to its consistency with the liability exemption regime established by the e-Commerce Directive. Article 14 of the e-Commerce Directive is often invoked by platforms claiming that, as mere hosts, they are not to be held liable for the information hosted, even when this violates consumer protection and constitutes an unfair commercial practice.[66] Even though in principle, this liability regime and the relevant consumer protection acquis communautaire should apply in a complementary manner,[67] it is unlikely to occur in practice. Instead of being complementary, the UCPD and the e-Commerce Directive seem to be alternatives of one another, if not conflicting. First, by expressly requiring that platforms comply with the UCPD's obligations the UCPD Guidance is only going to increase the number of clashes between platforms invoking Article 14 safe harbor for their hosting activity and internet users who, harmed by the misleading content hosted, complain about a lack of compliance with the UCPD provisions. Second, and more importantly, online platforms that fall within the scope of the UCPD and comply with the professional diligence requirements by adopting measures that allow them to intervene in the content, or by implementing filtering systems, might no longer fall within the safe harbor regime. Third, the interpretation of the obligations imposed by the UCPD is not in line with the prohibition on general monitoring as per Article 15 of the e-Commerce Directive. In fact, Article 5 of the UCPD introduces a "duty of activation" that could easily morph into a general obligation to carry out fact finding.[68]

---

[61] UCPD Guidance, *supra* note 42, at 114.

[62] *Id.*

[63] UCPD, *supra* note 56, at Art. 7(4).

[66] e-Commerce Directive 2000/31/EC, *supra* note 6, at Art. 14.

[67] *Id.*, at Art. 3(1).

[68] UCPD, *supra* note 56, at Art. 5.

B. IPRs Infringing Content in the DSM Strategy

Modernizing the enforcement of IPRs is one of the promises of the DSM Strategy that goes hand in hand with the issue of tackling illegal content online. In fact, one of the many consultations carried out by the European Commission within the DSM Strategy concerned the need to evaluate and modernize the current legal framework for IPR enforcement.[70] The primary topic was to ensure that Directive 2004/48/EC ('IPR Enforcement Directive') actively contributed to ensure a safe online environment for business operators and consumers. The consultation, however, also addressed a number of tangent issues, including the intermediaries' role in enforcing IPRs against illegal content online. [71]

As a result of the public consultation, in 2017 the Commission adopted the IPR Enforcement Guidance, which has as its main aim the optimization of the IPR Enforcement Directive.[72] In doing this, the Commission recalls the several initiatives on tackling illegal content online and strives to coordinate with them its actions on IPR enforcement. While declaring that the safe harbor regime stays intact, it attempts to clarify the intermediaries' responsibility in detecting and removing illegal online content, including content infringing IPRs through the adoption of autonomous technological systems that monitor and filter content online.

In particular, referring to filtering systems by intermediaries, the IPR Enforcement Guidance draws the line between, on the one hand, an injunction requiring a specific content to be removed from a website and, on the other, a broader injunction potentially obliging an intermediary to actively monitor all content made available on its platform.[73] The Commission specifies that the latter is prohibited according to Article 15 of the e-Commerce Directive as it would prompt intermediaries to "install and operate excessively broad, unspecific and expensive filtering systems of the type and in the circumstances at issue in the *Scarlet Extended* and *SABAM* cases."[74] However, it also states that where appropriate and within the limits of the above-mentioned provisions the adoption of a reasonable *duty of care* to detect and prevent certain specific types of illegal activities may be imposed.[75]

---

[70] EUROPEAN COMMISSION, Public Consultation on the Evaluation and Modernisation of the IPR Enforcement Framework (2015), https://ec.europa.eu/growth/content/public-consultation-evaluation-and-modernisation-ipr-enforcement-framework-0_en

[71] *See* Council Directive 2004/48/EC of April 30 2004 on The Enforcement of Intellectual Property Rights, 2004, O.J. (L 157) 23.

[72] IPR Enforcement Guidance, *supra* note 41, at 38.

[73] *Id.* at 20.

[74] *Id*.

[75] *Id*.

This move towards algorithmic enforcement of IP law in the IPR Enforcement Guidance sits somewhat uncomfortably with the e-Commerce Directive. Similar to the case of misleading content,[76] it raises the issue of how a filtering system can comply with the prohibition on general monitoring. Eventually, these algorithmic enforcement provisions seem to share the same internal clash: algorithmic enforcement complies with trend for adopting and utilizing technological systems that enable platforms to tackle illegal content pursuant to the DSM strategy but this does not coherently fit with the currently untouched conditional liability regime pursuant to the e-Commerce Directive.

C. Harmful Content in the DSM Strategy

A public consultation on the review of the AVMSD took place between 6 July 2015 and 30 September 2015.[77] The objective was to make the European audio-visual media landscape fit for, and aligned to, the 21st century media framework. As a result of the consultation, in May 2016 the Commission proposed a revision of the AVMSD,[78] which also encompassed provisions on combating hate speech and dissemination of harmful content to minors. This introduced new obligations for AVMS operators as they now became involved first-hand in tackling illegal content—in particular, content harmful to minors and hate speech.

The proposed revision of the AVMSD did not raise as much concern as the proposal for the DSM Directive.[79] The amended version was finally adopted in November 2018 by the European Parliament and the Council.

Before illustrating how the revised AVMSD encourages the use of technology for compliance, it is first necessary to turn to the definition of video-sharing platforms as they are now the specific addressees of the new set of obligations. Video-sharing platforms are defined as commercial services addressed to the public, where the principal purpose of the service (or an essential functionality of it) is devoted to providing programs and user-generated videos to the general public.[80] The scope of such service must then be that of informing, entertaining or educating, whereby the means adopted for

---

[76] *See infra* Section II A.

[77] *See generally* EUROPEAN COMMISSION, *Report on the Public Consultation on the Review of the Audiovisual Media Services Directive (AVMSD)* (2016), https://ec.europa.eu/digital-single-market/en/news/report-public-consultation-review-audiovisual-media-services-directive-avmsd.

[78] *See generally Proposal for a Directive of the European Parliament and of the Council Amending Directive 2010/13/EU on the Coordination of Certain Provisions Laid Down by Law, Regulation or Administrative Action in Member States Concerning the Provision of Audiovisual Media Services in View of Changing Market Realities*, *supra* note 47.

[79] *See infra* Section III.

[80] AVMSD, see supra note 39, at Art. 1.

it should be through electronic communications networks; and, finally, the content should be organized in a way determined by the provider of the service, in particular by displaying, tagging and sequencing.[81]

Such video-sharing platforms fall within the regulation introduced by the newly-adopted Chapter IXa, entitled "Provisions applicable to Video-Sharing Platform Services". In particular, Article 28b of the AVMSD brings in an obligation for Member States to ensure that appropriate measures are taken by video-sharing platforms to protect: first, minors from content which may impair their physical, mental or moral development; second, the general public from content containing incitement to violence or hate speech; and third, still the general public from content the dissemination of which constitutes criminal offence, namely "public provocation to commit a terrorist offence", "child pornography" and "racism and xenophobia".

Measures are appropriate "in light of the nature of the content in question, the harm it may cause, the characteristics of the category of persons to be protected as well as the rights and legitimate interests at stake".[82] Moreover, they should be practicable and proportionate, taking into account the size of the video-sharing platform service and the nature of the service it provides.[83] In addition, and more importantly, such measures should not lead to any *ex ante* control or upload-filtering of content insofar such actions are not in compliance with Article 15 of the e-Commerce Directive.

In particular, appropriate measures may include reporting/flagging operation systems, operating systems through which video-sharing platform providers explain to their users what effect has been given to the reporting and flagging systems, age verification systems, content rating systems, parental control systems, media literacy measures and tools, and raising users' awareness of those measures and tools, and finally, easy-to-use and effective procedures for the handling and resolution of users' complaints to the video-sharing platform provider in relation to the implementation of the measures.[84] At first sight, having examined the appropriate measures suggested, the push towards the adoption of algorithmic systems here seems quite modest. Yet, turning to Article 28b one grasps the full algorithmic enforcement potential of these new provisions. In fact, Article 28b(6) permits Member States to impose on video-sharing platform measures that are more detailed or stricter than those mentioned above. This is the most controversial part of Article 28b as it raises the question as to what measures could be stricter than the measures above mentioned if not filtering systems that prevent harmful content and hate speech

[81] *Id*.
[82] *Id*. at Art. 28(b)(3).
[83] *Id*.
[84] *Id*.

from being uploaded. It seems that these can only be technological systems that autonomously detect illegal content and block its appearance online.

The scenario is even more complicated as, even here, the adoption of these stricter measures needs to be reconciled with the ban on general monitoring obligations set by Article 15 of the e-Commerce Directive. In fact, when adopting stricter measures online intermediaries should still "comply with the requirements set out by […] Articles 12 to 15 of Directive 2000/31/EC," which means in turn that such measures "shall not lead to any *ex ante* control measures or upload-filtering of content which do not comply with Article 15 of Directive 2000/31/EC".[85] Now, it is still to be determined how an autonomous system monitoring the content posted to a platform can detect harmful to minor content or hate speech without the recourse to a general monitoring mechanism. Although the AVMSD maintains that the regime introduced by the e-Commerce Directive is untouched,[86] the issue of how to reconcile the new duties with the ban imposed by Article 15 remains highly critical.

D. Online Terrorist Content in the DSM Strategy

As part of the strategy to tackle illegal content online, from 30 April to 25 June 2018 the Commission conducted a public consultation on terrorist content. This initiative resulted in the proposal for a Regulation on preventing the dissemination of terrorist content online—the TERREG. This proposed Regulation encompassed stringent rules for online intermediaries in relation to content of terrorist nature, such as a one-hour deadline for content to be removed following a removal order from national competent authorities, a duty of care for all platforms to ensure they are not misused for the dissemination of terrorist content, and *proactive* measures on the side of platforms to protect their users from terrorist abuse.[87]

The proposed text of the regulation was strongly opposed for its impact on fundamental rights and in particular, freedom of expression.[88] It also attracted the concern of three independent experts of the United Nations Human Rights Council.[89] In this debate, highly problematic was Article 6 which

---

[85] *Id*. at Art. 28b(6).

