

La régulation de la modération des contenus en ligne, une affaire de droits humains¹ ?

PAR

Alejandra MICHEL

*Chercheuse senior, Responsable de l'unité Médias au CRIDS/NaDI
Maître de conférences à l'UNamur*

Résumé

La régulation de la modération des contenus en ligne est une problématique complexe qui implique d'équilibrer des droits et des intérêts multiples. Pour contrer le pouvoir considérable des plateformes sur le débat public, le législateur multiplie leurs obligations. La responsabilisation des acteurs du numérique par rapport à l'expression des internautes n'est toutefois pas sans risque pour les libertés d'expression et d'information. Face à cette inflation réglementaire, des voix s'élèvent pour replacer les standards internationaux (et régionaux) des droits humains au cœur des pratiques de modération des contenus.

Abstract

Regulating online content moderation is a complex issue that involves balancing multiple rights and interests. To counter the considerable power of platforms on public debate, legislators are multiplying their obligations. However, the accountability of digital actors with respect to the expression of Internet users is not without risk for freedoms of expression and information. Faced with this regulatory inflation, voices are being raised to put international (and regional)

¹ Cette publication a été réalisée avec le soutien financier du programme de recherche et d'innovation Horizon 2020 dans le cadre de la convention de subvention n° 958339 (DENiM). La publication ne reflète que l'opinion de son auteure et la Commission européenne ne peut être tenue responsable de l'usage qui en serait fait. L'auteure tient à remercier Cécile de Terwangne et Amélie Lachapelle pour leur relecture et leurs conseils avisés. La rédaction a été arrêtée en avril 2022, avant l'adoption définitive du *Digital Services Act* (ci-après « DSA »).

human rights standards back at the heart of content moderation practices.

Introduction

La modération des contenus en ligne est affaire de droits humains. Se pencher sur cette problématique nécessite forcément d'avoir à tenir compte de droits et d'intérêts divers, ce qui la rend à la fois complexe et passionnante. L'on pense d'emblée à la liberté d'expression de l'auteur d'un contenu censuré, au droit à l'information de l'ensemble des internautes qui souhaitent accéder aux contenus faisant l'objet de mesures de modération, ainsi qu'à la liberté d'entreprise des acteurs du numérique qui se voient parfois imposer des obligations à l'opposé de leurs intérêts économiques. L'on doit également considérer les droits des victimes de contenus illégaux ou dommageables diffusés sur les plateformes : les droits de propriété intellectuelle des titulaires de droits sur des contenus protégés, les droits à la non-discrimination, au respect de la vie privée et à la réputation des personnes, mais aussi la protection des consommateurs. Citons encore le droit à la protection des données des internautes visés par des mesures d'enquête et l'intérêt des États à assurer la sécurité intérieure, l'ordre public et la lutte contre les infractions pénales. Force est donc de constater que les pratiques de modération des contenus s'inscrivent dans un imbroglio de droits humains.

À l'heure actuelle, les acteurs du numérique jouissent d'un pouvoir considérable sur le débat public et leurs pratiques de retrait de contenus font régulièrement l'objet de critiques acerbes. En raison de ces dérives, des voix se sont élevées pour réformer le cadre réglementaire de la modération des contenus et pour accentuer les obligations et la responsabilité des géants du numérique. Néanmoins, en la matière, l'équilibre à respecter entre les droits et les intérêts en présence est loin d'être une mince affaire et la réflexion sur la mise en conformité des pratiques de modération avec les principes clés des droits humains n'en est qu'à ses balbutiements.

Après avoir mis en exergue la position des plateformes au regard de l'expression des internautes (I), la présente contribution s'intéresse au renforcement du rôle que les législateurs confèrent à ces dernières dans le cadre de la lutte contre les contenus illégaux (II). Nous exposons ensuite les risques pour le droit fondamental à la liberté d'expression suscités par la responsabilisation accrue des acteurs du numérique et par la multiplication des obligations de modération des contenus qui leur sont imposées tant par les instruments hori-

zontaux que verticaux (III). Enfin, avant de conclure, nous mettons en lumière les appels engagés d'acteurs de la société civile et d'académiques visant à replacer les standards des droits humains au cœur de toute action de régulation de l'expression et de la modération des contenus (IV).

