

L'impact des algorithmes pour le filtrage ou la modération de contenus en ligne

«Filtres de téléchargement»¹

RÉSUMÉ

L'étude porte sur le filtrage automatisé des contenus en ligne. Elle aborde le filtrage automatisé, un aspect de la modération des contenus créés par les utilisateurs. Elle présente les technologies de filtrage actuellement déployées pour traiter différents types de médias, texte, images ou vidéos. Elle examine les principales questions essentielles du cadre juridique actuel et formule des propositions de réglementation dans le contexte d'une future **législation de l'Union sur les services numériques**.

Contexte

L'étude porte sur le filtrage automatisé des contenus en ligne. Elle aborde le filtrage automatisé, un aspect de la modération des contenus créés par les utilisateurs. Elle présente les technologies de filtrage actuellement déployées pour traiter différents types de médias, texte, images ou vidéos. Elle examine les principales questions essentielles du cadre juridique actuel et formule des propositions de réglementation dans le contexte d'une future **législation de l'Union sur les services numériques**.

Vieille de 20 ans, la **directive sur le commerce électronique** a joué un rôle positif dans le développement de l'économie numérique et de l'environnement de l'information en ligne, mais elle est désormais appliquée dans un contexte technologique, économique et social qui a très largement évolué: les entreprises de l'internet sont devenues des acteurs mondiaux, avec d'énormes ressources financières et technologiques, et l'accès à l'information ainsi que les interactions sociales se déroulent aujourd'hui en ligne.

Le **projet de législation sur les services numériques** devrait tenir compte du nouveau contexte technologique, économique et social. Une question essentielle concerne la réglementation des services numériques, en particulier ceux qui relèvent de la catégorie des plateformes en ligne, à savoir les services numériques dont l'objectif est de faciliter l'interaction numérique entre les utilisateurs

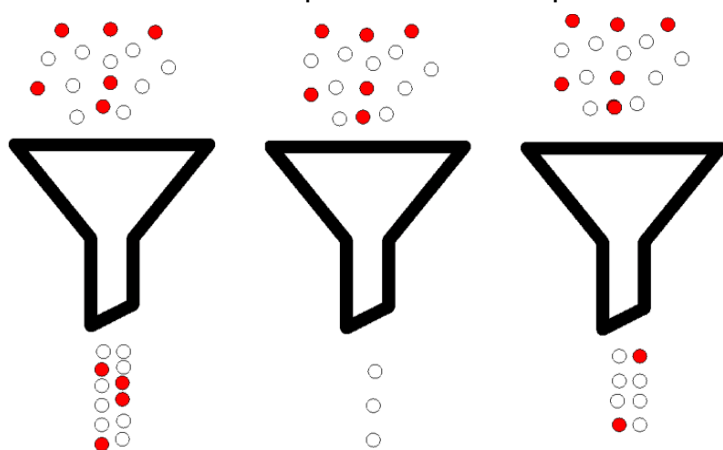
¹ Etude complète en anglais : [https://www.europarl.europa.eu/ReqData/etudes/STUD/2020/657101/IPOL_STU\(2020\)657101_EN.pdf](https://www.europarl.europa.eu/ReqData/etudes/STUD/2020/657101/IPOL_STU(2020)657101_EN.pdf)



(entreprises ou particuliers). En particulier, les plateformes de contenu créé par les utilisateurs permettent aux utilisateurs de s'exprimer, de créer ou de transmettre des informations et des créations culturelles ou bien d'y accéder, et de mener des interactions sociales, mais elles ouvrent également la porte à des comportements préjudiciables: les incivilités, l'agressivité dans les échanges personnels, la désinformation dans la sphère publique, le sectarisme et la polarisation en politique, ainsi que les illégalités, l'exploitation et la manipulation.

Pour prévenir les comportements illicites et préjudiciables en ligne, il faut une *modération*, à savoir une gestion active des plateformes visant à garantir, dans la mesure où cela est raisonnablement possible, des interactions productives ainsi qu'un comportement prosocial et licite des utilisateurs. La modération doit faciliter la coopération et prévenir les abus, sans quoi les communautés en ligne ont tendance à devenir dysfonctionnelles, à être victimes de «pollution» (ou spam), etc.

L'étude porte sur un aspect essentiel de la modération d'aujourd'hui, à savoir le filtrage automatisé destiné à classer et, par la suite, à supprimer ou à exclure les contenus créés par les utilisateurs. Il faut des filtres automatisés pour surveiller la quantité considérable de contenus téléchargés en ligne et



détecter les contenus (potentiellement) illicites et abusifs. Toutefois, l'utilisation de ces filtres présente des risques, car elle peut conduire à l'exclusion de contenus précieux; elle peut aussi affecter la liberté d'expression, l'accès à l'information et le dialogue démocratique.

La directive sur le commerce électronique préconise une démarche pratique de la modération, et en particulier du filtrage: elle protège les prestataires de la responsabilité de contenus créés

illégalement par les utilisateurs, tout en interdisant aux États d'imposer aux prestataires une obligation générale de surveillance. Cette approche a été remise en question ces dernières années, étant donné que de nouvelles technologies de filtrage ont vu le jour.

Objectif

L'objectif de cette étude était de proposer une analyse approfondie des questions d'ordres technologique et juridique liées à la réglementation du filtrage en ligne. Elle présente les technologies de filtrage actuellement déployées pour traiter différents types de médias, texte, images ou vidéos. Examen de leurs forces et faiblesses:

- La recherche, le hachage et le relevé d'empreintes digitales des métadonnées sont utilisés afin d'identifier de manière fiable des copies d'œuvres numériques connues;
- La liste noire est utilisée pour trouver des expressions indésirables;
- Des techniques avancées de traitement du langage naturel sont utilisées pour en décoder la signification et le contexte;
- Plusieurs techniques, souvent fondées sur l'IA, sont utilisées pour identifier des images, ou des combinaisons de texte et d'images indésirables.

