Filtrage automatique et libertés : peut-on sortir d'un Internet centralisé ?

Lionel Maurel

Filtrage automatique et libertés : peut-on sortir d'un Internet centralisé ?

Par **Lionel MAUREL**
CNRS

Que ce soit pour lutter contre la contrefaçon, la haine en ligne ou l'apologie du terrorisme, on assiste aujourd'hui à la multiplication des propositions visant à imposer aux grandes plateformes centralisées des mesures de filtrage automatique des contenus.

Cette formule est souvent présentée comme une manière de combler une lacune de la réglementation en obligeant les géants du web à assumer une responsabilité à laquelle ils se dérobent. Depuis le début des années 2000, un statut protecteur a en effet été mis en place au profit des intermédiaires techniques pouvant revendiquer la qualité d' «hébergeurs ». À l'inverse des « éditeurs », ces acteurs diffusent des contenus postés par leurs utilisateurs et ne sont pas frontalement responsables des publications s'opérant *via* leurs services[1]. Il est en outre normalement interdit aux États de les obliger à mettre en place une surveillance généralisée des contenus circulant sur leurs serveurs[2]. Leur responsabilité n'est engagée que s'ils ne réagissent pas rapidement pour retirer des contenus leur ayant été signalés comme illicites (système dit du *notice and take down*).

Ces règles ont été instaurées pour opérer un compromis entre protection de la liberté d'expression et responsabilité des intermédiaires techniques. Elles ont longtemps constitué l'une des pierres angulaires sur laquelle le développement d'Internet s'est appuyé. C'est notamment en partie grâce à cette réglementation que le web dit 2.0 – celui des réseaux et des médias sociaux – a pu prendre son essor à partir du milieu des années 2000, avec des sites emblématiques comme Wikipédia, Flickr, YouTube, Facebook, Twitter et tant d'autres.

Néanmoins, cet équilibre fragile s'est peu à peu rompu, à mesure que le phénomène de « plateformisation » de l'Internet a engendré de multiples conséquences négatives face auxquelles le législateur a été sommé de réagir. Un argument souvent invoqué pour justifier des réformes s'appuie sur le fait que les grandes plateformes, type GAFAM, ne peuvent plus être considérées comme de simples intermédiaires passifs. Elles joueraient en effet un rôle actif dans la diffusion et la hiérarchisation des contenus, quand bien même elles ne sont pas à l'origine de ceux-ci[3].

---

[1] Ces règles découlent de la directive dite eCommerce du 8 juin 2000, « relative à certains aspects juridiques des services de la société de l'information », transposée en France par la loi pour la confiance dans l'économie numérique (LCEN) adoptée en 2004. Sur ce sujet, voir BAYART B. (2018), « Intermédiaires techniques ».
https://www.les-crises.fr/intermediaires-techniques-par-benjamin-bayart/

[2] Principe posé par la Cour de Justice de l'Union européenne dans sa décision SABAM rendue en 2013. Voir REES M. (2013), « UE : la CJUE bloque le filtrage généralisé chez les hébergeurs »
https://www.nextinpact.com/archive/69013-cjue-sabam-hebergeur-filtrage-blocage.htm

[3] Voir, en ce sens, le rapport suivant remis par le Conseil supérieur de la Propriété littéraire et artistique (CSPLA) (2017), « La protection du droit d'auteur sur les plateformes numériques »

À partir des années 2010, on a d'abord assisté à l'adoption d'une série de réglementations visant à imposer le blocage de sites en passant par des injonctions adressées aux Fournisseurs d'Accès Internet (FAI). Une telle technique peut être mise en œuvre soit par voie judiciaire, soit par voie administrative, cette dernière hypothèse soulevant de nombreuses questions quant aux risques de censure et d'arbitraire liés au contournement du juge. Introduit initialement en matière de lutte contre les contenus pédopornographiques, le blocage n'a cessé de s'étendre dans des domaines aussi divers que le jeu en ligne, la contrefaçon, les propos sexistes ou homophobes ou, plus récemment, l'apologie du terrorisme[4].

Le blocage relève encore d'un paradigme dans lequel l'information circule sur des sites susceptibles d'être bloqués, ce qui n'est plus le cas lorsque l'essentiel des communications migrent vers des plateformes d'intermédiation. Ce poids croissant de la centralisation des échanges fait que l'attention se tourne à présent vers le filtrage, en changeant au passage substantiellement les termes du débat. Car si le blocage est déjà problématique vis-à-vis de la liberté d'expression, il reste fondamentalement une technique impliquant la décision humaine. Ce n'est plus le cas avec le filtrage qui provoque un glissement vers une application automatisée du droit, s'appuyant sur des technologies comme les algorithmes ou l'intelligence artificielle. Filtrer les plateformes revient à sous-traiter à des machines l'application de normes complexes et de notions souvent définies de manière floue dans les textes, avec à la clé des répercussions préoccupantes sur la garantie des droits fondamentaux[5].