[86] *Id*. at Recital (48).

[87] Proposal for a Regulation of the European Parliament and of the Council on Preventing the Dissemination of Terrorist Content Online, *supra* note 48, at Art. 6.

[88] *See generally* van Hoboken, *supra* note 48.

[89] *See* David Kaye, Joseph Cannataci & Fionnuala Ní Aoláin, *Mandates of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression: The Special Rapporteur on the Right to Privacy and the Special Rapporteur on the Promotion and Protection of Human Rights and Fundamental Freedoms While Countering Terrorism* (2018), https://spcommreports.ohchr.org/TMResults Base/DownLoadPublicCommunicationFile?gId=24234.

required the adoption of proactive measures that would detect, identify and expeditiously remove or disable access to terrorist content and would also prevent the re-upload of content that has been removed or disabled. The proposal expressly referred to the use of automated tools.

These proactive measures raised a twofold concern which mirrors what was already mentioned in respect of misleading, harmful and IPR infringing content. First, these measures were deemed to violate Article 15 of the e-Commerce Directive as they would amount to a general obligation to monitor.[90] In addition, such general monitoring and filtering of content uploaded by users risk to block content without any form of due process even before such content is published. This would reverse the well-established presumption that States, not individuals, bear the burden of justifying restrictions on freedom of expression. Second, these proactive measures could also deprive platforms from the protection against liability for third-party content as per Article 14 of the e-Commerce Directive.[91]

Following this debate, the version finally adopted in April 2019[92] carries several amendments. On the one hand, the modifications mitigate the impact of the new rules on platforms' operativity and their freedom to conduct business by limiting the measures and actions that platforms should undertake, and on the other hand, the amended rules increase the safeguards for fundamental rights by narrowing the definition of terrorist content.[93]

The aim remains that of tackling terrorist content online as this is "part of a broader problem of illegal content online, which includes child sexual exploitation, illegal commercial practices and breaches of intellectual property".[94] Broadly speaking, the TERREG addresses online hosting services through which terrorist content is disseminated. More specifically, it applies to information society services storing user-provided information which is also made available to the public, irrespective of whether this activity is of a mere technical, automatic and passive nature.[95] Importantly, the TERREG also applies to hosting service providers who fall within the definition of video-sharing platforms provided in the AVMSD.[96]

---

[90] *Id*. at 9.

[91] *Id*.

[92] TERREG, *supra* note 40.

[93] Riis & Schemer, *supra* note 3, at 2.3.

[94] TERREG, *supra* note 40, at Recital (1b).

[95] *Id.* at Recital (10), (clarifying that "providers of information society services include social media platforms, video streaming services, video, image and audio sharing services, file sharing and other cloud services to the extent they make the information available to *the public* and websites where users can make comments or post reviews.").

[96] *Id.* at Art. 2(1)(1).

The broad categories of intermediaries identified must still comply with a wide array of duties of care provisions, removal orders, referrals and "specific measures".[97] In particular, "specific measures" now replaces the "proactive measures" proposed by the Commission. The measures should operate without prejudice to the e-Commerce Directive and should be "effective, targeted and proportionate, paying particular attention to the risk and level of exposure to terrorist content, the fundamental rights of the users, and the fundamental importance of the right to freedom of expression and the freedom to receive and impart information and ideas in an open and democratic society."[98]

Beside the change in the terminology, Article 7 of TERREG also carries several other amendments such as the fact that it is now clearly stated that intermediaries are not obliged to monitor or filter the content,[99] but the obligation to withdraw the illegal content within an hour remains.[100]

IV. ALGORITHMIC COPYRIGHT ENFORCEMENT IN THE DSM

The use of technology to enforce the law is not a new concept,[101] neither is the use of automatic systems to enforce copyright law in the entertainment industry.[102]

The enforcement of copyright law online has always entailed a certain level of automation. However, over the last few years, automation has been progressively replaced with autonomy. While the former automated systems typically run within a well-defined set of parameters and are trained in what tasks they can perform, autonomous systems learn and adapt to the surrounding environments. In other terms, what distinguishes automation from autonomy is the amount of adaptation, learning and decision-making.

In the following sections the story of technologies for copyright enforcement is articulated in three phases, each representing a step from automation towards autonomy. The last step is algorithmic copyright enforcement, which is currently heading towards the adoption of artificial

[97] *Id.* at Recital (16) and Art. 6.
[98] *Id*. at Art. 6(1).
[99] *Id.* at Art. 3(1)(a).
[100] *Id.* at Art. 4 (2).
[101] *See generally* Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016).
[102] *See generally* Toni Lester & Dessislava Pachamanova, *The Dilemma of False Positives: Making Content ID Algorithms More Conducive to Fostering Innovative Fair Use in Media Creation*, 51 UCLA ENT. L. REV. 24 (2017).

intelligence (AI) and machine learning-based programs to tackle illegal content online.[103]

A. The Early Phase of Algorithmic Copyright Enforcement: The Robo-notice Regime

The first form of automation that copyright enforcement experienced was the adoption of automated notices by right holders on the basis of internet monitoring systems. To this, online platforms responded by adopting systems that automatically removed the signaled content without any form of human intervention.[104]

The reason for the adoption of automated systems is usually found in the increase of unauthorized material online that in turn resulted in difficulties to humanly monitor the spread of unauthorized content. Once human monitoring of content online became impractical, right holders started resorting to the development and use of automated systems that crawl the internet to detect allegedly illegal content. Thereafter, notices were automatically sent to hosting platforms.

On the other hand, to match this trend, the automation of the notices triggered automation in the taking down of the indicated content. In order to limit the risks of falling outside the safe harbor regime online platforms responded to the notices by automatically taking down the content ("automated take-down").[105] As a result, the higher the number of notices, the higher the amount of take-downs.

This phenomenon, that goes under the name of "robo-takedown regime"[106] has exacerbated the general public frustration already provoked by the N&TD and increased the amount of content erroneously taken down.[107]

---

[103] *See* James Vincent, *Instagram is Using AI to Detect Bullying in Photos and Captions*, THE VERGE (Oct. 9, 2018, 9:00 AM), https://www.theverge.com/2018/10/9/17954658/instagram-ai-machine-learning-detect-filter- bullying-comments-captions-photos.

[104] *See* Daniel Seng, *The State of the Discordant Union: An Empirical Analysis of the State of DMCA Takedown Notices*, 18 VA. J. L. & TECH. 369, 414-417 (2014); Martin Husovec, *The Promises of Algorithmic Copyright Enforcement: Takedown or Staydown: Which Is Superior? and Why*, 42 COLUM. J.L. & ARTS 53 (2018) (describing the phenomenon).

[105] Michael W. Carroll, *Pinterest and Copyright's Safe Harbors for Internet Providers*, 68 U. MIAMI L. REV. 421, 424 (2014).

[106] Zoe Carpou, *Robots, Pirates, and the Rise of the Automated Takedown Regime: Using the DMCA to Fight Piracy and Protect End-Users*, 39 COLUM. J.L. ARTS 551, 552-53 (2016).

[107] Jennifer M. Urban & Laura Quilter, *Efficient Process or "Chilling Effects"? Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 621, 623 (2006); *see generally* Bar-Ziv, Sharon & Niva Elkin-Koren, *Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown*, 50 CONN. L. REV. 339 (2018).

Automation significantly impacts the scope of permitted uses that third parties are allowed to rely on. It questions the traditional balance that copyright entails between protection of creative works and authorized uses by third parties.

The matter has also been addressed in court, at least in the US. In the *Lenz* case, the practice of using algorithms to send automated notices raised the question as to whether right holders should rely on the same algorithms used to identify potential infringement to also make a judgment about fair use. On this point, the Ninth Circuit affirmed that "fair use … is wholly authorized by the law"[108] which makes it compulsory for right holders to consider, before issuing a takedown notice, whether the possibly infringing content falls within fair use.[109] However, assessing fair use may prove problematic for any automated system. Distinguishing critical reviews, parodies, and transformative remixes from infringing reuses of copyrighted material often involves the kind of contextual decision-making that is already difficult for humansbut proves to be even more complicated, if not impossible, for algorithms.[110]

In addition to the evidence that non-infringing content is constantly removed pursuant to robo-takedown requests, the robo-notice regime also neglects the issue of due process. Indeed, the number of counternotices adopted in response to the algorithmic take-down represents an even smaller percentage than the overall number of counternotices sent and received.[111] In sum, the rise of mass notice sending via automated systems provokes immediate questions of both accuracy and due process.[112]

B. The Advanced Phase of Algorithmic Copyright Enforcement: The Voluntary Filtering Regime

The second level of automation came when some platforms voluntarily started to technologically implement agreements signed with right holders—

---

[108] Lenz v. Universal Music Corp., 815 F.3d 1145, 1151 (N.D. Cal. Apr. 8, 2008).

[109] Matthew Schonauer, *Let the Babies Dance: Strengthening Fair Use and Stifling Abuse in DMCA Notice-and-takedown Procedures*, 7 I/S: J.L. & Pol'y For Inf. Soc'y 135, 156 (2011).

[110] Matthew Sag, *Internet Safe Harbors And The Transformation Of Copyright Law*, 93 Notre Dame L. Rev. 499, 531 (2017); *see* Lydia Pallas Loren, *Deterring Abuse of the Copyright Takedown Regime by Taking Misrepresentation Claims Seriously*, 46 Wake Forest L. Rev. 745, 747 (2011) (criticizing the automation of notice and take-down); *see also* Schonauer, *supra* note 104, at 156.

[111] Carpou, *supra* note 106, at 573; *see* Annemarie Bridy & Daphne Keller, *U.S. Copyright Office Section 512 Study: Comments in Response to Notice of Inquiry* (Mar. 30, 2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757197.

