

# Automatic filtering and freedoms: Can we break free from a centralized Internet?

Lionel Maurel,  
*Centre National de la Recherche Scientifique (CNRS)*

## **Abstract:**

Whether at the French or European levels, plans are springing up for regulating the big centralized platforms through requirements about the automatic filtering of contents in order to fight on line against counterfeits, hate or terrorism. These technical measures are likely to imperil the exercise of fundamental freedoms, in particular by deeply modifying the rules on the responsibility of technical intermediaries. Paradoxically, they also risk reinforcing the domination of the giants (like GAFAM) by compromising the possibility of re-decentralizing uses on the Internet.

Whether to fight against counterfeits, online hatred or terrorism, proposals are proliferating for forcing the big centralized platforms to adopt measures for automatically filtering their contents. This is often claimed to be a way to make up for a lack of regulations, since the Web giants are thus forced to assume a responsibility that they have shirked till now.<sup>1</sup>

Since the start of the 21st century, the EU has granted a protective status to the “technical intermediaries” that claim to be “web hosting services”. Unlike the “publishers” of contents for the Web, these intermediaries are carriers of what their users post and are not directly responsible for what is published on line with the help of their services.<sup>2</sup> In addition, member states are normally not allowed to force these platforms to set up a system of general surveillance of the contents circulating through their servers.<sup>3</sup> They cannot be held liable unless they fail to react rapidly and take down the illegal contents reported to them — a procedure called NOTICE AND TAKE DOWN.

These rules represent a compromise between the protection of the freedom of expression and the responsibilities of Web hosting services. For a long time, they have served as a cornerstone for the Internet’s development. Thanks, in part, to these rules, the so-called Web 2.0 was able to emerge in the middle of the first decade of the new century with the social media and networks and the now symbolic websites of Wikipedia, Flickr, YouTube, Facebook and Twitter, to mention but a few.

However this unstable equilibrium has gradually come apart, as the “platformization” of the Internet has had many negative consequences that compel lawmakers to respond. An oft mentioned argument for reforming this system claims that the big platforms (*e.g.*, GAFAM: Google, Apple,

---

<sup>1</sup> This article, including quotations from French sources, has been translated from French by Noal Mellott (Omaha Beach, France). The translation into English has, with the editor’s approval, completed a few bibliographical references. All websites have been consulted in July 2019.

<sup>2</sup> These rules stem from EU Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 “on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market” available at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000L0031>. It was transposed into French law as the LCEN act “for confidence in the digital economy” in 2004. On this, see: BAYART B. (2018) “Intermédiaires techniques”, blog of 14 October reproduced at <https://www.les-crises.fr/intermediaires-techniques-par-benjamin-bayart/>.

<sup>3</sup> A principle formulated by the Court of Justice of the European Union (CJEU) in its SABAM decision in 2013. See REES M. (2012) “UE: la CJUE bloque le filtrage généralisé chez les hébergeurs”, 16 February, available at <https://www.nextinpact.com/archive/69013-cjue-sabam-hebergeur-filtrage-blocage.htm>.

Facebook, Amazon, Microsoft) can no longer be treated like merely passive intermediaries, since they actively diffuse and rank contents, even though they do not produce them.<sup>4</sup>

Starting in the 2010s, regulations were adopted that allowed for orders to be issued to Internet service providers (ISPs) instructing them to block websites. Such an order could come from the judiciary or public administration, although the latter possibility spurred questions about the risks of censorship and arbitrariness that would result from bypassing judges. Initially introduced to fight against pedopornography, the blocking of Web contents has continually expanded into fields as different as online betting, counterfeits, and contents that are sexist or homophobic or, recently, that glorify terrorism.<sup>5</sup>

The paradigm of these actions was that the information circulates on websites and that the latter can be blocked. However, this is no longer so since most communications have moved toward platforms with Web hosting services. Centralization has thus grown, and the focus is now shifting toward filtering (instead of blocking). This substantially alters the debate. Although blocking contents raises problems about freedom of speech, it fundamentally involves a human decision — what is no longer the case with filtering. Filtering slides toward an automated application of the law, an application based on algorithms or artificial intelligence. Filtering amounts to subcontracting to machines the application of complicated rules and concepts with often fuzzy definitions in legal texts. It will have repercussions on the upholding of fundamental rights.<sup>6</sup>

A generalization of filtering on the Internet, as is now being introduced in legal texts, might have paradoxical consequences. Far from being a means whereby states reassert their sovereignty opposite big platforms, filtering could, on the contrary, reinforce these platforms' dominant position by delegating to them essential duties of police and justice. Furthermore, the obligations related to filtering might drastically weaken the small players through whom a re-decentralization of the Internet might be achieved. To break out of this vicious spiral, other possibilities of regulation must be explored by tackling the very causes of this centralization.