La généralisation du filtrage sur Internet, telle qu'elle se dessine actuellement dans les textes, pourrait en outre avoir des conséquences paradoxales. Loin de constituer un moyen pour les États de réaffirmer leur souveraineté face aux grandes plateformes, le filtrage pourrait au contraire conduire à renforcer leur position dominante en leur délégant l'exercice de fonctions essentielles de police et de justice. Dans le même temps, les obligations de filtrage fragilisent drastiquement les petits acteurs à travers lesquels une redécentralisation des usages sur Internet pourrait encore s'opérer. Si l'on veut sortir de cette spirale infernale, ce sont d'autres pistes de régulation qu'il faut explorer, en s'attaquant aux causes mêmes de la centralisation.

Comment *Code Is Law* s'est renversé en *Law Is Code*

Code Is Law (« Quand le Code fait loi »), c'est le titre d'un célèbre article publié par le juriste américain Lawrence Lessig, en 2000, à propos des rapports entre le droit et la

---

https://cdn2.nextinpact.com/medias/cspla-rapport-mesures-techniques-sur-les-plateformes-numeriques.pdf

[4] Sur le blocage des sites terroristes, voir CNIL (2018), « Blocage des sites terroristes. 3e rapport de contrôle de la CNIL »
https://www.vie-publique.fr/actualite/alaune/blocage-sites-terroristes-3e-rapport-controle-cnil.html

[5] En ce sens, voir la position de David Kaye, rapporteur spécial à l'ONU pour la liberté d'expression, à propos des mesures de filtrage envisagées pour lutter contre la contrefaçon. KAYE D. (2019), "EU must align copyright reform with international human rights standards, says expert, United Nations Human Rights"
https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24298&LangID=E

technique à l'heure d'Internet[6]. Lessig soutenait la thèse selon laquelle la préservation des libertés en ligne dépend davantage de l'architecture technique des réseaux et des protocoles qui les soutiennent que des normes juridiques applicables.

Près de vingt ans après la publication de cet article séminal, on peut dire que la vision de Lessig a été en partie invalidée, notamment parce qu'il a sous-estimé la capacité du droit à rétroagir sur l'infrastructure technique. La trajectoire suivie par la plateforme YouTube est exemplaire à cet égard : cette plateforme a constitué un véritable laboratoire où le filtrage du Web a été conçu et expérimenté. L'année suivant son rachat par Google en 2006, YouTube s'est trouvé pris dans un procès intenté par le groupe multimédia Viacom pour violation du copyright[7]. Pressentant que les juges risquaient de remettre en cause son statut d'hébergeur et d'engager sa responsabilité, YouTube a choisi d'étendre la plainte par un règlement amiable de la plainte. Mais estimant courir un danger potentiellement mortel, l'entreprise a développé un système de reconnaissance des contenus – dénommé Content ID – destiné à « pacifier » ses relations avec les ayants droit.

Le dispositif utilise un robot pour identifier des œuvres protégées en ligne (images, musiques, vidéos) sur la base d'empreintes fournies à la plateforme par les titulaires de droits. Ces derniers peuvent également paramétrer les conséquences en cas d'identification d'une correspondance, avec trois options :

- bloquer la vidéo et sanctionner l'internaute responsable de sa publication ; - s'approprier la rémunération publicitaire associée ;
- ou laisser la vidéo circuler.

YouTube espérait à l'origine que Content ID favoriserait la diffusion et la réutilisation des contenus sur la plateforme, mais, dans les faits, cet espoir a largement été déçu : les titulaires de droits ont surtout utilisé Content ID à des fins de surveillance et de contrôle, tandis que l'algorithme de reconnaissance des contenus a souvent été pointé du doigt pour générer de nombreux faux positifs[8].

Mais c'est surtout sur le plan juridique que les conséquences de cette solution technique ont été importantes. À mesure qu'une plateforme croît en taille, il devient de plus en plus difficile pour elle de respecter les obligations liées à son statut d'hébergeur, en particulier l'obligation de retirer manuellement des contenus signalés. L'utilisation d'un dispositif de filtrage permet d'automatiser le processus, sans perdre sa qualité d'hébergeur, puisque le repérage est effectué par une machine et n'implique pas que l'entreprise ait connaissance directement des contenus diffusés par ses soins. Pour autant, l'esprit de la réglementation sur la responsabilité des intermédiaires est détourné, puisqu'elle visait à l'origine à éviter justement que les plateformes soient contraintes à exercer une surveillance généralisée.