[112] *See* Jennifer M. Urban, Joe Karaganis & Brianna L. Schofield, Notice and Takedown in Everyday Practice 116 (Mar. 29, 2016) (UC Berkeley Public Law Research Paper No. 2755628).

namely, entertainment majors—to share the revenues generated by the phenomenon of user-generated content.[113]

The classic example is the Content ID system developed by YouTube.[114] Broadly speaking, this and similar systems match whatever content is uploaded by users with the content for which right holders claim copyright and require protection.[115] Platforms then enable right holders to decide what to do when the content uploaded matches their content—to either block it, monetize it, or just monitor it.[116] In practical terms, the system works by comparing existing and newly uploaded contents to "index files" of video or audio material provided by right holders. If a user-uploaded video is matched with an audiovisual work in the reference file, the respective right holder is notified and has the choice as to what, if any, action to take.[117]

These systems are the result of the pressure exercised by right holders (in particular, the music industry) on intermediaries and the attempt to involve them even more in the fight against piracy.[118] In the pursuit of the liability rules within the e-Commerce Directive, platforms not only agree to perform a gatekeeping function, but they also encode that function in algorithms and software.[119] By doing this, they go even beyond what is required by law, namely the N&TD under the DMCA,[120] or the expeditious removal of the infringing

---

[113] *See* Edward Lee, *Warming Up to User-Generated Content*, U. ILL. L. REV. 1459 (2008).

[114] *See* Lauren D. Shinn, *Youtube's Content ID as a Case Study of Private Copyright Enforcement Systems*, 43 AIPLA Q.J. 359 (2015).

[115] *How Content ID Works*, YOUTUBE, https://support.google.com/youtube/answer/2797370?hl=en [https://perma.cc/CET5-3RHC] (last visited Nov. 7, 2019).

[116] *Id.*

[117] *Id.*

[118] *See, e.g.*, INTERNATIONAL FEDERAL FOR THE PHONOGRAPHIC INDUSTRY, *IFPI Global Music Report 2019* (Apr. 2, 2019), https://ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2019 [https://perma.cc/Z34U-Q9RN].

[119] Sag, *supra* note 110, at 538-41.

[120] Digital Millennium Copyright Act of 1998 (DCMA), 17 U.S.C. § 512(c)(1) (2010) ("A service provider shall not be liable...if the service provider--(A)(i) does not have actual knowledge that the material or an activity using the material on the system or network is infringing; (ii) in the absence of such actual knowledge, is not aware of facts or circumstances from which infringing activity is apparent; or (iii) upon obtaining such knowledge or awareness, acts expeditiously to remove, or disable access to, the material; (B) does not receive a financial benefit directly attributable to the infringing activity, in a case in which the service provider has the right and ability to control such activity; and (C) upon notification of claimed infringement as described in paragraph (3), responds expeditiously to remove, or disable access to, the material that is claimed to be infringing or to be the subject of infringing activity.").

content in the e-Commerce Directive.[121] Instead, they voluntarily adopt filtering systems that detect possibly infringing content uploaded by users.[122]

Compared with the robo-notice regime in the first stage, this second level of automation takes place in the absence of specific regulation which leads to an amplification of the risks present in the previous phase. The N&TD procedure, even when automated, occurs within a legal framework providing mandatory safeguards for users.[123] Instead, filtering systems voluntarily adopted by right holders are unregulated, which means that safeguards for third party users are entirely left to the discretion of platforms and depend on choices made by rights holders and platforms.[124]

In addition, filtering systems share the same concern raised by the robo-notice regime as to the permitted uses reserved to third users. Even here, the capacity of automated filtering systems to distinguish between infringing content and content that constitute a permitted use, such as for example parodies or educational content, is questionable. To address this issue, platforms have already started developing filtering systems that employ AI and machine learning to distinguish between infringing content and permitted content.[125] This is where the shift from automation to autonomy starts to materialize. The more sophisticated filtering systems become, the more they walk away from automation to enter the realm of autonomy, i.e. a realm in which the decision as to whether particular content amounts to a permitted use is left to an algorithm.

C. The Current Phase of Algorithmic Copyright Enforcement: Article 17 of the Directive on Copyright in the DSM

The last phase of copyright enforcement is represented by Article 17 of the DSM Directive. In a nutshell, this much controversial provision introduces the direct liability of "online content sharing service providers storing and giving access to large amounts of works"[126] for the content that they host. It imposes on them an obligation to either sign licensing agreements with right

---

[121] e-Commerce Directive, *supra* note 6, at Art. 14.

[122] *See supra* Section III. A.

[123] Sag, *supra* note 110, at 543

[124] *Id.* at 544.

[125] *See* Kalev Leetaru, *The Problem With AI-Powered Content Moderation Is Incentives Not Technology*, FORBES (Mar. 19, 2019, 2:34 P.M.), https://www.forbes.com/sites/kalevleetaru/2019/03/19/the-problem-with-ai-powered- content-moderation-is-incentives-not-technology/ [https://perma.cc/RU7K-JWCY].

[126] DSM Directive, *supra* note 7, at 112-113 ("Providers of services, such as not-for-profit online encyclopedias, not-for-profit educational and scientific repositories, open source software-developing and-sharing platforms, providers of electronic communications services as defined in Directive (EU) 2018/1972, online marketplaces, business-to-business cloud services and cloud services that allow users to upload content for their own use are not 'online content-sharing service providers' within the meaning of this directive.").

holders with respect to the content that the service provider stores, or to otherwise prevent protected content from (re)appearing online.

More in detail, since "online content sharing service providers perform an act of communication to the public", they should conclude fair and appropriate licensing agreements with right holders.[127] These agreements also cover "the liability for works uploaded by the users of such online content sharing services in line with the terms and conditions set out in the licensing agreement."[128] In case these agreements are not reached—that is, the licensing obligation is not complied with—another obligation is triggered. In order to avoid liability, hosting platforms shall, demonstrate that they not only have made "best efforts to obtain an authorisation" but, more significantly, that they have "made, in accordance with high industry standards of professional diligence, best efforts to ensure the unavailability of specific works and other subject matter for which the right holders have provided the service providers with the relevant and necessary information."[129] This last obligation implies the implementation of technological measures which, in a very similar fashion to the filtering systems developed in the previous phase, would screen what users upload. The system would then prevent the availability online of the works that match those indicated by right holders.[130]

In addition to the obligations above, Article 17(4)(c) also introduces a mechanism that can be defined as Notice & Stay Down (N&SD).[131] This entails that a single notification of an infringing work would now oblige an intermediary to forever prevent its reappearance on the platform. Even in the case in which the platform has actually managed to obtain a license—and in theory it is not obliged to adopt any technological measures pursuant to Article 17(4)(b)—in practice it is still mandated to adopt a system that will prevent content for which a notice has been sent from being uploaded again by another user.[132] This obligation can hardly be complied with if not through a system that filters the content that users constantly upload.

---

[127] *Id.* Art. 17(1), at 119.

[128] *Id.* Art. 17(2).

[129] *Id.* Art. 17(4)(a) and Art. 17(4)(b), at 119-120.

[130] MARTIN SENFTLEBEN, BERMUDA TRIANGLE–LICENSING, FILTERING AND PRIVILEGING USER-GENERATED CONTENT UNDER THE NEW DIRECTIVE ON COPYRIGHT IN THE DIGITAL SINGLE MARKET 4 (Apr. 4, 2019), https://www.ssrn.com/abstract=3367219, [https://perma.cc/E2T9-LPGS].

[131] Husovec, *supra* note 104.

[132] Martin Husovec, *General Monitoring of Third-party Content: Compatible with Freedom of Expression?*, 11 J. INTELL. PROP. L. & PRACT. 20 (2016); Giancarlo Frosio, *Why Keep a Dog and Bark Yourself? From Intermediary Liability to Responsibility*, 26 INT'L J. L. & INFO. TECH. 1 (2018).

Compared to Article 13 of the initial text of the DSM Directive proposed by the Commission in 2016,[133] in the current text the reference to technological measures of content recognition is not explicit. In this sense, Article 17 at first sight seems to take a step back and focus more on a licensing obligation rather than a filtering one. While Article 13 and its corresponding recitals (38) and (39) explicitly mentioned the use of technological measured to prevent illegal content from being uploaded, i.e. "effective content recognition technologies",[134] such mentions were all dropped in Article 17. However, the difference is more apparent than real. In this sense, unless exempted, online platforms are in any case forced to resort to the same filtering systems that they already started to voluntarily adopt in the previous phase.

The real change brought about by Article 17 is that it compels the adoption of filtering systems. It does this in two cases. This happens first, when a licensing agreement with right holders on the content uploaded by third parties on their platform cannot be reached. Here, pursuant to Article 17(4)(b), if platforms want to avoid being considered directly liable for communicating to the public the content uploaded by their users, they have to implement filtering systems to detect such content and block its upload. Absent an agreement, filtering systems become thus the standard to avoid liability. However, platforms already voluntarily adopt filtering systems to be sure to fall within the safe harbor set under Article 14 of the e-Commerce Directive.[135] Therefore, what the current Article 17 does is to make this practice compulsory to avoid not only secondary liability—as in the previous phase of algorithmic copyright enforcement—but also direct liability for infringement of the right of communication to the public. Second, by introducing the N&SD as a standard practice to be followed in any event, even in the presence of a license, Article 17 extends the adoption of filtering systems to all platforms, regardless of whether or not they have secured a license from the right holders. In other words, according to Article 17(4)(c), whenever a content is taken down, a platform should ensure that it is not re-uploaded by any third party. Again, as this is an obligation that can only be performed through filtering systems, these become the standard to comply with the N&TD introduced by Article 17 regime.

The adoption of Article 17—or better, its statutory request to adopt technologies to enforce the law—has legal and factual consequences. In the first place, its requests may amount to a general filtering obligation in violation of Article 15 of the e-Commerce Directive, regardless the proposition that "[t]he application of this Article shall not lead to any general monitoring

---

[133] *Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, *supra* note 46.
[134] *Id.* at 120.
[135] *See supra* Section III B-C.

obligation."[136] The DSM Directive implies that right holders can communicate information on the works—and other subject matter—which they do not want to be available online.[137] It is highly likely that these lists of works will be rather extensive and would eventually result in a quasi-general monitoring obligation on the side of the intermediaries.