I. Les plateformes, promotrices ou sanctionnatrices de la liberté d'expression ?

Depuis une dizaine d'années, le succès des plateformes ne cesse de s'accroître, attirant vers leurs services des millions, voire des milliards, d'utilisateurs à travers le monde². Les premiers «espaces personnels» et «forums de discussion», au *design* trivial, laissent désormais la place à des plateformes aux fonctionnalités extrêmement sophistiquées. Ces dernières se multiplient en façonnant chacune leurs services pour des utilisations de plus en plus spécifiques. À côté des incontournables réseaux sociaux dédiés à l'échange du quotidien, l'on retrouve des plateformes de partage de vidéos, des réseaux professionnels, des places de marché en ligne ou de vente de seconde main, des services d'échange dédiés aux jeux vidéo ou encore des plateformes de partage de contenus éphémères.

Dans ce contexte, les plateformes constituent un outil largement utilisé tant pour s'exprimer que pour s'informer. Elles deviennent ainsi un lieu privilégié pour le débat public à l'échelle mondiale. Ce constat a d'ailleurs poussé la Cour européenne des droits de l'homme à déclarer que «l'Internet est aujourd'hui devenu l'un des principaux moyens d'exercice par les individus de leur droit à la liberté d'expression et d'information : on y trouve des outils essentiels de participation aux activités et débats relatifs à des questions politiques ou d'intérêt public»³.

² Selon Statista, en janvier 2022, Facebook conservait sa place de *leader* en comptant plus de 2,9 milliards d'utilisateurs actifs par mois. Toujours dans le milliard d'utilisateurs suivaient YouTube (2,5), Instagram (1,4), le réseau social chinois WeChat (1,2) et TikTok (1). L'on retrouvait ensuite Snapchat, Pinterest et Twitter avec respectivement 557, 444 et 436 millions d'utilisateurs actifs par mois. Voy. «Classement des réseaux sociaux les plus populaires dans le monde en janvier 2022, selon le nombre d'utilisateurs actifs», 25 mars 2022, disponible sur <https://fr.statista.com/statistiques/570930/>.

³ Cour eur. dr. h., arrêt *Ahmet Yildirim c. Turquie*, 18 décembre 2012, § 54; arrêt *Cengiz e.a. c. Turquie*, 1^{er} décembre 2015, § 49. Par ailleurs, la Cour a déjà souligné, à plusieurs reprises depuis 2009, l'importance des sites web pour les libertés d'expression et d'information. Voy. Cour eur. dr. h., arrêt *Times Newspapers LTD c. Royaume-Uni (nos 1 et 2)*, 10 mars 2009, § 27. Dans le même sens, voy. notamment Cour eur. dr. h., Gde Ch., arrêt *Magyar Kétfarku Kutya Part c. Hongrie*, 20 janvier 2020, § 87.

Néanmoins, la démultiplication des atteintes aux droits d'autrui constitue le revers de la médaille. Les propos haineux, les insultes, les contenus violents, les brimades, les atteintes à la vie privée et les contrefaçons sont légion sur Internet. Pour enrayer ce phénomène, les plateformes sont amenées à intervenir davantage que par le passé sur le contenu. Entre les actions qu'elles entreprennent volontairement, les requêtes adressées par les États⁴, les injonctions émises par les autorités nationales et les signalements par les utilisateurs et les «signaleurs de confiance», les mesures de modération de contenus s'intensifient au détriment de la libre circulation des idées, des opinions et des informations. Ces inquiétudes sont renforcées par la pression qui est actuellement exercée par les politiques et les initiatives législatives, tant à l'échelon européen⁵ que national⁶, sur les acteurs du numérique dans le cadre de la lutte contre les

