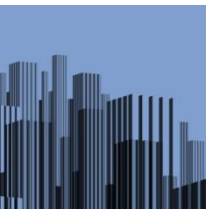
Journal of Liberty and International Affairs | Volume 7 · Number 3 · Supp. 1 · 2021 | eISSN 1857-9760
Published online by the Institute for Research and European Studies at www.e-jlia.com



Peer review method: Double-Blind
Accepted: September 16, 2021
Published: November 23, 2021
Original scientific article
DOI: https://www.doi.org/10.47305/JLIA2137136d

ALGORITHMS AND FUNDAMENTAL RIGHTS: THE CASE OF AUTOMATED ONLINE FILTERS

Matija Damjan
University of Ljubljana, Faculty of Law & Institute for Comparative Law, Slovenia
ORCID iD: https://orcid.org/0000-0001-6063-0328
matija.damjan@pf.uni-lj.si

Abstract: *The information that we see on the internet is increasingly tailored by automated ranking and filtering algorithms used by online platforms, which significantly interfere with the exercise of fundamental rights online, particularly the freedom of expression and information. The EU's regulation of the internet prohibits general monitoring obligations. The paper first analyses the CJEU's case law which has long resisted attempts to require internet intermediaries to use automated software filters to remove infringing user uploads. This is followed by an analysis of article 17 of the Directive on Copyright in the Digital Single Market, which effectively requires online platforms to use automated filtering to ensure the unavailability of unauthorized copyrighted content. The Commission's guidance and the AG's opinion in the annulment action are discussed. The conclusion is that the regulation of the filtering algorithms themselves will be necessary to prevent private censorship and protect fundamental rights online.*

Keywords: *Algorithms; Content Recognition; Upload Filters; Censorship; Human Rights; Intermediary Liability; AI*

INTRODUCTION: THE ROLE OF ALGORITHMS ONLINE

Algorithmic software tools increasingly tailor our online experience and thus shape our view of the world. Google's search algorithm was the key to its quick rise and eventual dominance over other search engines as it allowed its users to find the most relevant results on the web in a fraction of a second. If a piece of information published online is not indexed by Google's bots or has a low ranking in the presentation of Google's search results it will be effectively invisible to a vast majority of internet users. Algorithmic ranking and recommender systems play an essential role in social networks where they determine which posts will be displayed in a user's news feed, usually based on the user's interests and previous interactions (Llansó 2020, 1). All done to attract the user's attention, encourage the sharing of posts, and increase the time spent on the network and thus advertising opportunities. By playing on human psychology, algorithms close users into opinion bubbles where users are exposed only to

information confirming their pre-existing beliefs and in which hate speech and other harmful content can flourish as is most likely to be shared and liked.

Online advertising, controlled to a large part by Facebook and Google, is also based on algorithms following the consumer's preferences. Online shops will use algorithms to present goods that the user browsing their website is most likely to purchase (e.g., based on their browsing history, clicks, and previous purchases). Finally, algorithms are also used to select the information that users will be prevented from seeing. Online forums, reader comment sections, and social networks employ algorithmic tools to filter out profanities, ethnic slurs, insulting language, etc., from their user's posts (Krönke 2020, 147). Online video platforms utilize similar systems (such as YouTube's Content ID) to identify and take down copyrighted content that was posted without authorization by copyright owners. Of course, the use of filtering algorithms can go much further and be (mis)used for political purposes. China is well known for blocking from its internet users any information that may be seen as critical to its political system or its leaders.

The word algorithm is used here as a catchall for any set of computer-implementable instructions used to sort, rank, and filter information: from simple computer programs searching for specific pre-defined expressions to advanced artificial intelligence (AI) systems that can process large data sets to achieve goals used in automated applications (Wischmeyer, Rademacher 2020, vii). Since the internet is now the basic information substructure of modern society, any technology that selects or limits access to information online may interfere with the exercise of fundamental rights online, particularly the freedom of expression and information guaranteed by article 11 of the Charter of Fundamental Rights of the EU.