## **How “Code is law” has turned into “Law is code”**

“Code is law” is the title of a well-known article by Lawrence Lessig (2000) about the relations between the law and technology during the Internet era.<sup>7</sup> This American attorney argued that upholding fundamental freedoms on line depends more on the architecture of networks and the protocols underlying them than on applying legal rules.

Nearly twenty years after the publication of this seminal article, we can conclude that Lessig's vision has been, in part, invalidated, mainly because he underestimated the law's capacity for acting back on the technical infrastructure. YouTube is exemplary in this regard. This platform has been a laboratory that has designed filtering for the Web and experimented with it. The year following its acquisition by Google in 2006, YouTube was sued by Viacom, a multimedia group, for copyright

---

<sup>4</sup> CSPLA [Conseil Supérieur de la Propriété Littéraire et Artistique] (2017) “La protection du droit d'auteur sur les plateformes numériques. Les outils existants, les bonnes pratiques et leurs limites”, report head by O. Japot of 19 December, 55p. available via <https://cdn2.nextinpact.com/medias/cspla-rapport-mesures-techniques-sur-les-plateformes-numeriques.pdf>.

<sup>5</sup> On blocking websites that promote terrorism, see “Blocage des sites terroristes. 3<sup>e</sup> rapport de contrôle de la CNIL” on <https://www.vie-publique.fr/actualite/alaune/blocage-sites-terroristes-3e-rapport-contrrole-cnil.html>.

For the full report: LINDEN A. (2018) “Rapport d'activité 2017. La personnalité qualifiée prévue par l'article 6-1 de la loi n° 2004-575 du 21 juin 2004 créé par la loi n° 2014-1353 du 13 novembre 2014 renforçant les dispositions relatives à la lutte contre le terrorisme”, CNIL report (Paris: Commission Nationale Informatique et Libertés), 16p. available via [https://www.ladocumentationfrancaise.fr/ocfra/rapport\\_telechargement/var/storage/rapports-publics/184000339.pdf](https://www.ladocumentationfrancaise.fr/ocfra/rapport_telechargement/var/storage/rapports-publics/184000339.pdf).

<sup>6</sup> See the position of David Kaye, UN human rights expert, on filtering as a means of fighting against counterfeits: “EU must align copyright reform with international human rights standards, says expert” on <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24298&LangID=E>.

<sup>7</sup> LESSIG L. (2000) “Code is law: On liberty in cyberspace”, *Harvard Magazine*, January/February, 5p. available at <https://harvardmagazine.com/2000/01/code-is-law-html>.

infringement.<sup>8</sup> Lest the judge object to its status as a Web hosting service and hold it liable, YouTube chose to settle out of court. Realizing that the danger could eventually be mortal, YouTube sought to “pacify” its relations with copyright-holders by developing a system (Content ID) for recognizing contents.

The Content ID system uses a robot to identify protected works (images, music, videos) that copyright-holders have registered on line with the platform. Copyright-holders set the consequences for matches, *i.e.*, when the contents uploaded match a registered work. They have three choices:

- block the video and sanction the cybernaut who posted it;
- claim the income from the advertising associated with it; or
- let the video on line.

At the start, YouTube hoped that this system would foster the diffusion and reuse of contents on its platform. These hopes have been disappointed. Rightful owners have mainly used Content ID for purposes of surveillance and control. Meanwhile, this algorithm of content recognition has often been blamed for reporting false positives.<sup>9</sup>

Nonetheless, this technical solution had important legal consequences. As a platform grows in size, it is ever harder for it to fulfill its obligations as a Web hosting service, in particular the obligation to manually take down contents upon notification. Filtering automates the process without the platform losing its status as a hosting service. After all, a machine does the job of identification without implying that the firm is directly aware of the contents diffused by its services. However the intent of regulations about the responsibility of Web hosting intermediaries is not upheld, since, at the origin, these regulations sought to avoid forcing platforms to oversee the contents.

The paradox is, therefore, that a platform like YouTube has voluntarily opened the way toward automated filtering in order to cope with the higher volume of its business; but it has not been forced to do so by the law. Since an algorithm is capable of interpreting and applying rules (such as copyright law), Content ID has reversed “Code is law”, turning it into “Law is code”. The application of the law can be “encoded” and automated via algorithms. The YouTube precedent has had far-reaching effects, as filtering has gradually spread to online platforms — voluntarily (Dailymotion, Instagram, Facebook and Dropbox), or under pressure from rightful owners (as at SoundCloud, which put up a long resistance).<sup>10</sup>

## **What filtering does to fundamental freedoms**

Filtering is not without consequences. This technique has set off a process that jeopardizes fundamental rights on the Internet, starting with freedom of speech. Under the usual notice-and-take-down procedure, the party that considers an online content to be illicit has to notify the platform; and the latter has to evaluate — an analysis made by a human being — whether the request is founded. Under existing regulations, the hosting service may refuse to grant the request to take down a post if the contents are not “*manifestly illicit*”.<sup>11</sup> In case of disagreement, the party who notified the platform will have to sue, and a judge makes the final decision in the presence of all parties.