⁴ Certains gouvernements exercent des pressions sur les plateformes pour obtenir le retrait de contenus dérangeants ou recourent à des dispositions imprécises pour parvenir aux mêmes fins. À ce propos, voy. Rapport du Rapporteur spécial des Nations Unies sur la promotion et la protection du droit à la liberté d'opinion et d'expression, présenté lors de la 38^e session du Conseil des droits de l'homme, 6 avril 2018, A/HRC/38/35, §§ 13, 19, 20, 23, 39 et 49; N. ELKIN-KOREN et M. PEREL, «Guardian the Guardians: Content Moderation by Online Intermediaries and the Rule of Law», in G. Frosio (éd.), *The Oxford Handbook of Online Intermediary Liability*, Oxford University Press, Oxford, 2020, p. 673.

⁵ Parmi les instruments de l'Union européenne qui imposent des obligations de modération des contenus aux acteurs du numérique, voy. directive (UE) 2018/1808 du Parlement européen et du Conseil du 14 novembre 2018 modifiant la directive 2010/13/UE visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à la fourniture de services de médias audiovisuels (directive «Services de médias audiovisuels»), compte tenu de l'évolution des réalités du marché, *J.O.U.E.*, n° L 303 du 28 novembre 2018, p. 69; directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.U.E.*, n° L 130 du 17 mai 2019, p. 92; règlement (UE) 2021/784 du Parlement européen et du Conseil du 29 avril 2021 relatif à la lutte contre la diffusion des contenus à caractère terroriste en ligne, *J.O.U.E.*, n° L 172 du 17 mai 2021, p. 79; Proposition de règlement du Parlement européen et du Conseil du 15 décembre 2020 relatif à un marché intérieur des services numériques (Législation sur les services numériques) et modifiant la directive 2000/31/CE, COM(2020) 825 final (ci-après «Proposition de règlement DSA»).

⁶ Au sein de l'Union européenne, nous pouvons citer les initiatives allemande et française. En Allemagne, le législateur a adopté en 2017 la *Netzwerkdurchsetzungsgesetz* (plus connue sous l'acronyme «*NetzDG*») destinée à améliorer l'application de dispositions pénales existantes sur les réseaux sociaux. La principale obligation qui leur est imposée consiste en la suppression ou le blocage des contenus manifestement illégaux en 24 heures et des contenus illégaux en sept jours. Cette loi a été vivement critiquée pour les risques qu'elle pose pour la liberté d'expression. Voy. notamment B. HOLZNAGEL, «La loi d'application sur les réseaux – L'approche allemande pour lutter contre les 'fausses nouvelles', la violence et le discours terroriste dans les réseaux sociaux», in Fl. Sauvageau, S. Thibault et P. Trudel (dir.), *Les fausses nouvelles: nouveaux visages, nouveaux*

→

contenus illégaux et dommageables, d'autant qu'il n'est pas toujours aisé de tracer la frontière entre expressions licite et illicite.

Les plateformes s'érigent dès lors en véritables « gatekeepers » de la sphère publique et endosseraient selon certains, au regard des contenus en ligne, à la fois le rôle de législateur qui définit les contenus légitimes, de juge qui détermine la légitimité d'un contenu au cas par cas et d'organe administratif qui exécute les mesures adoptées à l'égard des contenus⁷. En tant que promotrices et sanctionnatrices de l'expression, l'influence et le pouvoir des plateformes sur le débat public sont énormes. En effet, au-delà de la distinction binaire entre soutien et censure de l'expression, la manière dont les acteurs du numérique façonnent leurs services impacte considérablement tant la disponibilité, la visibilité, l'accessibilité, la « trouvabilité » ou encore la prééminence des contenus diffusés par les utilisateurs en ligne⁸. Ce faisant, ces derniers influencent les conditions mêmes dans lesquelles la pensée à la base de l'expression est forgée.