On a more general basis, there is an issue of consistency between the DSM Directive and the safe harbor regime. In particular, there is a general lack of coordination between Article 17 of the DSM Directive and Article 14 of the e-Commerce Directive as interpreted by the CJEU case law in relation to active and passive hosting providers.[138] Specifically, according to Article 17(3) platforms that are addressees of the new rules do not fall anymore within Article 14 of the e-Commerce Directive. This carves out an exception in the safe harbor regime, which may raise issues as to where to draw the line between platforms for which the CJEU interpretation of active/passive hosting providers still applies and to assess those which fall within Article 17 of the new Directive.

In the second place, the system introduced by Article 17 shifts algorithmic enforcement from ex post to ex ante. In fact, while before the adoption of Article 17, ex post enforcement, i.e. the robo-notice regime,[139] coexisted with cases in which platforms voluntarily through filtering systems prevented content from being uploaded, [140] the regime under Article 17(4)(b) bring in a form of ex ante filtering obligation according to what is asked by right holders. Besides, Article 17(4)(c) introduces a system which is partially ex ante and partially ex post. Ex post when the first notice is sent, and ex ante afterwards, in relation to all the possible re-uploaders of that same content. Consequently, this new regime advances the threshold of protection granted to copyright law: it statutorily requires the use of algorithms not only to enforce copyright law once a violation has taken place, but also to prevent a violation from taking place in the first place. Eventually, this blurs the line between algorithmic enforcement and algorithmic content regulation. Such a changing nature of algorithmic enforcement, i.e. more and more ex ante, transforms it from a compliance tool into one of regulation.

---

[136] DSM Directive, *supra* note 7, at 120.

[137] *Id.* at 120.

[138] Eleonora Rosati, *The CJEU Pirate Bay Judgment and Its Impact on the Liability of Online Platforms*, 39 EIPR 737, 745 (2017); Jane C. Ginsburg & Luke Ali Budiardjo, *Liability for Providing Hyperlinks to Copyright-Infringing Content: International and Comparative Law Perspectives*, 41 COLUM. J. L. & ARTS 153, 176 (2018); Jane C. Ginsburg, *The Court of Justice of the European Union Creates an EU Law of Liability for Facilitation of Copyright Infringement: Observations on Brein v. Filmspeler [C-527/ 15] (2017) and Brein v. Ziggo [C-610/15]* 1, 13 (Columbia Law & Econ. Working Paper No. 572, 2017).

[139] *See supra* Section III A.

[140] *See supra* Section III B.

Third, Article 17 also pushes towards the implementation of increasingly autonomous filtering systems. While in the voluntary filtering systems phase, safeguards as far as legitimate uses of copyright content as well as consideration of fundamental rights are not explicitly required, the DSM Directive expressly calls for certain exceptions or limitations not to be affected and the data of individual users to be protected in accordance with Directive 2002/58/EC and Regulation (EU) 2016/679.[141] This approach introduces within the European legal framework the principle jurisprudentially developed in the US though the *Lenz* case.[142] Such a step requires that filtering systems ought to develop in a way that takes into consideration third parties interests as well as fundamental rights. A development of this kind can only occur by relying more heavily on AI and machine learning, the concern of which will be illustrated in the following section.

V. THE BEAUTIFUL AND THE UGLY OF ALGORITHMIC COPYRIGHT ENFORCEMENT

Regardless of the general hype around algorithmic enforcement, and its implications for copyright law in particular, such systems raise a fair amount of concern. Several scholars have already tried to mitigate such discontent. In the following, I will consider the concerns and address the possible mechanisms that could tackle the alarming consequences of algorithmic copyright enforcement.

A. Algorithmic Copyright Enforcement and its Shortcomings

A first cluster of concerns deals with the convergence of all law enforcement functions[143] in the hands of intermediaries that are private companies.[144] Through the privatization of a function that is usually public, we witness a delegation of power to online intermediaries.[145] It has been observed that automated systems become the "de facto delegations of rulemaking power"

---

[141] DSM Directive, *supra* note 7, at Art. 17(9).

[142] *Lenz*, *supra* note 108.

[143] *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1253-1254 (2007) ("Although programmers building automated systems may not intend to engage in rulemaking, they in fact do so. Programmers routinely change the substance of rules when translating them from human language into computer code."), for the description of the combining between rulemaking and adjudication.

[144] *See* Salung Petersen, *Private Enforcement of IP Law by Internet Service Providers: Notice and Action Procedures*, *in* USER GENERATED LAW: RE-CONSTRUCTING INTELLECTUAL PROPERTY LAW IN A KNOWLEDGE SOCIETY 228, 242 (Thomas Riis ed., Edward Elgar Pub. 2016).

[145] See Alexa Capeloto, *Transparency on Trial: A Legal Review of Public Information Access in the Face of Privatization*, 13 CONN. PUB. INT. L.J. 19, 20-21 (2013), for a discussion on the privatization of public functions.

as they are those who encode the law into decision-making programs.[146] Besides, the convergence of the enforcement functions merges all phases of the traditional case-by-case, human-driven process of detection, prosecution, adjudication and execution of the law into a one step process, which is now technologically performed.[147] For instance, in the case of copyright law, private online platforms employ algorithms to detect the possibly infringing content and decide whether it infringes. Then, on the basis of this, content is automatically made available online or not. This process is characterized by an evident lack of accountability on behalf of private platforms for the way in which their algorithms enforce the law.[148] Users affected by the algorithmic decision making are not able to understand the reasons why the content they tried to upload failed to appear online and what is more, are not able to challenge such decisions.

A lack of accountability makes it also difficult to correct autonomous decision-making derived from biased algorithms. It is in fact known that algorithmic decision-making may result in biased outcomes whenever the computer system systematically and unfairly discriminates against certain individuals or groups of individuals in favor of others. Regardless of the origin of bias—were it a pre-existing bias rooted in social institutions, practices, and attitudes, or a technical bias deriving from technical constraints or considerations, or an emergent bias arising in a context of use—the more complex a system is, the more biases remain hidden in the code, which are difficult to pinpoint or explain. Biased systems, thus, offer no means for appeal.[149]

The main excuse for such a lack of accountability is that algorithms are opaque by nature.[150] This inherent opacity would derive from algorithms being innately non-transparent as they amount to complex code that even programmers cannot easily comprehend. Indeed, algorithms now combine more than one decision tree to generate the required outcome and this adds to the

---

[146] *See y* Citron, *supra* note 143, at 1296 ("Delegations to computer systems and their programmers . . . are more troubling than delegations to private contractors. . . [Because] code writers lack policy expertise.").

[147] *See* Niva Elkin-Koren & Maayan Perel, *Algorithmic Governance by Online Intermediaries*, *in* THE OXFORD HANDBOOK OF INSTITUTIONS OF INTERNATIONAL ECONOMIC GOVERNANCE AND MARKET REGULATION 1, 3 (Eric Brousseau, Jean-Michel Glachant & Jérôme Sgard eds., Oxford Handbooks Online 2019).

[148] *See* Helen Nissenbaum, *Accountability in a Computerized Society*, 2 SCI. & ENGINEERING ETHICS 25, 26 (1996) (raising the issue on the lack of accountability on private platforms). *See also* Perel & Elkin-Koren, *supra* note 101 (arguing specifically on accountability in algorithmic copyright law enforcement).

[149] *See* Batya Friedman & Helen Nissembaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFO. SYSTEMS 330, 331 (1996).

[150] *See* Paul B. de Laat, *Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?*, 31 PHIL. & TECH. 525, 526 (2018).

issues of opacity that of explicability or interpretability.[151] Predictive algorithms are becoming harder to decipher *ex post*, which in turn makes it particularly complicated for the addressees of decision-making operations to challenge the effect of such decisions.[152] This is further exacerbated in the case of self-learning algorithms that, after being fed with large amounts of data, develop on their own and react to the environment in ways that cannot be foreseen.[153] This feature of autonomy renders algorithms even less accountable.[154] The performance of tasks with less, or entirely without, supervision, coupled with the algorithm's capacity to adjust according to the data collected in the course of the operation, make it impossible to predict ex ante the decisions that they will take.[155] Furthermore, scrutinizing ex post the result of such operations becomes if not unfeasible extremely difficult.[156] In addition, as new data constantly pours in and algorithms continue to evolve, even if interpretation were viable, it would be bound to change all the time, i.e. it would reveal information that is cogent to a specific moment but it would not disclose the pattern followed in the decision-making.[157]

On top of the inherent opacity of algorithms, there is also a layer of opaqueness artificially created by those who develop and employ them. It is indeed the norm for private institutions to consider algorithms as their intellectual property. In fact, most algorithms are proprietary (e.g. Google Search and Facebook News Feed) covered primarily by trade secrets.[158] This is surely the case within the European Union, where the recently revised Trade Secrets Directive prevents the unauthorized acquisition, use, or disclosure of algorithms covered by trade secrets.[159] Such protection is available as long as the algorithm is not generally known or easily accessible, has commercial value, and the person who has control of it takes steps to keep it secret.[160] Similarly, in the US, a trade secret is granted broad protection in the form of secrets that derive actual or potential independent economic value as long as reasonable

---

[151] *Id.* at 536.
[152] *Id.*
[153] *Id.*
[154] *Id.*
[155] *Id.* at 537.
[156] *Id.*
[157] *Id.* at 538.
[158] Guido Noto La Diega, *Against the Dehumanisation of Decision-Making:Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, 9 JIPITEC 3, 10 (2018).
[159] Directive 2016/943 of the European Parliament and of the Council of 8 June 2016 on the Protection of Undisclosed Know-How and Business Information (Trade Secrets) Against Their Unlawful Acquisition, Use and Disclosure, 2016 O.J (L 157), Art. 4.
[160] Mariateresa Maggiolino, *EU Trade Secrets Law and Algorithmic Transparency*, BOCCONI LEGAL STUD. RES. PAPER NO. 3363178, 1, 8 (2019), https://papers.ssrn.com/abstract=3363178.

