

# Kluwer Copyright Blog

## Can algorithmic justice be just? ...if it's quick enough?

Marcella Favale (Bournemouth University) · Friday, October 22nd, 2021

On the 16<sup>th</sup> of October 2020, one year ago, a middle-school teacher, Samuel Paty, was beheaded by a terrorist who would not know of his existence if not for a number of videos posted on social media, against which Mr. Paty had filed for defamation with the local police. Yet, a law against publishing heinous content online was approved in France on the 13<sup>th</sup> of May last year. But in June, the Constitutional Council had repealed the article requiring take down of unlawful content within 24 hours, on the basis that it would trump freedom of expression.



Photo by Rafael Garcin on Unsplash

In light of this gruesome event, the decision to repeal the article has sparked heated political debate. The topic of the liability of internet intermediaries has never been so contentious.

Internet platforms have enjoyed immunity both in EU law and overseas. Two years ago, a new [Copyright Directive](#) (CDSMD) entered into force. It was due to be implemented by Member States this last June. This piece of legislation prompted criticism because it requires enhanced responsibility for Internet platforms that do not remove illegal content from their social media quickly enough.

But how quick must an action be to be done “quickly enough” (e.g. expeditiously)?

ISPs (Internet Service Providers) argue that by being “mere” intermediaries, they could not control the content that their subscribers were publishing online, and therefore could not be responsible for the unlawful actions taking place on their platforms. Rightsholders argued in response that intermediaries would often benefit from infringing activities. Hence, their provision of services could not be considered

totally neutral and therefore intermediaries should be held accountable.

Increasingly, the ecosystem of intermediaries is composite and multi-faceted. ISSP include search engines, auction platforms, e-commerce platforms, product comparison services, internet payment systems, self-publishing platforms, social media, etc.

The legislation however did not catch up with the evolution of technology and related business models. The definition of “hosting” as a legal concept has been subjected to increasing strain, especially in the last decade. Therefore, the issue of the responsibility of intermediaries has received a lot of attention from international organizations and legislators.

The [CDSMD](#) entered into force on the 7<sup>th</sup> of June 2019. The much-discussed Art 17 states that service providers should obtain permission from rightsholders to allow access to copyright content on their platforms, or they will be held accountable for the illegal uploading of infringing content on their websites unless they prove that they have made all their best efforts to get rightsholders’ authorisation or act expeditiously upon receiving substantiated notice of infringement from rightsholders. In doing so, intermediaries should ensure they account for the protection of some copyright exceptions, such as for criticism, quotation and parody. Also, they should put in place complaints and redress mechanisms that are available to users in case of disputes over content moderation. General monitoring for potential infringement is excluded. Also, a new [EU Regulation](#) obliges platforms to remove terrorist content from their online services within one hour. This new law applies from the 7<sup>th</sup> of June 2021.

Currently, two new pieces of EU legislation are underway to horizontally streamline platforms’ filtering duties ([the Digital Services Act](#) and [the Digital Markets Act](#)). But the debate on these new norms is still running, notably about different types of illegal content and whether they deserve different treatment.

Their aim is to “set out uniform rules for a safe, predictable and trusted online environment, where fundamental rights enshrined in the Charter are effectively protected.” To this end, the new DSA will establish a network of Digital Service Coordinators that will assist and supervise the enforcement of these norms within Member States. It will establish a duty to act following an order of the judiciary, but also a system of Notice and Action for private individuals, along with an internal Complaint Handling system. Prohibition of general monitoring is confirmed by the Regulation, and no further detail is provided on delays for taking down allegedly illegal content, beyond the usual mandate to “act expeditiously”.

Among copyright scholars, [Senftleben and Angelopoulos](#) have made the argument that expressing content moderation duties across EU Directives and across different areas of law such as copyright, trademark and defamation may prove challenging for the inherent differences among “the scope of rights and the characteristics of infringement”.

In the field of defamation, the CJEU established in [Glawischnig](#) that Art. 15 of the E-commerce directive does not preclude injunctions against intermediaries to take down (by relying on automated technologies in the case of identical content) the content

that has been declared illegal by the competent authority.

Therefore, while potentially (depending on national law) an injunction is needed to take down defamatory content, no order injunction is routinely needed to take down copyright-infringing content and to filter further instances of the same content. What is the justification for this disparity? It should be, arguably, all about the balance of the fundamental rights involved, which the platform **needs to respect and enforce**. In cases of defamation, freedom of expression is systematically deployed as a defence, whereas in relation to copyright this is not always the case. Hence, in practice, copyright-infringing material is more likely to be expeditiously removed upon simple request of the rightsholder, while defamatory content needs to undergo the scrutiny of a Court of Law. Of course, from a practical point of view, copyright infringement may be easier to detect and identify automatically, as it mainly entails visual or audial comparison between content, whereas defamatory material, at least in the first instance, requires specific examination.