←

défis – Comment déterminer la valeur de l'information dans les sociétés démocratiques?, Presses de l'Université de Laval, Laval, 2018, pp. 197-214; ARTICLE 19, « Germany: The Act to Improve Enforcement of Law in Social Networks », 2017, disponible sur www.article19.org/wp-content/uploads/2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf. Le législateur français a rencontré davantage de difficultés que son homologue allemand dans sa volonté de légiférer. La « loi Avia » envisageait d'imposer des mesures fortes aux plateformes, dont la suppression ou le blocage des contenus manifestement illégaux en 24 heures et des contenus terroristes et pédopornographiques en 1 heure. Toutefois, ces mesures n'ont pas passé le cap du Conseil constitutionnel, qui les a déclarées contraires à la liberté d'expression. Le législateur français a remis le couvert en août 2021 avec la loi confortant le respect des principes de la République, vivement critiquée en raison du fait qu'elle « anticipe » l'adoption du DSA en imposant aux plateformes des obligations très similaires à celles discutées au niveau de l'Union. Voy. la loi n° 2021-1109 du 24 août 2021 confortant le respect des principes de la République, *J.O.R.F.*, 25 août 2021; M. REES, « Le recyclage de la loi Avia n'est pas au goût de la Commission européenne », 16 juin 2021, disponible sur www.nextinpact.com/article/47441/le-recyclage-loi-avia-nest-pas-gout-commission-europeenne. La France a également adopté un cadre légal destiné à lutter contre les pratiques de désinformation en période électorale. Voy. la loi organique française n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information, *J.O.R.F.*, 23 décembre 2018. À ce propos, voy. E. DERIEUX, « Lutte contre les 'fausses informations' en France », in Fl. Sauvageau, S. Thibault et P. Trudel (dir.), *Les fausses nouvelles...*, op. cit., pp. 217-230; M. HANOT et A. MICHEL, « Entre menaces pour la vie en société et risques réglementaires, les fake news : un danger pour la démocratie? », in Y. Pouillet (coord.), *Vie privée, liberté d'expression et démocratie dans la société du numérique : espoirs ou menaces?*, coll. du CRIDS, n° 47, Larcier, Bruxelles, 2020, pp. 178-184.

⁷ N. ELKIN-KOREN et M. PEREL, op. cit., p. 669.

⁸ T. MCGONAGLE, « Free Expression and Internet Intermediaries: The Changing Geometry of European Regulation », in G. Frosio (éd.), *The Oxford Handbook of Intermediary Liability Online*,

→

II. Le rôle des plateformes dans la lutte contre les contenus illégaux : vers un renversement du paradigme ?

Ces dernières années, l'intensification du pouvoir des acteurs du numérique sur les contenus diffusés par les internautes provoque un changement dans le modèle définissant jusqu'alors leur rôle dans la lutte contre les contenus illicites. Leur puissance pousse les législateurs à s'appuyer davantage sur eux que par le passé. Même si le régime d'exemption des intermédiaires établi par la directive 2000/31/CE est toujours largement défendu et maintenu, l'introduction successive de nouvelles dispositions en matière de modération des contenus suscite des inquiétudes pour la stabilité du régime ainsi que pour la liberté d'expression en ligne.

A. De l'exonération de responsabilité des plateformes...

Soucieux de ne pas brider le potentiel offert par les services numériques de favoriser la liberté d'expression, le législateur européen a introduit, en 2000, un régime d'exemption de responsabilité des intermédiaires aux articles 12 à 15 de la directive 2000/31/CE sur le commerce électronique que l'on retrouvera dorénavant dans le DSA⁹. Il assure notamment la protection de la liberté d'expression en empêchant, dans certaines circonstances, que la responsabilité des fournisseurs de services de simple transport, de mise en cache et d'hébergement ne puisse être engagée en raison des contenus diffusés par les utilisateurs par l'entremise de leurs services. L'idée majeure derrière ce régime, perçu depuis son établissement comme l'une des pierres angulaires de l'économie numérique¹⁰,

←

Oxford University Press, Oxford, 2020, pp. 479-480. Dans le même sens, voy. J. BARATA, «La législation sur les services numériques et le pouvoir des médias sociaux de réguler le discours: obligations, responsabilités et protections», in M. Cappello (éd.), *Décrypter la législation sur les services numériques*, coll. Iris Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2021, pp. 8-9.