efforts to maintain their secrecy is made.[161] Besides acting at state level, a lawsuit can also be brought at federal level as the Defend Trade Secret Act of 2016 allows an owner of a trade secret to sue in a federal court individuals or organization suspected of stealing confidential information.[162] Although differently defined according to the jurisdiction at stake, the broad notion of trade secret is likely to encompass algorithms.[163] Surely, the practice is right now that of keeping them the most secret possible.[164]

A first corollary of opacity and limited—or non-existent—accountability is the lack of due process, i.e. the absence of meaningful opportunities for affected individuals to contest algorithmic decisions.[165] In a system where algorithms become decision-makers and govern important aspects of individual lives, lack of due process could pave the way to a "new feudal order of unaccountable reputational intermediaries".[166]

Algorithmic copyright enforcement is emblematic of such a risk. In principle, the N&TD—at least in the US version which has become the de facto standard—provides for a counter-notice to be issued by a user whose content has been taken down. This system has however been labelled as ineffective in granting users the right to react to algorithmic decisions as the possibility is theoretically available but practically unfeasible.[167] Users are not inclined to respond to a notice with a counter-notice and tend to passively accede to the platforms' operations.[168]

Things get even worse as far as ex ante filtering and the N&SD are concerned. In both cases users are not given any possibilities to react and assert their rights as far as the uploaded content is concerned.[169] For example, although Article 17(9) of the DSM Directive requires some complaint and redress mechanisms to be adopted, it is for the platforms to define how these should work on the basis of a cooperation process between themselves.[170] This generates a disparity between the right of copyright holders and those of

---

[161] Sharon K. Sandeen & Christopher B. Seaman, *Toward a Federal Jurisprudence of Trade Secret Law*, 32 BERKELEY TECH. L.J. 829, 906 (2017).

[162] DEFEND TRADE SECRETS ACT 2016, 18 U.S.C. § 1836 (2016).

[163] Frank A. Pasquale, *Restoring Transparency to Automated Authority*, 9 J. ON TELECOMM. AND HIGH TECH. L. 235, 236 (2011).

[164] *Id.*

[165] S*ee* Citron, *supra* note 143, at 1253-54.

[166] Danielle K. Citron & Frank Pasquale, *The Scored Society: Due Process for Automatic Predictions*, 89 WASH. L. REV. 1, 19 (2014).

[167] Perel & Elkin-Koren, *supra* note 101, at 506.

[168] Urban, Karaganis & Schofield, *supra* note 112, at 44.

[169] *Id.*

[170] *See* Perel & Elkin-Koren, *supra* note 101, at 506.

users.[171] While the copyright holders are granted protection under a legal provision, the users are safeguarded merely through a possible self-regulatory procedure that online platforms should collaborate to set up.[172]

A second corollary of opacity lies in the general lack of public oversight on enforcement that is privately carried out.[173] As a matter of fact, there is no way to oversee what content—and on what grounds—is allowed on platforms and under what conditions.[174] Even when platforms publish transparency reports, it is simply impossible to monitor intermediaries' activities.[175] Public oversight is further impeded by the claim that transparency cannot be challenged as algorithms are valuable trade secrets.[176] In particular, the ex ante enforcement that algorithms enable limits the possibility to correct errors and this in turn significantly curbs the possibility for the public to intervene and have a voice.[177]

A further cluster of concerns relates to the establishment of technological normativity, i.e. the embodiment of norms in automated systems, devices and agents.[178] In particular, when technologies bear a "constitutive normativity," they determine users' actions as well as the way in which they can operate and act.[179] In this case, technology can end up not only conditioning individuals' behaviors but also altering the essence of the law as such.[180]

The impact of technological normativity on the law is evident in the case of copyright enforcement online. It is because of technological normativity that both the internal and external balances, typical of copyright law, are modified.[181] All jurisdiction encompass internal mechanisms that limit the scope of copyright, among which are exceptions and limitations in civil law countries and fair use or fair dealing in common law countries.[182] Similarly, on an external front, copyright protection finds boundaries in other branches of the law, such as fundamental rights and freedoms, particularly the freedom of expression and the right to privacy, but also competition law through which

---

[171] *See Id.*
[172] *See Id.*
[173] Elkin-Koren & Perel, *supra* note 147, at 13.
[174] *Id.*
[175] *Id.*
[176] *Id.*
[177] *Id.*
[178] Marcello D'Agostino & Massimo Durante, *Introduction: the Governance of Algorithms*, 31 PHIL. & TECH. 499, 501 (2018).
[179] *See* Mireille Hildebrandt, A VISION OF AMBIENT LAW 177-78 (Roger Brownsword & Karen Yeung eds., 2008), for the definition of the effects of technological normativity.
[180] *Id.*
[181] *See* Michal Birnhack, *Judicial Snapshots and Fair Use Theory*, 5 QUEEN MARY J. OF INTELL. PROP. 264, 268 (2015).
[182] *Id.* at 275.

compulsory licenses are usually imposed.[183] Internal and external mechanisms should not be understood as two separate spheres.[184] The need to resort to external balance arises when the internal safeguards do not adequately cater for instances other than granting protection to authors and right holders.[185] However, it has been stressed that this "inside/outside location metaphor" can be seen as a distinction between the types of considerations required to resolve a given dispute.[186] When internal safeguards are applicable, the conflict is between instances grounded within copyright law. When, on the other hand, the conflict is rooted in different grounds – such as copyright vis-a-vis freedom of expression—there is the need for balancing between different fundamental rights.[187]

Now, once copyright enforcement becomes algorithmic, the decision as to whether a content falls within a permitted use or represents an instance of freedom of expression is delegated to the autonomous system in place.[188] A balancing operation that is usually carried out by humans is adopted by an algorithm employing various degrees of autonomy, depending on the way in which it has been programmed.[189] Algorithms translate legal mandates into code. This implies that the interpretation of the law may, and most often is, affected by a variety of extrajudicial considerations, including the conscious and subconscious professional assumptions of program developers, as well as various private business incentives.[190] As a matter of fact, many are the cases in which algorithmic decision-making has resulted in a lack of consideration of permitted uses or fundamental freedoms.[191] These cases span from the famous *Lenz* case in the US in 2007, which saw YouTube taking down the "dancing baby video" featuring a child dancing in the family's kitchen to Prince's song "Let's Go Crazy by Prince,"[192] to the several cases in which song writers

---

[183] Bernt P. Hugenholtz & Ruth L. Okediji, *Conceiving an International Instrument on Limitations and Exceptions to Copyright*, Amsterdam Law Sch. Legal Studies Research Paper No. 2012-43, 1, 11–12 (2008), https://ssrn.com/abstract=2017629.

[184] Abraham Drassinower, *Exceptions Properly So-Called*, LANGUAGE AND COPYRIGHT 205, 230 (2009).

[185] *Id.*

[186] *Id.*

[187] *Id.* at 222 (defining exceptions properly so-called as the nexus of an encounter between copyright and other juridical interests).

[188] Perel & Elkin-Koren, *supra* note 101, at 517.

[189] *Id.*

[190] *Id.* at 517-18.

[191] *See* Sag, *supra* note 110, at 528.

[192] *Id.* at 527-28.

complained that the Content ID incorrectly identified their works as infringing.[193]

Over the years, the story of algorithmic copyright enforcement shows how the employment of increasingly autonomous systems has tilted the balance in copyright law, namely, "if copyrighted materials were once available unless proven to be infringing, today materials that are detected by algorithms are removed from public circulation unless explicitly authorized by the right holder".[194] To be on the safe side, online intermediaries implement algorithms that remove or impede the upload of allegedly infringing content. Therefore, the question arises as to whether it is really impossible for algorithms to consider permitted uses or it is more that intermediaries to be safe, employ algorithms that disregard permitted uses. The concern raised by the effects of technological normativity is thus twofold. On the one hand, it questions the way in which technology is used to comply with the law and, on the other hand, it discusses the aptness of technology to efficiently implement the law.

B. Proposals to Overcome the Shortcomings of Algorithm Enforcement

The above-mentioned concerns have already been raised by several scholars who have also proposed solutions to overcome them. Measures to cope with algorithmic enforcement in general – and algorithmic copyright enforcement in particular – aim at understanding the algorithmic decision-making process, providing sufficient opportunities to challenge an algorithmic decision and correcting erroneous or improper outcomes regarding online content. A fundamental prerequisite to achieve the above is having transparent, explicable and accountable algorithms. Perel and Elkin-Koren propose an intervention at several levels in order to enhance accountability—from individual users and the public's general involvement to an official regulatory activity.[195] In particular, with regard to the latter, standards of algorithm disclosure as well as reporting obligations should be in place.[196] Transparency would surely be increased if intermediaries are required to disclose the criteria their algorithms consider when determining copyright infringement. This should include quantitative thresholds (*i.e.*, what percentage of the copyrighted work must be used to cause content restriction) and fair use policies.[197] However, these same authors,[198] as well as other scholars in the field, caution about the drawback of transparency as this is just a means to accountability and

---

[193] Taylor B. Bartholomew, *The Death of Fair Use in Cyberspace: YouTube and the Problem with Content ID*, 13 DUKE L. & TECH. REV. 66, 69–70 (2015).

[194] Niva Elkin-Koren, *Fair Use by Design*, 64 UCLA L. REV. 1084, 1093 (2017).

[195] Elkin-Koren & Perel, *supra* note 147.

[196] *Id.* at 530-531.

[197] *Id.* at 530.

[198] Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 188 (2017).

not an end in itself.[199] Unveiling the algorithm's code is a way to make those employing them more accountable, but it does not constitute accountability in itself. An algorithmic process is in fact truly accountable when its stakeholders can intervene to change the algorithm, or its implementation. In other words, transparency is a fundamental step towards accountability when it enables due process and public oversight.