The adoption of two major pieces of legislation in this connection is currently underway in the European Union. The draft Digital Services Act (COM(2020) 825 final), proposed by the European Commission in December 2020, will require very large online platforms to implement specific measures to mitigate systemic risks, such as the spreading of harmful disinformation. It is hard to conceive how to do that apart from relying on algorithmic tools for content moderation or recommendation. The draft Artificial Intelligence Act (COM(2021) 206 final), proposed in April 2021, will lay down harmonized rules for the use of AI systems, including the prohibition of certain AI practices and transparency rules for AI systems. However, these are future legislative acts the precise content of which is not yet certain. This paper will focus on the rules governing a specific set of algorithmic online filters the use of which is (indirectly) mandated by the Directive (EU) 2019/790 on Copyright in the Digital Single Market (DSM Directive), which entered into force in 2019 and requires online platforms to make best efforts to ensure the unavailability of unauthorized copyright works uploaded by their users.

PROHIBITION OF GENERAL MONITORING OBLIGATIONS

To protect the nascent internet intermediary industry from excessive legal risks arising from potential liability for any illegal information transmitted or stored by the users of their services, the E-Commerce Directive (2000/31/EC) introduced a haven for online intermediaries in 2000. The providers of mere conduit and cashing were exempt from liability as long as they provide the services in a technically correct manner and do not in any way tamper with the information transmitted or stored. Hosting, however, is a wider category of online services consisting of longer-term storage of information provided by the recipient of the service. Apart from the hosting of websites and blogs, this includes social networks, online video and music platforms, online marketplaces, cloud computing services, etc. Since hosting providers have greater technical possibilities of reviewing and removing the hosted information, they were exempt from the liability in exchange for cooperation in removing any illegally hosted content once notified about the illegality under the notice-and-takedown system (Edwards 2009, 65).

As long as the internet intermediary service remains "of a mere technical, automatic and passive nature" (recital 42), its provider is not required to check the legality of the information transmitted or stored, or to actively search for any unlawful content. The E-Commerce Directive reinforces this principle by expressly prohibiting the Member States from imposing on intermediary service providers any general obligation to monitor the information which they transmit or store, or any general obligation actively to seek facts or circumstances indicating illegal activity (article 15). Member States may only require service providers to inform the competent public authorities of alleged illegal activities by their users. The prohibition of imposing general monitoring obligations has been an essential tenet of the EU's internet regulation for more than twenty years. As the only feasible manner of sifting through the mounds of data uploaded daily by the users of social networks and other online platforms is by using automated algorithmic tools, this rule effectively banned the Member States from prescribing the use of such filters. Service providers, however, are free to use sorting, ranking, recommending, and filtering algorithms for their business purposes if they choose so.

Whereas article 15 of the E-Commerce Directive bans the imposition of general monitoring obligations, article 14(3) allows national courts or competent administrative authorities to order the service provider to terminate or prevent an infringement in specific cases. On this basis, intellectual property rights holders have pushed to achieve court-ordered monitoring obligations aimed at specific service providers. The Member States' courts did not offer a uniform answer to the question of whether it is permissible for a court to order an internet agent to filter potentially infringing user content. In cases *Atari Europe* and *GEMA v. Rapidshare,* the German Federal Court held that a diligent hosting provider should set up a system of automated filtering of infringing

content after they have received notifications that the use of the hosting services violates the rights of third parties.

The EU Court of Justice (CJEU) did not follow this reasoning. The case *Scarlet Extended* (C-70/10) concerned the question of whether an internet access provider could be ordered to filter all data traffic preventively to prevent illegal transfers of copyrighted content. At the suggestion of the collective organization Sabam, a Belgian court ordered the internet access provider to set up a system that would prevent its customers from transferring music files using peer-to-peer software. The CJEU held that such an order infringed the prohibition of general monitoring obligations and would disproportionately interfere with the freedom of economic initiative of the provider concerned. Traffic filtering would also violate the fundamental rights of users, namely the right to the protection of personal data and the freedom to receive and impart information. If the filter did not distinguish illegal content from legal content well enough, its use would make it impossible to download some legal content, which is unacceptable (Edwards 2009, 81).