⁹ Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur («directive sur le commerce électronique»), *J.O.U.E.*, n° L 178 du 17 juillet 2000, p. 1.

¹⁰ Proposition de règlement DSA, préc., p. 3 et cons. 16. Ce régime est d'ailleurs maintenu aux articles 3, 4, 5 et 7 du DSA qui reproduisent les dispositions de la directive 2000/31/CE en y intégrant les précisions apportées par la Cour de justice de l'Union européenne sur les exemptions de responsabilité et sur les contours de l'interdiction des obligations générales de surveillance. Par

→

est d'éviter que les intermédiaires, par crainte d'avoir eux-mêmes à assumer la responsabilité d'une infraction, n'attendent aux contenus et ne modèrent des expressions licites.

Les plateformes rentrent dans la notion d'«hébergeur». À leur égard, le régime d'exemption de responsabilité s'articule autour de deux axes: d'une part, l'exonération de responsabilité dès lors qu'elles n'ont pas connaissance de la présence d'un contenu illicite ou qu'elles agissent promptement pour le supprimer ou en bloquer l'accès dès qu'elles en ont connaissance¹¹ et, d'autre part, l'interdiction de leur imposer une obligation générale de surveillance des contenus ou de recherche active des activités illicites¹². Les hébergeurs ne peuvent bénéficier de ce régime que vis-à-vis des contenus diffusés par leurs utilisateurs et lorsqu'ils agissent de manière totalement neutre en se limitant à un traitement technique et automatique des informations¹³.

B. ... à la responsabilisation accrue pour les contenus diffusés par les utilisateurs

Alors que l'idéologie reposait jusqu'il y a peu sur l'absence de responsabilité des plateformes pour les contenus postés par les utilisateurs, le paradigme semble se renverser au gré des actions de la politique du marché unique numérique de l'Union européenne¹⁴. Les instruments récents, tant horizontaux, verticaux, législatifs, d'autorégulation que de *soft law*, multiplient considérablement les exigences à charge des acteurs du numérique. Ces obligations de diligence raisonnable ont en principale ligne de mire le renforcement de leur

←

rapport à la directive, le DSA crée une exception au régime d'exemption de responsabilité pour les plateformes qui mettent en relation des consommateurs et des professionnels lorsque, en raison de la présentation des informations relatives à la transaction, le consommateur peut présumer que les informations, le produit ou le service sont fournis directement par la plateforme ou par une personne agissant sous son autorité ou son contrôle. Voy. Proposition de règlement DSA, préc., art. 5, § 3, et cons. 16 et 23.

¹¹ Directive 2000/31/CE, préc., art. 14, § 1^{er}, a) et b).

¹² Directive 2000/31/CE, préc., art. 15, § 1^{er}.

¹³ Les hébergeurs perdent le bénéfice de l'exemption de responsabilité dès lors qu'ils endossent un rôle actif leur permettant de contrôler ou de prendre connaissance des informations fournies par les utilisateurs ou que les informations sont directement fournies par eux ou sous leur responsabilité éditoriale. Voy. Proposition de règlement DSA, préc., cons. 18 et 20.

¹⁴ Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des Régions, «Stratégie pour un marché unique numérique en Europe», 6 mai 2015, COM(2015) 192 final.

responsabilité et la nécessité de les contraindre à rendre des comptes quant à leurs activités de modération des contenus.

1. Le renforcement du cadre horizontal de la modération des contenus

Adoptée aux balbutiements du Web 2.0, la directive 2000/31/CE, qui constitue le principal instrument horizontal visant l'ensemble des services de la société de l'information, ne contient en réalité aucune disposition spécifique sur la modération des contenus. Même si ses principes fondateurs demeurent essentiels de nos jours¹⁵, la modernisation, le renforcement et surtout l'harmonisation du cadre horizontal définissant le régime exonératoire des intermédiaires et les obligations qui leur sont imposées paraissaient inévitables pour le législateur de l'Union¹⁶.