---

<sup>8</sup> “Viacom International Inc. v. YouTube, Inc.” at [https://en.wikipedia.org/wiki/Viacom\\_International\\_Inc.\\_v.\\_YouTube\\_Inc.](https://en.wikipedia.org/wiki/Viacom_International_Inc._v._YouTube_Inc)

<sup>9</sup> LANGLAIS P.C. (2016) “Comment fonctionne Content ID?”, 20 September, on <https://scoms.hypotheses.org/709>.

<sup>10</sup> MAUREL L. (2013) “Filtrage: quand SoundCloud joue au RoboCopyright”, blog of 18 April on <https://scinfolex.com/2013/04/18/filtrage-soundcloud-fait-sa-police-du-copyright/>.

<sup>11</sup> For a recent application: Twitter refused to take down videos at the request of the comic Gad Elmaleh because it had doubts about whether they were “*manifestly illicit*”. See: PCS Avocat (2019) “Plagiat, contrefaçon et hébergeurs de contenus: Gad Elmaleh contre Twitter”, 27 February, on [https://www.pcs-avocat.com/plagiat—contrefacon-et-hebergeurs-de-contenus—gad-elmaleh-contre-twitter\\_ad240.html](https://www.pcs-avocat.com/plagiat—contrefacon-et-hebergeurs-de-contenus—gad-elmaleh-contre-twitter_ad240.html).

Filtering modifies this procedure since it is the platform that intervenes in advance to take down contents that an automatic analysis has reported as not complying with the law. The burden of proof is reversed, since the user is the party who has to object to his contents being taken down. Most platforms have internal procedures for handling appeals. In the case of YouTube however, copyright holders have the last say about whether to keep or lift automatic sanctions. The user will then have to sue to protect his rights, but this happens very seldom. For this reason, filtering has often been accused of amounting to a private system of police and justice.

The copyright directive approved by the European Parliament on 26 March 2019 goes even farther. Centralized platforms might be forced to filter contents *ex ante*.<sup>12</sup> Whereas a robot like Content ID intervenes once a video has been posted, this new directive implies switching to NOTICE AND STAY DOWN, since platforms would have controls performed upstream, when the contents are uploaded. As a result, it would be impossible to post contents at a time when YouTube's algorithm is still making mistakes.

These flaws in the operation of filtering are not just technical problems. They bring to light a probably unsurpassable limit in the process of automating the application of law. Most laws in Europe allow for exceptions to copyright law, in particular for remakes of protected contents for the purpose of parody. Content ID is structurally incapable of telling the difference between a simple rerun and a remake as a parody. To do so, it would have to be capable of detecting humor. The fact that machines are blind to such subtleties can have serious consequences (in, for example, the online fight against hatred or terrorism). The definition of incitement to hatred (or terrorism) is fuzzy in legal texts, and this could lead to serious abuses when the enforcement of regulations is subcontracted to algorithms. We are not surprised to observe that, despite the talk about technical solutions and the boasts about artificial intelligence, the big platforms still, in fact, rely heavily on human moderators (who often dwell in countries in the global South and have deplorable working conditions) to actually take down posts.<sup>13</sup>

These structural inadequacies have not, however, kept lawmakers from leaning toward an extension of filtering. Such is the case in discussions on an EU antiterrorist regulation, which would require taking down contents within one hour after notification by the judicial or administrative authorities of member states. This draft calls for platforms to take measures in advance so as not to circulate contents glorifying terrorism. This means filtering, which opponents of the regulation claim will be a step toward an "*automation of political censorship*".<sup>14</sup> The same line of reasoning is pursued in the bill of law about online hatred and cyberstalking now imagined by the French government. It advocates automated filtering... with the same foreseeable consequences on the freedom of expression.<sup>15</sup>

---

<sup>12</sup> REES M. (2018) "Pourquoi la directive Droit d'auteur peut aboutir à un filtrage de l'upload", 11 December, on <https://www.nextinpact.com/news/107399-pourquoi-directive-droit-hauteur-peut-aboutir-a-filtrage-upload.htm>.

<sup>13</sup> On this, see the documentary by R. Riesewieck and H. Block broadcast on Arte on 28 August 2018, "Les nettoyeurs du web".

<sup>14</sup> TRÉGUER F. (2019) "Vers l'automatisation de la censure politique", 27 February, on <https://www.laquadrature.net/2019/02/22/vers-lautomatisation-de-la-censure-politique/>.