The CJEU adopted similar reasoning in the case *SABAM v Netlog* (C-360/10) which concerned the social network Netlog, whose users shared on their profiles copyrighted music and audio-visual works from the catalog of the music collective organization Sabam. The collecting society requested that the operator of the online platform be ordered to prevent such unlawful use of copyrighted works. A Belgian court asked the CJEU whether it was permissible to order a hosting provider to set up a preventive system of filtering all information stored by the users to identify the works managed by Sabam and to prevent the unauthorized sharing of these works. The CJEU reiterated its view that the automatic filtering system would seriously infringe the service provider's freedoms of economic initiative while disproportionately interfering with users' rights to the protection of personal data and the freedom to receive and impart information. Accordingly, it held that the court should not order a hosting provider to establish a preventive system for filtering all user data.

Thus, it is an established position under the E-Commerce Directive that the duty of care cannot be interpreted in a way as to require intermediary service providers to set up an automated (algorithmic) system of filtering of any potentially illegal information uploaded or transmitted by their users. The article 14 requirement of the intermediary's actual knowledge or awareness of the unlawful information does not encompass any knowledge that the intermediary could obtain solely upon monitoring the hosted contents (Rowland, Kohl, Charlesworth 2012, 87). This applies even in cases of social networks and other mass platforms where it can be expected that a considerable share of user-uploaded content will infringe a copyright or other exclusive rights.

MOVE TOWARDS AUTOMATED CONTENT RECOGNITION IN COPYRIGHT LAW

Controversial Adoption of the DSM Directive

In the two decades since the adoption of the E-Commerce Directive, the role and influence of the main online platforms have grown immensely. Unlike vulnerable internet upstarts of the early 2000s, Facebook, Instagram, Twitter or YouTube are now internet giants generating vast advertising revenue at least indirectly derived from making available (unauthorized) copyrighted content uploaded by their users (Krönke 2020, 161). This situation has been met with increasing dissatisfaction by copyright holders as it both disturbed traditional channels for the distribution of copyrighted works as well as stymied the development of new paid online channels. The rightholders have pointed out that the technically neutral role of social networks and other interactive online platforms is questionable since their operators actively encourage users to publish and share their content, which generates high web traffic (Murray 2010, 107; Rowland, Kohl, Charlesworth 2012, 89). Since providing access to user-uploaded content is an essential part of the platform operator's business model, copyright holders increasingly demanded that the operators take a more active role in preventing copyright infringements.

The specific protection of copyright in online platforms was addressed in the DSM Directive, adopted in April 2019 after two years of tumultuous debate in which one of the most contentious issues was whether to mandate the use of automated upload filters to reduce the amount of copyright-infringing content uploaded on social networks. Article 13 of the initial Commissions proposal for a new directive (COM(2016) 593 final) required information society service providers who store and provide to the public access to large amounts of copyrighted content uploaded by their users to take measures to prevent the availability on their services of such content identified by rightholders. As an example of such measures, the Commission's proposal expressly mentioned the use of effective content recognition technologies, stressing that their use must be appropriate and proportionate. The use of content recognition technologies was also referred to in recital 39 of the proposal.

Prescribing the use of content recognition technologies (also referred to as upload filters) seems to go against the prohibition of general monitoring obligations from the E-Commerce Directive, which would lead to a significant overhaul of the EU's online liability rules. Whereas the publishers' and copyright holders' associations were generally supportive of the proposed solution, IT companies (including the internet giants) and many academics were firmly opposed. Critics have pointed out that algorithm-based automatic filtering is technically relatively inefficient. Experience with the use of algorithm-based automatic filtering tools (e.g., on YouTube) has shown that they are not very reliable even in the relatively simple task of recognizing copyrighted

content based on a digital fingerprint, let alone in considering the various limitations and exceptions to copyright. An additional concern is that the costs of operating filtering mechanisms may stifle small independent online platforms and thus increase the existing oligopoly of internet giants, most of them located outside the EU.

Article 17: Shadow Regulation?