Fin 2020, ce dernier a donc mis sur la table la proposition de règlement relatif à un marché intérieur des services numériques, plus connue sous le nom de *Digital Services Act* (ci-après « DSA »). Ce règlement, dont l'adoption est intervenue à l'automne 2022¹⁷, constituera l'instrument clé de l'Union en matière de modération des contenus en ligne en offrant « un cadre horizontal applicable à toutes les catégories de contenus, de produits, de services et d'activités sur

¹⁵ Le maintien des principes de la directive (notamment, la clause de marché intérieur, la liberté d'établissement et le régime d'exonération de responsabilité) a fait l'unanimité lors des discussions autour du DSA. Malgré la nécessité d'actualiser le cadre horizontal applicable aux services intermédiaires, ces principes, qui ont permis de façonner le marché unique numérique et qui ont joué un rôle crucial dans le développement des services numériques, sont toujours valables aujourd'hui. Voy., entre autres, Résolution du Parlement européen, « Législation sur les services numériques : améliorer le fonctionnement du marché unique », 20 octobre 2020, (2020/2018(INL)), principes généraux nos 2, 6 et 57 ; Proposition de règlement DSA, préc., pp. 2, 9 et 10, ainsi que cons. 9 et 16 ; Analyse d'impact accompagnant la Proposition de règlement DSA, 15 décembre 2020, SWD(2020) 348 final, pp. 110, 114, 115 et 150 ; A. DE STREEL et M. HUSOVEC, *The e-commerce Directive as the cornerstone of the Internal Market*, Study for the committee on Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020.

¹⁶ Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, « Façonner l'avenir numérique de l'Europe », 19 février 2020, COM(2020) 67 final, pp. 11, 12 et 14. Le DSA n'abroge toutefois pas la directive 2000/31/CE qui demeure toujours applicable, à l'exception du régime d'exemption de responsabilité des intermédiaires qui est déplacé dans le DSA.

¹⁷ Après un accord politique informel atteint en avril et approuvé par le Comité du marché intérieur du Parlement européen le 16 juin 2022, le DSA a été adopté le 19 octobre 2022 et publié au *Journal officiel de l'Union européenne* le 27 octobre 2022.

les services intermédiaires»¹⁸. Le législateur justifie son adoption principalement à deux égards. Premièrement, le cadre existant se révèle insuffisant pour agir de manière globale contre les contenus illicites dans l'ensemble de l'écosystème numérique : d'une part, les instruments spécifiques visent bien souvent uniquement une catégorie précise de contenus (œuvres protégées, contenus audiovisuels, contenus à caractère terroriste ou pédopornographique...) ou de services (plateformes de partage de vidéos, fournisseurs de services de partage de contenus en ligne...); d'autre part, le cadre actuel n'évoque que des règles de base en matière de transparence et de responsabilité, n'établit pas de véritables garanties procédurales entourant le retrait ou le blocage de contenus et contient des lacunes en termes d'*enforcement* (mécanisme de surveillance et de contrôle limité)¹⁹. Deuxièmement, dans une optique d'harmonisation, une action au niveau de l'Union européenne s'avère nécessaire pour pallier la fragmentation des règles relatives à la modération des contenus générée par les initiatives nationales. En effet, les législateurs nationaux s'engagent de plus en plus dans la voie de la régulation de l'expression en ligne²⁰, ce qui renforce malheureusement les disparités et l'incohérence et génère des obstacles à la fourniture de services d'intermédiation sur le territoire de l'Union²¹.