<sup>15</sup> "Mahjoubi et Schiappa croient lutter contre la haine en ligne en méprisant le droit européen", 14 February 2019, on <https://www.laquadrature.net/2019/02/14/mahjoubi-et-schiappa-croient-lutter-contre-la-haine-en-mepriant-le-droit-europeen/>.

## **A cause of concern for decentralized alternatives**

Filtering might, ironically, tighten the hold of the big platforms instead of loosening their grip. As pointed out, Google developed the first automated filtering techniques for its subsidiary YouTube. It invested more than \$100 million in Content ID.<sup>16</sup> Facebook has also devoted considerable funds to developing a filtering of its own, which will work, it is claimed, by using artificial intelligence.<sup>17</sup>

This sort of technology has high costs, which only the biggest players can lay out. Regulations with technical requirements for filtering contents will mechanically favor the biggest players to the detriment of the small ones, since only the former are capable of rolling out such systems. Worse yet, the giants of the Web, such as Google and Facebook, are the very parties who have, till now, developed the most effective technology of this sort; and they will, therefore, be able to sell their solutions to competitors and thus make them dependent. For this reason, it is fallacious to present filtering as a technique that will ensure the digital sovereignty of Europe over GAFAM. On the contrary, the generalization of filtering will but reinforce the digital ecosystem's dependency on the big American operators.

This is probably the reason that GAFAM is not opposing such requirements but, to the contrary, is actively lobbying for the deployment of filtering systems. Google's reason for opposing the copyright directive has less to do with the question of filtering (already implemented on YouTube) than with the directive's requirement of more royalties for copyright holders. In a recent editorial, Mark Zuckerberg, the founder of Facebook, did not oppose filtering, but argued for *"third-party bodies to set standards governing the distribution of harmful content"*.<sup>18</sup> Facebook has advanced by organizing its own in-house "court of appeals", which it claims to be an *"independent body"* for examining complaints when posts are taken down. This measure reflects its determination to institutionalize a private justice; and questions are eluded about the circumvention of official tribunals that the extension of filtering will bring about.

According to Félix Tréguer, these trends are much more the sign of a *"state-GAFAM merger"* than evidence of sovereign states taking the platforms back under control: *"If we think of the state not as a clearly identified bloc (as jurists do) but more like a set of practices and a rationality, which Michel Foucault called 'governmentality', then it is clear what these trends let us see, namely the incorporation of these private actors in the state — the cooptation of their infrastructures and the diffusion of their know-how for processing and analyzing the masses of data now crucial to contemporary forms of government. What we are seeing is, therefore, a merger much more than a competition between states and GAFAM, the latter seeking to take the place of governments."*<sup>14</sup>

"Platformization" is far from inevitable however. Uses can still be decentralized, thanks, in particular, to recent standards of interoperability.<sup>19</sup> Services such as Mastodon or Peertube — free, open alternatives to Twitter and Youtube respectively — prove that it is technically possible to propose a level of services comparable to the offers made by the big platforms but without centralizing contents. These alternatives federate a multitude of activities that, hosted by various actors (individuals, associations, administrations, firms), can communicate with each other. Framasoft, an association, has formed a coalition of actors (CHATONS)<sup>20</sup> capable of shared services in compliance with a charter, which lays down the major principles, such as the use of free software or the refusal to process personal data.

<sup>16</sup> See the Wikipedia article "Content ID (algorithm)" at [https://en.wikipedia.org/wiki/Content\\_ID\\_\(algorithm\)](https://en.wikipedia.org/wiki/Content_ID_(algorithm)).

<sup>17</sup> SOMINITE T. (2018) "AI has started cleaning up Facebook, but can it finish?", 18 December, on <https://www.wired.com/story/ai-has-started-cleaning-facebook-can-it-finish/>.

<sup>18</sup> ZUCKERBERG M. (2019), "The Internet needs new rules. Let's start in these four areas", *Washington Post*, 30 March, available at [https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f\\_story.html?utm\\_term=.fb9b915019d1](https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html?utm_term=.fb9b915019d1).

<sup>19</sup> See, in particular, the ActivityPub standard drafted by W3C on the interoperability of the "Social Web": <https://www.w3.org/TR/activitypub/>.

<sup>20</sup> CHATONS [Collectifs des Hébergeurs Alternatifs Transparents Ouverts Neutres et Solidaires]. See <https://chatons.org/>.

These possibilities show how a free and open Internet could be founded anew, by turning away from the platformization characteristic of the early 21st century. This alternative is what generalized filtering would nip in the bud. These small federations of actors are definitely unable to deploy filtering, but they would be exposed to the whiplash of liability, an unbearable consequence given the risks.

If public authorities really want to oppose GAFAM, they could use other legislative levers: the protection of personal data, the taxation of digital firms, or the fight against abuses of dominant position. Lawmakers' obsession with filtering mainly evinces their resignation to accepting domination by the big platforms. It is a way of coping with the consequences of this domination without tackling the real causes of the problem.