L'étude porte sur la précision des systèmes de filtrage:

- Ces systèmes reposent sur des méthodes fondées sur les probabilités, de sorte que toutes les erreurs ne peuvent être évitées.

- Il est difficile de déterminer ce qui peut être considéré comme une réponse correcte, étant donné que la «réalité de terrain» dépend des évaluations humaines sous-jacentes.

Compte tenu du faible niveau de fiabilité et de la subjectivité des filtres automatisés, leur fonctionnement devrait être contrôlable. L'étude examine les méthodes permettant de garantir la transparence et offrant des procédures de recours:

- La transparence prévoit que les personnes concernées et la société dans son ensemble soient informées du processus de filtrage;
- Il faut des procédures de recours et d'indemnisation pour permettre aux utilisateurs de porter plainte et pour remédier aux défaillances.

L'étude examine également l'accessibilité et les coûts des technologies de filtrage. Ces technologies sont principalement utilisées par de grands acteurs, qui développent souvent de puissants systèmes internes, mais certaines solutions sont également accessibles aux petites entreprises.

Enfin, l'étude porte sur la réglementation du filtrage. Elle formule tout d'abord plusieurs prémisses qu'il convient de prendre en compte:

- Le filtrage automatisé ne devrait pas être découragé, car il s'agit d'un élément essentiel d'une modération en ligne efficace;
- Le filtrage n'est pas suffisamment fiable, même lorsqu'il est développé et mis en œuvre de bonne foi pour lutter contre les contenus abusifs;
- Le filtrage prosocial autorisé ne se limite pas aux contenus illicites; il peut, à juste titre, viser tout type de contenu contestable au sein d'une communauté en ligne particulière;
- Bien que le filtrage justifié ne se limite pas aux contenus illicites, il ne devrait pas reposer sur des choix aléatoires ou arbitraires des propriétaires et des modérateurs de plateformes;
- Aucun ensemble de règles d'application purement mécanique ne peut déterminer avec certitude ce qui est juridiquement acceptable ou inacceptable sur une plateforme.
- Des normes de diligence raisonnable peuvent être utilisées pour évaluer le comportement des prestataires;
- Une réglementation inappropriée peut donner lieu à un filtrage excessif ou insuffisant, de sorte que des contenus précieux se trouvent dès lors inaccessibles ou que des contenus préjudiciables restent en ligne;
- Une incertitude plane sur les obligations de contrôle du contenu interdites par le droit de l'Union.

À ce titre, certaines options stratégiques sont présentées:

- Il conviendrait d'envisager une mise à jour des principes généraux relatifs à l'immunité des prestataires tels qu'ils figurent dans la directive sur le commerce électronique. L'on pourrait également préciser dans quelle mesure le droit de l'Union protège les prestataires contre l'imposition d'obligations légales de surveillance, de suppression ou de blocage de contenus.
- Il convient également de préciser qu'un engagement à assurer une modération, et en particulier à filtrer les contenus illicites ou abusifs, ne devrait avoir aucun lien avec les immunités ou autres avantages accordés aux fournisseurs.
- Il convient de prévoir des recours procéduraux contre les suppressions en ligne, afin que ceux qui téléchargent des contenus externalisés filtrés le fassent en toute connaissance de cause,

reçoivent des explications et puissent contester ces décisions en obtenant des réponses de la part de personnes «humaines».

- Les pouvoirs publics devraient se pencher sur le filtrage en ligne, au moyen de règlements spécifiques, ainsi que sur les décisions prises dans des affaires controversées. Elles pourraient être coordonnées par des organes existants de l’Union ou par une autorité nouvellement créée.
- Il convient par ailleurs de soutenir les petites et moyennes entreprises. Elles ne devraient être considérées comme responsables uniquement de l’absence d’adoption de mesures (y compris de filtrage) qui sont de leur ressort, tant sur le plan technologique que sur le plan économique.
- Il convient de mettre sur pied une approche à l’échelle de l’Union pour réglementer le filtrage, tout en tenant compte de la diversité nationale dans l’évaluation de la licéité des contenus en ligne.
- Mais il convient avant tout de mener sur la modération en ligne, et en particulier sur le filtrage, un large débat qui associe non seulement les autorités politiques et administratives, mais aussi la société civile et le monde universitaire.

Clause de non-responsabilité et droits d’auteur. Les opinions exprimées dans le présent document sont celles des auteurs et ne reflètent pas nécessairement la position officielle du Parlement européen. La reproduction et la traduction sont autorisées, sauf à des fins commerciales, moyennant mention de la source, information préalable du Parlement européen et transmission d’un exemplaire à celui-ci. © Union européenne, 2020.

Auteurs externes: Prof. Giovanni Sartor, Institut universitaire européen de Florence.

Coécrit par le professeur Giovanni Sartor et le docteur Andrea Loreggia, sous la supervision du professeur Sartor.

Administrateur responsable de recherche: Udo BUX

Assistance éditoriale: Monika LAZARUK

Contact: poldep-citizens@europarl.europa.eu

Ce document est disponible en ligne à l’adresse suivante: <https://www.europarl.europa.eu/committees/fr/supporting-analyses/sa-highlights>

PE 657.101

Version imprimée ISBN 978-92-846-7188-5 | doi:10.2861/150225 | QA-02-20-825-FR-C

PDF ISBN 978-92-846-7183-0 | doi:10.2861/628984 | QA-02-20-825-FR-N