Dans l'attente de la version consolidée, l'objectif de la présente contribution n'est pas de livrer une analyse substantielle du DSA²². L'un des axes majeurs de cet instrument réside dans l'imposition d'une série d'obligations de diligence pour la fourniture de services intermédiaires, selon une «approche graduée» qui prend en compte tant le risque engendré que la taille et les moyens de l'acteur concerné. Alors qu'un socle d'obligations minimales s'applique à la totalité des intermédiaires, des obligations additionnelles sont tour à tour imposées aux hébergeurs, aux plateformes en ligne et aux très grandes plate-

¹⁸ Voy. Proposition de règlement DSA, préc., pp. 4 et 7. Le DSA vient compléter la directive 2000/31/CE en visant uniquement les fournisseurs de services intermédiaires, peu importe leur lieu d'établissement, dès lors qu'ils fournissent leurs services au sein de l'Union. Il s'agit en définitive des acteurs dont l'utilisation biaisée de leurs services pave la voie aux contenus illicites et qui sont par ailleurs en mesure d'agir pour aider à les combattre.

¹⁹ Proposition de règlement DSA, préc., pp. 4, 5 et 8.

²⁰ Voy. *supra*, note n° 6, sur les initiatives allemande et française.

²¹ Proposition de règlement DSA, préc., pp. 2, 6 et 7, et cons. 2, 4, 34 et 41.

²² Nous relevons néanmoins, dans le cadre de cette contribution, certaines des dispositions qui nous paraissent présenter un risque au regard de la liberté d'expression en ligne. Voy. *infra*, point III.

formes en ligne²³. Parmi les principales obligations du DSA, l'on peut citer l'inclusion dans les conditions générales d'informations relatives aux restrictions à l'utilisation des services (y compris, quant aux politiques, aux procédures, aux mesures et aux outils utilisés pour la suppression ou le blocage de contenus) ainsi que l'application diligente, objective et proportionnée de ces conditions générales²⁴, l'instauration d'un «mécanisme de notification et d'action» permettant le signalement des contenus illicites²⁵, le renforcement des voies de recours pour contester les mesures de modération avec l'élaboration d'un système interne de traitement des réclamations²⁶ et la possibilité de faire appel au règlement extrajudiciaire des litiges²⁷, le renforcement de la transparence avec notamment la publication de rapports sur les activités de modération de contenus et de versions anonymisées des décisions adoptées²⁸, ainsi qu'une série d'obligations spécifiques relatives aux comportements abusifs²⁹, à la notification des soupçons d'infractions pénales³⁰, à la vente en ligne³¹, aux systèmes de recommandation³² ou encore à la publicité en ligne³³.

²³ Pour éviter d'imposer aux petits acteurs une charge excessive susceptible d'entraver leur liberté d'entreprendre, le législateur exempte les micros et les petites entreprises de la majorité des exigences prévues par le DSA. Voy. notamment Proposition de règlement DSA, préc., art. 16 et cons. 43.

²⁴ Proposition de règlement DSA, préc., art. 12 et cons. 38. Cette obligation s'appliquera à l'ensemble des fournisseurs de services intermédiaires.

²⁵ Proposition de règlement DSA, préc., art. 14 et cons. 40 et 41. Cette obligation s'appliquera aux hébergeurs. Le DSA détermine avec précision les mentions minimales qui doivent être exigées lors d'un signalement ainsi que les conditions procédurales applicables au traitement des signalements.

²⁶ Proposition de règlement DSA, préc., art. 17 et cons. 44. Cette obligation s'imposera aux plateformes en ligne.

²⁷ Proposition de règlement DSA, préc., art. 18 et cons. 44. Cette obligation s'imposera aux plateformes en ligne.

²⁸ La transparence constitue l'un des piliers du DSA. Les exigences de transparence portent sur des aspects variés et leur nombre dépend du type d'acteur. L'on retrouve des obligations relatives aux conditions générales, aux décisions adoptées, notamment en termes de modération de contenus ou de traitement des réclamations, aux voies de recours disponibles, au coût du règlement extrajudiciaire des litiges, à l'utilisation de moyens automatisés pour la prise de décision, au nombre d'utilisateurs et au point de contact ou au représentant légal. Voy. Proposition de règlement DSA, préc., art. 12, § 1^{er}, 13, 14, §§ 5 et 6, 15, 17, § 4, 18, §§ 3 et 5, 23, §§ 2 et 3, 24, 30 et 33.