As far as due process is concerned, Citron has suggested the introduction of a "technological due process," by which accountability would be enhanced through the establishment of procedures ensuring that predictive algorithms live up to some standard of review and revision.[200] Consequently, this would guarantee its fairness and accuracy.[201] In particular, the author suggests the adoption of automated systems to generate audit trails that record the facts and rules supporting their decisions, with the aim of providing a comprehensive history of the decisions made in a case.[202] The author also proposes that the source codes of such automated systems are released to the public as this would prevent inadvertent and procedurally defective rulemaking.[203] Opening up the source code would expose how a system works. Moreover, before the release for mass use, such systems should be tested, and texting should continue to run even during their implementation, as well as take place every time policies change.[204] Finally, the ways to allow the public to participate in the building of automated decision systems should be envisaged when such systems are adopted by administrative agencies.[205] The same principles are adopted and further developed also by Crawford and Schultz, who look for a fair and reasonable procedure in the case of privacy harm.[206] The authors suggest the introduction of an external regulator or audit body which might investigate complaints and provide mediation or adjudication.[207]

Similarly, as transparency itself is not enough to enable public oversight, Perel and Elkin-Koren advocate that users need to resort to self-help as a more efficient means of ensuring accountability of algorithms.[208] Using algorithmic copyright enforcement as a case study, the authors demonstrate how "tinkering"

---

[199] Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH.103 (2018).
[200] Citron, *supra* note 143.
[201] *Id.*
[202] *Id.* at 1305.
[203] *Id.* at 1308.
[204] *Id.* at 1310.
[205] *Id.* at 1312.
[206] Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 BOSTON COLL. L. REV. 93 (2014).
[207] *Id.* at 49.
[208] Elkin-Koren & Perel, *supra* note 198, at 181.

can provide a valuable research methodology to explore the hidden practices of algorithms. Therefore, legal intervention should be called for in favor of "tinkering" activities. For example, a statutory immunity could be considered for researchers and users tinkering a safe harbor to understand its functioning.[209] In other words, "tinkering" is formulated as the "freedom to understand, discuss, repair, and modify the technological devices you own,"[210] which, in the case of algorithms, morphs into "an important tool to proactively check the credibility, fairness, and trustworthiness of algorithms that cannot be adequately reviewed through traditional means of transparency."[211] In other words, tinkering encourages public oversight as it enables the public to exercise judgement and demand that algorithmic enforcement complies with the public interest.

Moving to the limits of technological normativity, a way of overcoming the imbalances that it may generate is that of designing the technology in a way that encompasses and balances all the interests at stakes. For example, in the case of copyright, Elkin-Koren proposes a fair use by design, which could be achieved by "translating fair use considerations into a set of instructions that can be executed on certain data sources, and writing them in programing language that is readable by computers."[212] This could be made more efficient by the latest development in AI.[213] However, given the challenges of applying AI and machine learning to detect fair use, the author suggests that a new approach to fair use is needed—an approach that encompasses a shift from a case-by-case analysis to a legal scrutiny of the AI system as such, and of its procedures, measures and error rate.[214]

On the other hand, Dan Burk highlights the risks of algorithmic fair use.[215] As fair use is an ex ante indeterminate legal standard translating its values in code, the embedding of these same values in public behavior and consciousness is not achievable. As a consequence, algorithmic fair use would carry the very real possibility of adapting new media participants to its own biases. Eventually, the original balanced fair use as envisaged by the legislature would be progressively altered, and the public would no longer benefit from the fair use as originally conceived.[216]

---

[209] *Id.* at 217.

[210] Edward Felten, *The New Freedom to Tinker Movement* (2013), https://freedom-to-tinker.com/blog/felten/the-new-freedom-to-tinker-movement/.

[211] Elkin-Koren & Perel, *supra* note 198, at 199-200.

[212] Elkin-Koren, *supra* note 194, at 1095.

[213] *Id.* at 1096-1097.

[214] *Id.* at 1099.

[215] Dan L. Burk, *Algorithmic Fair Use*, 86 U. CHI. L. REV. 283 (2019).

[216] *Id.* at 285.

Now, it is incontestable that a new framework for addressing algorithmic governance must be developed, possibly by "rethinking the role of courts and of judicial oversight"[217] or by establishing a sound governance system of autonomous systems. At the same time, taking fair use as an example, the social costs of an algorithm implementing an ex ante determination of the law should be of concern and be addressed in such a framework itself.

VI. TOWARDS A BALANCED ALGORITHMIC COPYRIGHT ENFORCEMENT

The debate on algorithmic copyright enforcement is extensive, in particular when framed within the algorithmic accountability discussion.[218] Similarly vast is the literature on the solutions to make algorithms more accountable, fair and transparent.[219] Nonetheless, these do not yet amount to wide-ranging practices. Therefore, in the following sections I formulate a proposal for a more balanced algorithmic copyright enforcement. To avoid reinventing the wheel, I draw from other fields of law, such as data protection and environmental law, adapt some of the solutions that have already been offered and fill the gaps that need to be filled in the pursuit of comprehensiveness and consistency.

My proposal consists in the adoption of a principle for a balanced algorithmic copyright enforcement, which translates into the need for open record policies and a right to explanation. This in turns is coupled with the obligation of a right-based impact assessment and a right to audit. While devised with algorithmic copyright enforcement in mind, this regulatory toolkit can hopefully provide insight for a more balanced algorithmic society overall.

A. The Principle of a Balanced Algorithmic Copyright Enforcement

The majority of literature on algorithmic copyright enforcement focuses on how negatively they can impact individuals and envisages tools to enhance algorithms accountability,[220] which, in many cases concern the enhancement of transparency.[221] However, the essential question is not just how to make algorithms transparent and accountable to affected individuals. It pertains to the broader issue of how to encourage private companies implementing algorithms to develop norms, policies, and technical tools that will make algorithmic

---

[217] Elkin-Koren, *supra* note 194, at 1100.

[218] *See* Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC. 973 (2016); Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1 (2017).

[219] *See generally* Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. Online 189 (2017); *see also* Citron, *supra* note 143.

[220] *Id*.

[221] Ananny & Crawford, *supra* note 218.

enforcement and decision making in general more balanced—that is, capable of contemplating and overseeing the interests of all parties involved.

This ambitious goal requires placing algorithmic decision-making within a sound legal framework, which clearly establishes the necessity that technology is not used to comply solely with one set of rules such as copyright law. Instead, technology should take into consideration the interests and values present in the overall system holistically.

Generally speaking, one may take this principle for granted. Yet, as algorithmic copyright enforcement in the DSM market demonstrates, this is certainly not the case. Article 17 of the DSM Directive clearly obliges online content-sharing service providers to offer protection to copyright works and, to this end, imposes the twofold obligation of licensing or filtering. Provided that a platform complies with either of the two limbs, it can escape liability for infringement of the communication right.[222] The same provision also calls for safeguards to be introduced in order to avoid limiting third parties' permitted uses under the copyright exceptions and limitation regime. In particular, Article 17(7) provides that "[t]he cooperation between online content-sharing service providers and right holders shall not result in the prevention of the availability of works or other subject matter uploaded by users, which do not infringe copyright and related rights, including where such works or other subject matter are covered by an exception or limitation." Following the same lines, Recital 70 recommends that "[u]sers are allowed to upload and make available content generated by users for the specific purposes of quotation, criticism, review, caricature, parody or pastiche" as these are "particularly important for the purposes of striking a balance between the fundamental rights laid down in the Charter of Fundamental Rights of the European Union, in particular the freedom of expression and the freedom of the arts, and the right to property, including intellectual property rights. Moreover, an effective complaint and redress mechanism to support use for such specific purposes is equally important.

The DSM Directive calls for safeguards to be introduced with the aim of making copyright enforcement balanced. Yet, the precise specification of these safety measures is left to the discretion of online intermediaries or rather to the collaboration between them and right holders. This generates a clear imbalance between the protection of copyright works, detailed at a legislative level, and protection of other interests such as freedom of information or permitted uses, that are left at the level of self-regulation, namely upon discretion of those same intermediaries deploying algorithms to avoid being liable. Such a distortion is the first obstacle in achieving a balanced algorithmic copyright enforcement. It needs to be offset by the adoption of a clear principle

[222] *See supra* Section III C.

that details the steps of the process that intermediaries would adopt in order to accomplish a balanced algorithmic copyright enforcement.

In fact, the principle of a balanced copyright enforcement is necessary but alone is not enough. There is a further need for a more intense regulatory involvement in oversight and accountability as far as technological measures of compliance are concerned.[223] As these choices cannot be left in the hands of those who are at the same time forced to enforce copyright protection, the following sections put forward a set of measures aiming at limiting the discretion of online intermediaries in enforcing copyright law and rendering them accountable for their operations. The ultimate goal of this regulatory toolkit is that of achieving a form of "balance by design".

B. Algorithmic Explainability: From Open Record Policies to a Right to Explanation

A balanced algorithmic copyright enforcement requires algorithmic explainability. This is the possibility of understanding the question answered by the algorithm and the manner in which it arrives at it. Explainability can be achieved by demanding firms employing autonomous systems to adopt open record policies and by granting a right to explanation to individuals that are affected by automated decision-making.

Broadly speaking, an open record policy entails that online intermediaries reveal the relevant structures, logic, and policies underlining the adopted algorithms. Proposed by Brauneis and Goodman as an empirical project for better governance of smart cities, the open record policy could be extended to private firms involved in algorithmic enforcement.[224] In this way, one could establish a record generation practice creating information on the design, the procurement, and the implementation of algorithmic processes.[225]

The adoption of an open record policy has the advantage of overcoming the transparency problems, in particular when it is achieved by disclosing the proprietary code.[226] Yet, given the complexity of such algorithms, in the

---

[223] Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 729 (2010).

[224] Brauneis and Goodman, *supra* note 199, at 136.

[225] *Id.* at 175-176; s*ee also* Tal Z. Zarsky, *Transparent Predictions*, 4 U. TEX. L. REV. 1504, 1521 (2013) (proposing a "call for mapping transparency" based on (1) the collection of data and aggregation of datasets, (2) data analysis, and (3) actual strategies and practices for using the predictive models, effectiveness of which could be measured by both the way they are applied ex ante and their final impact ex post).