**No specific legislation exists** in EU and US law against online hate speech and the most utilised legal remedies against this growing behaviour are lawsuits for defamation and harassment. Prosecuting these cases is **very difficult** because of freedom of speech. Also, pending a court ruling on the legal or illegal nature of the speech, the platform does not have a duty to act, even when the author of the content is known. The sheer lapse of time allows the hateful content, potentially inciting to violence, to reverberate across the globe, unhindered. The traditional recourse to off-line justice is therefore powerless to address the most dangerous forms of this illegal behaviour, at least until expeditious remedies are available specifically against online defamation/hate speech.

It has been argued that automatic filtering is a good solution, indeed an ideal solution, when the technology is effective enough to balance all the rights at stake. It is not difficult, given the current state of technology, to conceive the implementation of an algorithm that identifies a picture, a song, a video online, which corresponds to a declared copyright-protected work. It is arguably more difficult for the algorithm to realize whether this is a reproduction exempted from copyright protection because of criticism, quotation or parody... but thanks to Machine Learning this is not unconceivable.

However, how can we distinguish automatically a case of political defamation, from one based on religious, ethical, racial motivation? We can assume that an algorithm can identify hate speech from vocabulary related to religion, race, sex/sexual orientation and disability. But can the right “expeditiousness” of action to be taken against potentially illegal behaviour be assessed by technological measures? Can the consequences of illegal behaviour be assessed by a robot?

Could an algorithm have foreseen the beheading of a good high-school teacher?

Of course, some of the rules developed under current practices and jurisprudence can assist. For example, according to the proposed DSA Regulation expeditious treatment would be given to notices when the information is provided by trusted flaggers; and users that are likely to infringe due to recidivism will be more easily blocked.

Certainly, in the case of religious hate speech these norms could be successfully applied to some “influencers”. However, any action will have to be balanced against Freedom of Religion, which is also a fundamental right.

It all seems to simmer down to the balance between fundamental rights.

When it comes to defamation, the analysis becomes more difficult. On the one hand, we have cases pitting multinational corporations against civil rights groups, where freedom of expression, underpinning the defence of the environment or the economy, is pitched against the financial interests of multinationals. In other cases, critical forms of expression (parodies) are aimed at corporations in the framework of cultural debate. And then, of course, we have defamation cases involving public figures, where freedom of expression is instrumental to a healthy democratic debate.

Be that as it may, the far-reaching influence of unmoderated (or insufficiently moderated) social media has produced defamation cases against private individuals on delicate topics such as religion, sex, race, disability, etc. These are defined as “hate speech” and are a potential trigger for physical violence, as the – hopefully – landmark case of Mr. Paty shows. Also, and even more dangerously, we have cases of slanderous statements inducing social unrest and attack on democratic institutions, even producing [physical violence](#). Finally, if we enlarge the picture so as to include unverified statements (known as “fake news”) maliciously diffused to sway public opinion and potentially [impacting on geopolitics](#), we realize that the consequences of these different instances of illegal behaviour are simply not comparable.

“Doing Nothing” is not a conceivable option. In this sense, the initiative of the EU Commission to produce new regulations is welcome. However, a more nuanced approach, prompting a range of mandatory actions to be taken by platforms, could be more appropriate to fight different illegal behaviours, with potentially very different consequences. For example, immediate action could be required from platforms to take down the most potentially dangerous content (instead of “expeditious action”). This, of course, will require a consensus on what is “most potentially dangerous” content, which will prove contentious. But this is not a good reason to give up the challenge.

To sum up, the road to horizontal harmonization of algorithmic justice for large platforms appears strewn with many obstacles, but this is a worthy pursuit. Meanwhile, significant investment in research and development will have to be factored in to make technological measures capable of performing these duties.

Robots, as judges, will have to be able to assess the potential consequences of various illegal behaviour and take proportionate action accordingly. In the meantime, reliance on human judgment will be unavoidable.

This post is based on my working paper that you can find [here](#). Comments are welcome!

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe [here](#).

## Kluwer IP Law

The **2021 Future Ready Lawyer survey** showed that 81% of the law firms expect to view technology as an important investment in their future ability to thrive. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?

Learn how **Kluwer IP Law** can support you.



This entry was posted on Friday, October 22nd, 2021 at 10:53 am and is filed under [Copyright](#), [Digital Single Market](#), [European Union](#), [Liability](#). You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.