---

[6] LESSIG L. (2000), "Code is Law. On Liberty In Cyberspace. Harvard Magazine" http://harvardmagazine.com/2000/01/code-is-law-html

[7] "Viacom International Inc. v. YouTube, Inc." https://en.wikipedia.org/wiki/Viacom_International_Inc._v._YouTube,_Inc.

[8] Voir LANGLAIS P.-C. (2016), « Comment fonctionne Content ID ? » https://scoms.hypotheses.org/709

Le paradoxe est donc qu'un acteur comme YouTube a volontairement ouvert la voie au filtrage automatique, pour faire face à la massification de son service, mais sans y être contraint par la loi. En montrant qu'un algorithme était capable d'interpréter et d'appliquer une législation comme celle du droit d'auteur, Content ID a provoqué un renversement de *Code Is Law* en *Law is Code* : la mise en œuvre du droit peut être « codée » et automatisée *via* des algorithmes. Le précédent de YouTube a ensuite fait tache d'huile et les dispositifs de filtrage se sont étendus progressivement à d'autres services en ligne, sur une même base volontaire pour Dailymotion, Instagram, Facebook ou même Dropbox, ou sous la pression des ayants droit, comme sur la plateforme SoundCloud, qui a longtemps cherché à résister à leur introduction[9].

Ce que le filtrage du web fait aux libertés

Le filtrage n'est pas une technique anodine et il provoque une fragilisation des droits sur Internet, au premier rang desquels la liberté d'expression. Dans un système classique de *notice and take down*, c'est normalement à celui qui estime qu'un contenu est illicite de procéder à un signalement et l'intermédiaire technique doit apprécier, au terme d'une analyse réalisée par un humain, si cette plainte est fondée. La réglementation prévoit normalement que l'hébergeur peut refuser de faire droit à la demande de retrait si le contenu en cause n'est pas « manifestement illicite[10] ». En cas de désaccord, c'est normalement au juge d'avoir le dernier mot dans le cadre d'un procès contradictoire, qui devra être intenté par la personne à l'origine du signalement.

Le filtrage modifie ce cheminement dans la mesure où c'est la plateforme qui intervient proactivement pour retirer des contenus jugés non conformes au droit en vertu d'une analyse automatique. La charge de la preuve est alors inversée, puisque c'est à l'utilisateur de contester le retrait s'il estime être dans son bon droit. Les plateformes ont en général instauré des voies de recours internes, mais si l'on prend l'exemple de YouTube, ce sont au final les ayants droit qui ont le dernier mot et décident de maintenir ou lever les sanctions automatiques. L'utilisateur doit alors prendre sur lui de saisir le juge pour faire valoir ses droits, ce qui intervient en pratique extrêmement rarement. C'est la raison pour laquelle ces dispositifs sont souvent accusés de mettre en œuvre des formes de police et de justice privées.

En matière de droit d'auteur, la directive Copyright adoptée par le Parlement européen le 26 mars dernier va encore plus loin, en risquant d'imposer aux plateformes centralisées un filtrage *a priori* des contenus[11]. Jusqu'à présent, un robot

[9] Voir MAUREL L. (2013), « Filtrage : quand SoundCloud joue au Robocopyright. S.I.Lex » : https://scinfolex.com/2013/04/18/filtrage-soundcloud-fait-sa-police-du-copyright/

[10] Pour une application récente, voir cette affaire dans laquelle Twitter a refusé de faire droit à une demande de retrait de vidéos formulée par l'humoriste Gad Elmaleh en invoquant un doute sur leur caractère « manifestement illicite » : PCS Avocat (2019), « Plagiat, contrefaçon et hébergeurs de contenus : Gad Elmaleh contre Twitter »
https://www.pcs-avocat.com/plagiat--contrefacon-et-hebergeurs-de-contenus---gad-elmaleh-contre-twitter_ad240.html

[11] Voir REES M. (2018), « Pourquoi la directive Droit d'auteur peut aboutir à un filtrage de l'upload, Next INpact »

comme Content ID intervenait après que les vidéos étaient postées sur la plateforme, mais ce nouveau texte implique le passage à une logique de *notice and stay down* et le contrôle opéré par la plateforme pourrait ~~s'opérer~~ intervenir en amont au moment du chargement. Il en résulterait une impossibilité de poster les contenus, alors que l'algorithme de YouTube est à l'origine de nombreuses erreurs.