[226] Perel & Elkin-Koren, *supra* note 101, at 509.

majority of cases a disclosure obligation would not really enable the understanding of how the algorithm indeed works. Instead of resorting to the code disclosure, an algorithm can still be rendered transparent enough by revealing some combination of mathematical and logical notation, natural language and information regarding the way in which data is selected, including the rules of operation chosen and the steps taken to validate those choices.[227] Beside the model design choices, the data selection, the factor weighting and validation elements, the records should also indicate the questions to which the algorithm provides an answer. This is crucial in the assessment of the algorithm's effectiveness, fairness, and political acceptability.[228]

Essentially, an open policy record is a key element in the process of making algorithms explainable. This becomes even more evident when coupled with the legibility test developed by Malgieri and Commande.[229] According to the authors, the algorithm's inherent functionality and its strict logic should be kept distinct from its contextual use, consequences, and impact.[230] Indeed, it is from the understanding of both the technical architecture of an algorithm (the functionality and the logic involved) and its contextual implementation (the significance of a decision-making and its envisaged consequences) that its legibility derives and in turn generates algorithmic explainability.[231]

An open record policy provides an ex ante means of making algorithmic copyright enforcement more balanced. This needs to be complemented with an ex post tool in the hand of the individuals affected by the algorithmic decision-making. Such an ex post tool should encompass a set of individual rights as already done in other branches of law such as data protection, namely, the right to obtain explanation as to the algorithm affecting the individual in question. For example, Article 22 of the GDPR introduces a right to receive "meaningful information about" the logic, significance and consequences of algorithmic data processing, when a decision is the result of a completely automatic process.[232] Now, there is an extensive debate as to what right, if any, Article 22 introduces. In principle, the article merely states that a "data subject shall have the right not to be subject to a decision based solely on automated processing … which produces legal effects concerning him or her or similarly significantly affects

---

[227] Brauneis and Goodman, *supra* note 199, at 130.

[228] *Id.* at 131.

[229] Gianclaudio Malgieri & Giovanni Comande, *Why a right to legibility of automated decision-making exists in the general data protection regulation*, 7 INT'L DATA PRIVACY L. 243, 259-61 (2017).

[230] *Id.* at 258.

[231] *Id.* at 258-9.

[232] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 679) ('GDPR').

him or her."[233] But, recital 71 of the GDPR clearly refers to "the right to obtain … an explanation of the decision reached" through the processing of personal data.[234] While many commentators, adopt a broad reading of Article 22 and conceptualize it as a right to understand the reason,[235] others read it more literally as the right to exercise the decision to not be subject to any autonomous decision-making.[236]

The reality is that the tool provided by Article 22 is very narrow,[237] as it amounts to a mere right to be informed.[238] Thus, it is unlikely that it would be of any use in tackling the imbalances of algorithmic copyright enforcement. To put this in context, consider the case of an Internet user who seeks to understand the reasons of an automatic take down or the automatic prevention of one of hers uploads. Several requirements need to be met so that the user can actually resort to Article 22. In the first place, there must be some processing of her personal data such as the IP address. Then, the take-down or the impossibility of uploading should be the result of a decision *solely* based on automated processing. This in itself entails three further elements: first, there must be a decision; second, the decision must result from automated processing; and, third, there must be no human intervention in the process. Eventually, the decision must produce legal effects concerning the user, or significantly affect her by, for example, limiting her freedom of expression. If all of these requirements are present, a user can rely on Article 22 provided that none of the exceptions envisaged by the GDPR apply. In this respect, the user needs not to have explicitly agreed to the automatic decision-making;[239] such automatic decision-making needs not to be necessary for entering into or performing a contract between the user and the platform;[240] nor needs the automated decision-making to have been authorized by a Member State national law.[241] Moreover, since a user likely consents to automatic decision-making when adhering to the

---

[233] *Id*. at Art. 22(1).

[234] *See* GDPR, at Recital (71).

[235] *See generally* Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L. J. 189, 217 (2019); Andrew D. Selbst & Julia Powles, *Meaningful Explanation and the Right to Information*, 7 INT'L DATA PRIVACY L. 233 (2017); Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-making and a "Right to Explanation"*, 38 AI MAG 50, 55 (2017).

[236] *See* Sandra Wachter, Luciano Floridi & Brent Mittelstadt, *Why a Right to Explanation of Automated Decision-making does not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76 (2017).

[237] *See* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a "Right to an Explanation" Is Probably not the Remedy You are Looking for*, 16 DUKE L. & TECH. R. 18, 44-5 (2017).

[238] Wachter et al., *supra* note 236, at 89-90.

[239] *See* GDPA, *supra* note 232, at Art. 22(2)(c).

[240] *Id* Art. 22(2)(a).

[241] *Id.* Art. 22(2)(b).

platform's terms of use, it results that Article 22 cannot provide a tool for transparency in the case where algorithmic copyright enforcement leads to an erroneous—or at least debatable—takedown of a user's uploaded content.

Despite the narrow applicability of Article 22, the overall solution championed by the GDPR as to the set of individual rights conferred to the data subjects may be suitable to introduce a right to explanation.[242] Article 22, when coupled with Articles 13(2)(f), 14(2)(g) and 15(1)(h), may offer a valuable model for a more balanced algorithmic enforcement. In fact, considered together, these provisions constitute a right to be made aware of the existence of automated decision-making, which includes profiling pursuant to Article 22(1) and (4). Furthermore, what is envisaged is the right to receive meaningful information about the logic involved in such processing, as well as its significance and its consequences for the data subject. In particular, Articles 13 and 14 introduce "notification duties," whereby information must be provided when the processing of personal data begins. Moreover, Article 15 stipulates that one has the right to access all those pieces of information at any moment. Consequently, since the information to be provided as part of the notification duty concerns "the existence of automated decision-making, including profiling and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject", it enables the understanding of both the architecture and the implementation of the algorithmic decision-making.[243]

Interestingly, the interpretation of "meaningful information" offered by the Article 29 Data Protection Working Party ('A29WP') in the *Guidelines on Automated Individual Decision-Making and Profiling for the GDPR* closely resembles what an open policy record would entail.[244] According to the A29WP's opinion, meaningful information concerns the logic involved. In most cases this "logic" will require controllers to provide details such as the information used in the automated decision-making process, including the categories of data used in a profile and the source of that information, the manner in which any profile used in the automated decision-making process is built—including any statistics employed in the analysis— the reason a profile is considered relevant in the automated decision-making process, and how it is used for a decision concerning the data subject.[245] Even though Article 22

---

[242] *See* Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The Gdpr's "Right To Explanation" Debate And The Rise Of Algorithmic Audits In Enterprise*, 34 BERKELEY TECH. L. J. 144, 153, 167 (2019).

[243] Malgieri & Comande, *supra* note 229, at 255.

[244] ARTICLE 29 DATA PROTECTION WORKING PARTY, Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679 (2017), file:///C:/Users/mlm26/Downloads/wp251rev01_enpdf%20(1).pdf

[245] *Id.* at 25.

applicability is limited to algorithmic decision-making that processes personal data, it offers a model to consider when aiming for algorithmic explainability.

In sum, a more balanced algorithmic enforcement would require both ex ante and ex post regulatory tools, *i.e.* generating records depicting how the algorithm is designed, how it works, and what it is intended for, as well as furthering the conferring of a right to explanation on users.

C. A Right-Based Impact Assessment and a Right to Audit: Safeguarding the Safeguards

In practice, explainability is not limited to revelations to the public or the individual user upon its request. Instead, it includes putting in place internal company procedures of oversight, combined with the right to react when affected individuals raise concerns. From an operational point of view, an open-record policy and a right to explainability should correspond to an obligation to perform an impact assessment, which in turn would enable auditing by affected parties.

Impact assessments are certainly not novel—they are often used to understand the risks generated by activities and identify ways of managing them. In fact, impact assessments are widely used in environmental law,[247] and are becoming more and more popular in data protection regulation.[248]

The GDPR, for example, specifically requires that when processing is done through new technologies and is likely to result in high risks for data subjects, a prior Data Protection Impact Assessment (DPIA) must be carried out.[249] This requirement stems from the notion of privacy by design and aims at building privacy-aware or privacy-friendly systems, starting from the beginning of the process of technology design rather than tackling privacy issues at the end of it.[250] Within the same lines, the A29WP describes the DPIA as a process for building and demonstrating compliance by systematically examining automated processing techniques to determine the measures necessary to

---

[247] *See generally* Tseming Yang, *The Emergence of the Environmental Impact Assessment Duty as a Global Legal Norm and General Principle of Law*, 70 HASTINGS L. J. 525 (2019).

[248] *See generally* A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements*, 2015 U. ILL. L. REV. 1713 (2015) (discussing the problem with mass surveillance and comparing the need for impact statements of privacy violations with that of environmental impact statements); Alessandro Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, 34 COMPUTER L. AND SECURITY R. 754, 772 (2018) (proposing a Human Rights, Ethical and Social Impact Assessment for data processing).

[249] GDPR, *supra* note 232, at Art. 35.

[250] *See* Edwards & Veale, *supra* note 237, at 59.

"manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data."[251]

More specifically, the GDPR proposes impact assessments primarily centered on data quality and data security.[252] This derives from the procedural approach focused on data management risk assessment, which characterizes the development of data protection in Europe. The core principle of EU data protection is that of providing data subjects with control over their own information. As this information amounts to data of various nature, processed for various ends, the only viable approach is a procedural one. In this way one can consistently secure all the stages of data processing, from data collection to communication of data to third parties.[253]

However, considering that a procedural and risk management approach leaves aside the actual nature of the safeguarded interests, it may not be the most suitable form of impact assessment in the case of algorithmic copyright enforcement. Copyright law entails a specific set of rights, freedoms and values that should be considered beyond—or at least beside—the technology deployed to enforce. A copyright impact assessment needs to adopt a more right-oriented approach and focus also on the individual and societal impact of the technologies. The nature of the technology has direct relevance in the assessment process as it determines the most appropriate measures to take when safeguarding the rights and values impacted. A right-based impact assessment thus encompasses the potential negative outcomes on a variety of fundamental rights and principles while also taking into account the ethical and social consequences of algorithmically enforced copyright law.