After the discussion of several drafts of the contentions provisions in the European Parliament, the EU's legislative process resulted in today's article 17 of the DSM Directive, which tightens the liability rules of a new sub-category of online intermediaries: online content-sharing service (OCSS) providers. These are hosting providers whose main task is to store and give the public access to a large amount of copyrighted content uploaded by its users, which the service provider organizes and promotes for profit-making purposes.

When an OCSS provider gives the public access to copyrighted content uploaded by its users, this qualifies an act of communication to the public or an act of making available to the public by the service provider itself. This means that the service provider must obtain appropriate authorization for such use by the copyright holders, for instance by concluding a licensing agreement. Content-sharing platforms can no longer avoid liability for copyright infringements only by responding to takedown notices but must demonstrate that they have made best efforts to obtain authorization or, failing that, best efforts to ensure the unavailability of the unauthorized copyrighted content. OCSS providers must also make best efforts to prevent any future upload of the infringing content already removed upon receiving a takedown notice (Spindler 2020, 139).

The DSM Directive states that the application of article 17 should not lead to any general monitoring obligation, but due to the enormous amount of users' posts on content-sharing platforms it is hardly conceivable how the removal of all illegal content and the prevention of its future uploads could be achieved otherwise than by using automated filtering tools (Solmecke, Herr 2019; Spindler 2020, 16). Hence, although the Directive's provisions do not expressly mention content recognition technologies, they indirectly mandate their use, which is often referred to as shadow regulation. The conditions for the use of content recognition algorithms should be defined by the guidance provided by the European Commission and through the high industry standards referred to in article 17. Due to the potential conflict with human rights, the courts will certainly play an important role.

The Commission's Guidance

The suspicion that automated algorithmic content recognition will be the preferred, even if not legally mandated manner of complying with the content-sharing platform's best-efforts obligation under the DSM Directive was confirmed by the Commission's Guidance on article 17 (COM(2021) 288 final), issued in June 2021. The document stresses that the best-efforts provision should be implemented in a technologically neutral manner so that OCSS providers are free to choose the solution to comply with their obligations. However, the Commission also points out that the stakeholder dialogue showed that content recognition technology is commonly used today to manage the use of copyrighted content, even if it cannot be considered as the market standard for smaller service providers. The assessment of whether an OCSS provider has made its best efforts concerning specific protected content should be made on a case-by-case basis, according to the proportionality principle, considering the type, size, and audience of the service; the availability of suitable and effective means and the related costs; and the type of content uploaded by users.

The Commission's guidance resembles the CJEU's reasoning in joined cases *YouTube* and *Cyando* (C-682/18 and C-683/18), which was decided based on liability rules from the E-Commerce Directive, but after the adoption of the DSM Directive. The court assessed whether the video hosting platforms have taken 'credible and effective measures to counter copyright infringements after having been notified by the rightholder of specific violations. From the enumeration of various technical measures that might be considered sufficient in this regard, one can conclude that the CJEU does not consider upload filters as the only appropriate technological measure to prevent illegal uploads (Reda, Selinger 2021). The court also stressed that considering the particular importance of the internet to freedom of expression and information, a fair balance must be sought between, on the one hand, the protection of the intellectual property right and, on the other, the right to freedom to conduct a business enjoyed by service providers and the right to freedom of expression and information enjoyed by internet users (paras 65 and 138).

Poland's Action for the Annulment of Article 17

The ECJ is expected to provide further guidance on the acceptability of algorithmic content filtering when deciding on the action for the annulment of article 17 of the Directive lodged by Poland (C-401/19). Poland claims that the imposition of the obligation to make best efforts to ensure the unavailability and future uploads of infringing content require in effect that OCSS providers carry out prior automatic filtering of content uploaded online by users. Such preventive control mechanisms undermine the essence of the right to freedom of expression and information and do

not comply with the requirement that limitations imposed on that right be proportional and necessary. Advocate General Saugmandsgaard Øe concluded in his opinion delivered on 15 July 2021 that OCSS providers are under an obligation to engage in preventative monitoring; however, that obligation is specific, not general. The AG conceded that the contested provisions of the directive might indirectly force OCSS providers to use content recognition tools to filter the user-uploaded content, particularly where its employees would not be able to check all or most of the uploads. This obligation interferes with freedom of expression and information but remains compatible with the Charter of Fundamental Rights. In AG's understanding, OCSS providers are not authorized preventively to block all content that reproduces the copyrighted works but must block only manifestly infringing content. Conversely, in all ambiguous situations where exceptions and limitations to copyright might apply (e.g., short extracts or transformative works) the content concerned cannot be the subject of a preventive blocking measure since this could cause irreparable damage to freedom of expression (Rosati 2021).