Ces failles dans le fonctionnement du filtrage ne sont pas uniquement des problèmes techniques : elles révèlent aussi une limite sans doute indépassable dans l'application automatisée du droit. La plupart des législations européennes autorisent par exemple la reprise de contenus protégés dans le cadre de parodies sur la base d'une exception au droit d'auteur. Or, Content ID est structurellement incapable de distinguer la simple reprise d'une œuvre d'une parodie, car il faudrait pour cela qu'il soit en mesure de détecter l'humour. L'aveuglement des machines à ces nuances peut avoir des conséquences plus graves, lorsqu'on prétend les employer dans le cadre de la lutte contre la haine en ligne ou contre l'apologie du terrorisme. La définition de l'incitation à la haine – et plus encore celle du terrorisme – reste extrêmement floue dans les textes juridiques, ce qui est susceptible d'entraîner de graves dérives dès lors que l'on sous-traite la régulation à des algorithmes. Il n'est pas étonnant d'ailleurs que, malgré les discours empreints de « solutionnisme » vantant les mérites de l'intelligence artificielle, les grandes plateformes ont en réalité toujours largement recours à des modérateurs humains – souvent implantés dans des pays du Sud et travaillant dans des conditions déplorables – pour procéder à des retraits manuels[12].

Ces insuffisances structurelles n'empêchent pourtant pas les législateurs d'envisager favorablement l'extension du champ d'application du filtrage. C'est le cas par exemple dans le Règlement anti-terroriste actuellement en discussion au niveau de l'Union européenne, qui entend imposer le retrait en une heure de contenus signalés par des autorités judiciaires ou administratives des pays de l'Union. Le texte prévoit que les plateformes devront prendre des mesures proactives pour éviter la circulation des contenus faisant l'apologie du terrorisme, ce qui conduira à la mise en place d'un filtrage dénoncé par les opposants au règlement comme un pas vers une « automatisation de la censure politique[13] ». La même logique se retrouve dans la loi de lutte contre la haine en ligne et le cyber-harcèlement envisagée actuellement par le gouvernement français, qui remet en avant l'idée du filtrage automatique avec les mêmes conséquences prévisibles sur la liberté d'expression[14].

Une menace préoccupante pour les alternatives décentralisées

---

https://www.nextinpact.com/news/107399-pourquoi-directive-droit-dauteur-peut-aboutir-a-filtrage-upload.htm

[12] En ce sens, voir le documentaire suivant : RIESEWIECK R. et BLOCK H. (2018), « Les nettoyeurs du web ».

[13] TRÉGUER F. (2019), « Vers l'automatisation de la censure politique » https://www.laquadrature.net/2019/02/22/vers-lautomatisation-de-la-censure-politique/

[14] La Quadrature du Net (2019), « Mahjoubi et Schiappa croient lutter contre la haine en ligne en méprisant le droit européen » https://www.laquadrature.net/2019/02/14/mahjoubi-et-schiappa-croient-lutter-contre-la-haine-en-meprisant-le-droit-europeen/

De manière ironique, le filtrage pourrait contribuer à renforcer la domination des grandes plateformes décentralisées, bien plus qu'à desserrer leur emprise. Comme nous l'avons montré, c'est l'entreprise Google qui a mis au point les premiers dispositifs de filtrage automatisé pour sa filiale YouTube, en investissant plus de 100 millions de dollars pour le développement de Content ID[15]. Facebook a également consacré des dépenses considérables pour élaborer son propre dispositif de filtrage, présenté comme fonctionnant grâce à l'intelligence artificielle[16].

L'accès à ces technologies représente donc un coût considérable, que seuls les plus grands acteurs sont en mesure de supporter. Si la réglementation imposait le recours à ces mesures techniques de contrôle, elle favoriserait mécaniquement les plus grands acteurs au détriment des plus petits, car seuls les premiers seraient capables de déployer ces dispositifs de filtrage. Pire encore, ce sont des Géants du Web comme Google ou Facebook qui ont développé à ce jour les technologies les plus performantes dans le domaine et ces entreprises seront donc en mesure de vendre leurs solutions à leurs concurrents, de manière à les rendre dépendants de leur système. C'est pourquoi il est fallacieux de présenter le filtrage comme une manière pour l'Europe d'imposer sa souveraineté numérique aux GAFAM, puisqu'au contraire la généralisation de ces dispositifs ne ferait que renforcer la dépendance de l'écosystème numérique aux grands opérateurs américains.