A right-based impact assessment ought to be carried out during each phase of the algorithm development and deployment. This will achieve ex ante and ex post evaluations of the ways in which third parties and society overall are affected.[254] During the design and development stage, an ex ante impact assessment enables the evaluation of how the algorithm at stake works, it ensures that it functions as intended and identifies problematic processes or assumptions. This analysis provides an opportunity to modify the algorithm design at an early stage, to build in rights compliance, to include monitoring mechanisms from the beginning, or to stop the development of an algorithm whenever right-related concerns cannot be addressed. An ex post impact

---

[251] *See* ARTICLE 29 DATA PROTECTION WORKING PARTY, Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing is "Likely to Result in a High Risk" for the Purposes of Regulation (2016) 679 final (Oct 4, 2017), https://iapp.org/media/pdf/resource_center/WP29-GDPR-DPIA-guidance_final.pdf.

[252] Mantelero, *supra* note 248, at 757.

[253] *Id*. at 767.

[254] Lorna McGregor, Daragh Murray & Vivian Ng, *International Human Rights Law as a Framework for Algorithmic Accountability*, 68 ICLQ 309, 330 (2019).

assessment, conducted during the deployment stage, enables monitoring an algorithm's effects during operation. This entails observing if the oversight and safeguards set during the design and development phase are indeed able to identify and respond to rights violations once the algorithm is deployed. This ability to react to violations is key, as any effective impact assessment requires that problematic processes must be capable of being reconsidered, revised or adjusted.[255]

A balanced algorithmic copyright enforcement can be achieved by placing a tool in the hands of affected users that allows them to challenge the algorithmic decision-making and address violations of their rights. Such a tool works as a complement to the right-based impact assessment and could also be at the disposal of researchers and authorities to test the processes deployed and exercise oversight. While a right to explanation enables affected individuals to comprehend whether the law is algorithmically enforced in a balanced manner, it does not provide a tool to react to a lack of balance and to challenge it.[256] Therefore, beyond the right to explainability—which still plays an important role—an effective regulatory toolkit should also consider the adoption of a right to have the algorithm audited. This, again, was contemplated by the A29WP Guidelines in relation to data processing, where, in the context of algorithmic decision-making, the authority envisioned substantial third-party oversight by suggesting the implementation of third-party audits.[257] Indeed, Recital 71 of the GDPR could be interpreted in the sense of required auditing.[258] Though, to avoid the misuse of an auditing request, the specifics of it should be devised in such a way to not constitute an excessive burden on companies developing and deploying algorithms. For example, it could be envisaged as a consumer association or as a collective right that can be acted upon when a substantial number of individuals claim to have been affected.For example, a recent proposal for a directive for the protection of the collective interests of consumers provides consumers with a redress mechanism by strengthening their access to justice.[260] At the same time, these provisions strive to strike the necessary balance between access to justice and procedural safeguards against

---

[255] *Id.*

[256] *See generally* Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, U. of Colorado Law Legal Studies Research Paper No. 19-9 (2019) (discussing the need for a collaborative approach to regulation between developers and researchers); Edwards & Veale, *supra* note 237, at 44 (discussing the shortcomings of existing algorithmic oversight tools).

[257] *See* Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679 *supra* note 244, at 32.

[258] Malgieri & Comande, *supra* note 229, at 9 and 24 (arguing that Articles 13(2)(f), 14(2)(g) and 15(1)(h) introduce a duty to perform an *auditing* of decision-making algorithms).

[260] *See* European Commission, *Proposal for a Directive of the European Parliament and of the Council on Representative Actions for the Protection of the Collective Interests of Consumers, and Repealing Directive 2009/22/EC*, COM(2018) 184 final.

abusive litigation which could unjustifiably hinder the ability of businesses to operate in the Single Market.[261]

In sum, from an operational point of view, online platforms algorithmically enforcing copyright law should follow procedures to ensure that this occurs in a balanced manner. Beyond designing algorithms in a way that they are balanced by design and implement the safeguards needed to contemplate all the interests at stake, companies should also adopt procedures, such as ex ante and ex post right-based impact assessments and auditing, which when requested by groups of affected individuals would provide oversight in algorithmic copyright enforcement.

VII. CONCLUSION

Starting from the proposition that the algorithmic society is here to stay, this article formulates a proposal for a balanced algorithmic copyright enforcement. The current development of autonomous systems and self-learning algorithms indicates that technology will continue to provide an increasingly sophisticated tool of compliance. In such a setting it becomes crucial to govern adequately the development that we are witnessing.

To this end, having surveyed the European policies on tackling illegal content online within the DSM Strategy,[262] I provided evidence of a strong trend towards algorithmic enforcement in the various provisions that have been adopted. In addition, I also discussed how this algorithmic enforcement is in turn morphing into algorithmic content regulation in European law.[263]

The case of algorithmic copyright enforcement demonstrates how technology protects copyright law online and its evolution over the years—from automated instruments removing content signaled as infringing by rightholders, to autonomous systems filtering allegedly infringing content and preventing its reappearance online. These activities initially took place within the safe harbor regime which would shield online intermediaries from liability for third parties' infringing content; however, under the recent DSM Directive a new phase has entered. Online intermediaries have become directly liable for third parties' infringing content unless they adopt systems that prevent the infringement from taking place. In other words, they are at risk of liability unless they algorithmically enforce copyright law through autonomous systems capable of detecting and preventing the upload of infringing content. This shift from automation to autonomy has been paralleled with a shift from ex post algorithmic copyright enforcement—as per the robo-notice and take down – to ex ante algorithmic copyright enforcement—as per the autonomous filtering

[261] *Id.* at 19

[262] *See* discussion *supra* Section I.

[263] *See* discussion *supra* Section II.

systems. The aim of avoiding infringement instead of remedying it is what brings about a regime of algorithmic content regulation.[264]

This trend towards algorithmically enforcing legal rules and content regulation raises wide concerns, which this work has summarized with a focus on algorithmic copyright enforcement. While several efforts have already been made to try and overcome the algorithmic copyright enforcement shortcomings,[265] a coordinated and balanced action is still absent.

Besides the approaches proposed in literature, European legal framework promotes several hard and soft legal instruments that can be used to achieve a more balanced algorithmic copyright enforcement. The various principles, recommendations, suggestions and tools already adopted are however scattered in numerous policies and actions introduced by the European institutions over the years.

In this article, I contribute to the existing debate on algorithmic copyright enforcement by proposing a regulatory toolkit that could also provide insights for a better algorithmic society overall.[266] The starting point is the principle for a more balanced algorithmic copyright enforcement regime which translates into the need for open record policies and a right to explanation. This is coupled with the obligation of a right-based impact assessment and a right to audit.

The regulatory toolkit proposed herein represents a first step towards a more coordinated approach to algorithmic copyright enforcement. It introduces the discussion on "balance by design", i.e. the engineering of online platforms from their beginning in a way that takes into consideration all the rights involved. However, while balance by design can be a very attractive concept as it strives to achieve a more balanced algorithmic copyright enforcement *ab initio*, it needs to be carefully inserted within a sound legal framework[267] and complemented with a series of further instruments.[268]

Despite the need to govern algorithms through both specific measures and a sound legal framework, some of the broader concerns raised by algorithmic enforcement and regulation may prove hard to eliminate. Specifically, algorithmic enforcement and regulation lead to legal

[264] *See* discussion *supra* Section III.
[265] *See* discussion *supra* Section IV.A.
[266] *See* discussion *supra* Section V
[267] *See* discussion *supra* Section V.A.
[268] *See* discussion *supra* Section V.B and C.

automation.[269] Whatever semi-optimal system we may achieve, legal automation runs the risk of jeopardizing social reflexivity and collective and individual autonomy. In other terms, there is the imminent danger of unintended consequences on the way in which we conceive the law and interact with it.

First, social reflexivity (*i.e.*, the public scrutiny, discussion and evaluation of norms) is a crucial factor for an accurate understanding and application of the law.[270] Legal automation, being a largely impersonal process, is less capable of taking into account context-specific factors. It thereby risks altering the relationship between law and facts and undermining the exact circumstances of the case. In other words, legal automation is likely to eliminate the distinction between rules and standards. Rules establish basic instructions for behavior and require consistent treatment of similar cases, eliminating the need to reconsider recurring issues. Standards, on the other hand, permit decision-makers to tailor an outcome to the facts. For example, changing circumstances brought about by technology can be taken into account.[271] At least for the time being, automated systems are primarily apt to apply rules rather than standards. This is due to the way in which an algorithm works to predetermine an outcome for a set of facts. This promotes a decontextualized assessment over context-dependent decision making. Consequently, the complexity of the legal system is reduced and tends to isolate and set in stone notions that are instead subject to evolution.

Second, legal automation may also impair collective and individual autonomy by modeling social conduct by design.[272] The non-compliance with a set of prescriptions triggers certain effects. However, when enforcement becomes algorithmic—and, specifically, when the algorithm works ex ante, preventing the violation from taking place as in the case of algorithmic copyright enforcement—the law is no longer just a set of rules, it becomes a set of effects that apply regardless of whether an infringement has occurred.

As a result, the discussion on the modelling of a truly balanced algorithmic enforcement must go hand in hand with the conversation on its

---

[269] *See generally* Ugo Pagallo & Massimo Durante, *The Pros and Cons of Legal Automation and its Governance*, 7 EUROPEAN JOURNAL OF RISK REGULATION 323 (2016) (discussing the notion of legal automation).

[270] Marcello D'Agostino & Massimo Durante, *Introduction: The Governance of Algorithms*, 31 PHILOSOPHY & TECHNOLOGY 499, 501 (2018).

[271] Citron, *supra* note 143, at 1303

[272] Ugo Pagallo, *Cracking Down on Autonomy: Three Challenges to Design in IT Law*, 14 ETHICS AND INF. TECH. 319, 321 (2012).

unintended consequences. This should encompass both the way in which the law is conceived as well as how law and society interact one with one another.