Further Conflict of Automated Filtering with Fundamental Rights

In AG Saugmandsgaard Øe's opinion, any filtering algorithms under DSM Directive should be able to protect the fundamental rights exercised through the various limitations and exceptions to copyright prescribed by the Member States in cases where reasons of a public interest override the rightsholders' interests and refrain from blocking such non-infringing content. This seems optimistic considering the current technical level of content-recognition algorithms which are mainly limited to identifying content identical to the provided sample and often fail even at that task (Dawson 2018). It remains to be seen whether the more advanced algorithms will be able to recognize effectively the highly contextual instances where such exceptions and limitations might apply (such as parody, quotation, or incidental inclusion). Romero Moreno proposes that upload filters should be targeted specifically at copyright infringement on a commercial scale, which are more easily recognizable, ensuring the proportionality of the measure (Romero Moreno 2020, 164).

The problem will be further exacerbated if the statutory requirements for automated filtering are eventually expanded to other types of illegal content, such as terrorist materials, hate speech, child pornography, etc., where the recognition of illegal information and the protection of lawful communication might be even more difficult.

Perhaps the rapid development of AI-based software tools will increase the ability of automated contextual recognition of infringing versus non-infringing content. However, AI-based algorithms carry with them the black box problem: their content policies are difficult to understand and, due to their self-learning features, the precise criteria they use to identify, select, or classify information are constantly evolving and

may not be well understood even by the operators themselves (Wischmeyer 2020, 77). This makes it difficult any effective *ex post* judicial control over content filtering, whereas *ex ante* procedural hurdle to censorship is completely removed by automation (Llansó 2020, 3-4).

Even if the content filtering algorithms perform perfectly, however, the setting up of technical infrastructure for permanent monitoring of all internet content is dangerous. Free internet is an essential information infrastructure of modern society. The practice of scanning all online content for any possible illegalities is eerily similar to the manners of totalitarian states and the suspicion will linger that filtering algorithms could be misused for political or for commercial purposes. Hence the warning of the internet pioneers that the DSM directive takes an unprecedented step towards the transformation of the internet from an open platform for sharing and innovation into a tool for the automated surveillance and control of its users (Cerf *et al.* 2018).

CONCLUSION

The use of content recognition and other content sorting algorithms online is a reality that will not go away, regardless of the law. Evermore complex algorithms will be used to sort out the ever-increasing amounts of information. This is increasingly recognized by the CJEU's case law and in the EU's legislation, although both remain based on the principle of prohibition of general monitoring obligations. To protect the exercise of fundamental rights online, the operation of the algorithms will have to be regulated, and copyright law is just the first field where such attempts have been made in legislation. However, the regulation of filtering algorithms should not simply amount to delegating the task of censoring the internet to private service providers who are then free to determine themselves what information they will block (Institut Suisse 2017, 17-22). The intermediaries' neutral role in handling users' data is essential to preserve the internet's role as a public information infrastructure rather than just an offering of commercial electronic services completely within their provider's ambit and responsibility. To ensure democratic control of the internet, the operation of algorithms should be transparent, including transparency into what elements of the underlying data were important in developing the classifier of an algorithm (Llansó 2020, 5). The draft Artificial Intelligence Act contains transparency obligations for certain AI systems, but these would not apply to the content-filtering algorithms discussed here as they do not directly interact with humans, use biometric data or generate or manipulate content. Rather than using shadow regulation, the copyright legislation should expressly regulate the filtering algorithms used on content-sharing platforms and require their transparency. This would also allow the courts to preserve their role of assessing whether the measures strike a balance between the fundamental rights.

COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments:
Not applicable.

Funding:
The author's research for this article was supported by the Slovenian Research Agency (ARRS) under the research program P5-0337 "Legal challenges of the information society" and the research project J5-3107 "The development and use of artificial intelligence in the light of negative and positive obligations of a State to ensure the right to life".

Statement of human rights:
This article does not contain any studies with human participants performed by any of the authors.

Statement on the welfare of animals:
This article does not contain any studies with animals performed by any of the authors.

Informed consent:
Not applicable.

REFERENCES

1. Cerf, Vint, *et al*. 2018 "Article 13 of the EU Copyright Directive Threatens the Internet, letter to the President of the European Parliament", *Electronic Frontier Foundation,* 12 June. https://www.eff.org/files/2018/06/12/article13letter.pdf
2. Dawson, Aimee. 2018."Facebook censors 30,000-year-old Venus of Willendorf as 'pornographic'." *The Art Newspaper,* 27. February. https://www.theartnewspaper.com/news/facebook-censors-famous-30-000-year-old-nude-statue-as-pornographic
3. Edwards, Lilian. 2009. "The Fall and Rise of Intermediary Liability Online." In *Law and the Internet*, edited by Lilian Edwards and Charlotte Waelde, Oxford: Hart Publishing.
4. Institut suisse de droit compare. 2017.*Comparative Study on Blocking, Filtering, and Takedown of Illegal Internet Content.* Lausanne: Council of Europe.
5. Llansó, Emma J. 2020. "No Amount of 'AI' in Content Moderation Will Solve Filtering's Prior-Restraint Problem." *Big Data & Society*, 7:1-6.
6. Murray, Andrew. 2010.*Information Technology Law: The law and society.* Oxford: Oxford University Press.
7. Reda, Julia, and Joschka Selinger. 2021. "YouTube/Cyando – an Important Ruling for Platform Liability – Part 1." *Kluwer Copyright Blog*, 1 July, http://copyrightblog.kluweriplaw.com/2021/07/01/youtube-cyando-an-important-ruling-for-platform-liability-part-1.
8. Romero Moreno, Felipe. 2020 "Upload filters and human rights: implementing Article 17 of the Directive on Copyright in the Digital Single Market." *International Review of Law, Computers & Technology*, 34:2, 153-182.
9. Rosati, Eleonora. 2021. "AG Øe advises CJEU to rule that Article 17 is COMPATIBLE with the EU Charter of Fundamental Rights and should not be annulled Thursday." *TheIPKat*, 15 July. https://ipkitten.blogspot.com/2021/07/ag-e-advises-cjeu-to-rule-that-article.html.
10. Rowland, Diane, Uta Kohl, and Andrew Charlesworth. 2012. *Information technology law*. 4th edition. London, New York: Routledge.
11. Solmecke, Christian, and Anne-Christine Herr. 2019. "Rechtliche Analyse der Pro- und Contra Argumente zu Artikel 13 der geplanten EU Urheberrechtsnovelle." Wilde Beuger Solmecke Rechtsanwälte, 19 March. https://www.wbs-law.de/wp-content/uploads/2019/03/Analyse-Artikel-13-Version-1.2-WILDE-BEUGER-SOLMECKE-Rechtsanw%C3%A4lte.pdf.
12. Spindler, Gerald. 2020."Copyright Law and Internet Intermediaries Liability." In *EU Internet Law in the Digital Era: Regulation and Enforcement,* edited by Tatiana-Eleni Synodinou *et al*.Cham: Springer Nature.

13. Wischmeyer, Thomas and Timo Rademacher. 2020. "Preface: Good Artificial Intelligence." In *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer and Timo Rademacher. Cham: Springer.
14. Wischmeyer, Thomas. 2020. "Artificial Intelligence and Transparency: Opening the Black Box." In *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer and Timo Rademacher. Cham: Springer.
15. Krönke, Christoph. 2020. "Artificial Intelligence and Social Media." In *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer and Timo Rademacher. Cham: Springer.