C'est sans doute la raison pour laquelle les GAFAM ne luttent pas contre l'imposition des mesures de filtrage, mais au contraire en sollicitent activement le déploiement. Si Google s'est opposé à la directive sur le droit d'auteur, c'est moins sur la question du filtrage – qu'il a déjà implémenté sur YouTube – que parce que le texte entend lui imposer le paiement de redevances supplémentaires aux ayants droit. Dans une récente tribune parue dans *Le Journal du Dimanche* intitulée « Quatre idées pour réguler Internet », Mark Zuckerberg – le fondateur de Facebook – ne remet pas en cause la logique du filtrage, mais plaide pour la mise en place de voies de recours confiées à des « organismes tiers chargés de définir des standards sur la diffusion des contenus violents et haineux[17] ». L'entreprise a d'ailleurs déjà pris les devants en organisant en son sein une cour d'appel, présentée comme « indépendante », pour examiner les réclamations suite aux retraits de contenus. Cette évolution traduit la volonté d'institutionnaliser la justice privée dont Facebook est devenu l'opérateur, mais sans remettre en question le contournement des tribunaux étatiques qui accompagnent l'extension du filtrage.

Pour Félix Tréguer, ces évolutions sont le signe d'une « fusion État-GAFAM » bien davantage qu'une reprise en main des plateformes par les États souverains[18] :

« Si l'on pense l'État non pas comme un bloc aux contours clairement identifiés (à la manière des juristes) mais davantage comme un ensemble de pratiques et une rationalité que Michel Foucault désignait comme la "gouvernementalité", alors il est

[15] Voir Wikipédia "Content ID"
https://en.wikipedia.org/wiki/Content_ID_(algorithm)
[16] SOMINITE T. (2018), "AI has started cleaning up Facebook, but can it finish ?"
https://www.wired.com/story/ai-has-started-cleaning-facebook-can-it-finish/
[17] ZUCKERBERG M. (2019), « Quatre idées pour réguler Internet », *Le Journal du Dimanche* :
https://www.lejdd.fr/Medias/exclusif-mark-zuckerberg-quatre-idees-pour-reguler-internet-3883274
[18] TRÉGUER F., *op. cit.*

clair que ce que ces évolutions donnent à voir, c'est l'incorporation de ces acteurs privés à l'État ; c'est la cooptation de leurs infrastructures et la diffusion de leurs savoir-faire dans le traitement et l'analyse de masses de données désormais cruciales dans les formes contemporaines de gouvernement. C'est donc une fusion qui s'opère sous nos yeux, bien plus qu'une concurrence entre les États et les GAFAM qui chercheraient à se substituer aux gouvernements. »

Pourtant, la « plateformisation » est loin d'être une fatalité et une décentralisation des usages reste possible, notamment grâce aux développements récents des standards d'interopérablité[19]. Des services comme Mastodon ou Peertube – alternatives libres et ouvertes respectivement à Twitter et à Youtube – montrent qu'il est possible sur le plan technique d'offrir des niveaux de services comparables à ceux des grandes plateformes sans passer par une centralisation des contenus. Ces alternatives fonctionnent sur la base d'une multitude d'instances fédérées entre elles, pouvant être hébergées par une diversité d'acteurs (individus, associations, administrations, entreprises) et capables de communiquer entre elles. Avec son initiative CHATONS[20], l'association Framasoft a mis en place une coalition d'acteurs capables de soutenir ces services partagés, dans le respect d'une Charte fixant de grands principes, comme la non-exploitation des données personnelles ou le recours aux logiciels libres.

Ces perspectives montrent la manière dont un Internet Libre et Ouvert pourrait être refondé, en tournant la page de la « plateformisation » qui aura marqué le début du XXI^{e} siècle. C'est précisément cette alternative que le filtrage généralisé pourrait tuer dans l'œuf, car il est clair que ces petits acteurs fédérés sont dans l'incapacité de déployer des mesures de filtrage. À défaut, ils se trouveraient alors exposés à une responsabilité de plein fouet, insoutenable pour eux en raison des risques encourus.

Si les pouvoirs publics veulent réellement lutter contre les GAFAM, il y a bien d'autres leviers législatifs à activer : protection des données personnelles, lutte contre les abus de position dominante ou fiscalité des entreprises du numérique. L'obsession du législateur pour le filtrage est avant tout le signe d'une résignation à l'existence même de la domination des grandes plateformes : une façon de lutter contre les conséquences de la domination de ces acteurs sans s'attaquer aux causes réelles du problème.

[19] Voir en particulier le standard ActivityPub élaboré par le W3C pour l'interopérabilité du « Web Social »
https://www.w3.org/TR/activitypub/
[20] Collectifs des Hébergeurs Alternatifs Transparents Ouverts Neutres et Solidaires
https://chatons.org/