²⁹ Proposition de règlement DSA, préc., art. 20.

³⁰ Proposition de règlement DSA, préc., art. 21.

³¹ Proposition de règlement DSA, préc., art. 22.

³² Proposition de règlement DSA, préc., art. 29.

³³ Proposition de règlement DSA, préc., art. 24 et 30.

Autre pan du cadre horizontal de la modération des contenus, la directive 2018/1808 modifie la directive 2010/13/UE sur les services de médias audiovisuels en introduisant des exigences spécifiques à charge des fournisseurs de services de partage de vidéos³⁴. Ces derniers doivent prendre toutes les mesures appropriées³⁵ pour protéger, d'une part, les personnes mineures de tous les contenus susceptibles de nuire à leur épanouissement physique, mental ou moral et, d'autre part, l'ensemble des utilisateurs des contenus terroristes et pédopornographiques, des discours de haine racistes ou xénophobes et des incitations à la violence ou à la haine visant un groupe de personnes ou un membre d'un groupe sur la base d'un motif de discrimination³⁶.

2. La multiplication des instruments verticaux de modération des contenus

Récemment, à côté des instruments horizontaux, le cadre vertical de la modération des contenus a également fortement évolué. Généralement, les instruments verticaux prévoient des dispositions spécifiques de modération applicables à une catégorie d'acteurs déterminée, pour un type particulier de contenu illégal.

Pour rappel, au niveau du droit de l'Union, quatre catégories de contenus illégaux font consensus : les contenus relatifs aux abus sexuels sur les enfants³⁷,

³⁴ Directive (UE) 2018/1808, préc., art. 28ter, § 1^{er}.

³⁵ Il s'agit, selon les besoins, de l'insertion de ces restrictions et des éventuelles mesures de modération de contenus dans les conditions générales, de la mise en place d'un système de signalement des contenus dommageables pour les mineurs et illégaux pour l'ensemble des utilisateurs, de l'établissement d'un système de traitement des signalements (*whistleblowing*), de la possibilité d'introduire des réclamations, de la vérification de l'âge des utilisateurs pour l'accès aux contenus dommageables, de système de contrôles parentaux ou encore de l'éducation aux médias et de la sensibilisation. Voy. directive (UE) 2018/1808, préc., art. 28ter, § 3, sous a), d), e), f), h), i) et j).

³⁶ L'article 28ter vise les motifs de discrimination de l'article 21 de la Charte des droits fondamentaux de l'Union européenne: la discrimination fondée sur le sexe, la race, la couleur, les origines ethniques ou sociales, les caractéristiques génétiques, la langue, la religion ou les convictions, les opinions politiques ou toute autre opinion, l'appartenance à une minorité nationale, la fortune, la naissance, le handicap, l'âge ou l'orientation sexuelle.

³⁷ Directive 2011/93/UE du Parlement européen et du Conseil du 13 décembre 2011 relative à la lutte contre les abus sexuels et l'exploitation sexuelle des enfants, ainsi que la pédopornographie et remplaçant la décision-cadre 2004/68/JAI du Conseil, *J.O.U.E.*, n° L 335 du 17 décembre 2011, p. 1.

les discours de haine racistes ou xénophobes³⁸, les contenus terroristes³⁹ et les infractions aux droits de propriété intellectuelle⁴⁰. L'on peut ajouter une cinquième catégorie de contenus illégaux à cette énumération : les dispositions en matière de protection du consommateur dont l'infraction se matérialise par le biais d'un contenu (par exemple, les pratiques commerciales déloyales et les clauses abusives)⁴¹. Au-delà de ces catégories, il n'existe aucun consensus au sein de l'Union sur d'autres contenus illégaux. Le caractère illégal, dommageable ou légal des autres contenus (discours de haine autres que racistes et xénophobes, désinformation, propos controversés...) varie d'une législation nationale à l'autre⁴². Cette précision est importante et a un impact considérable sur la détermination des mesures de modération, puisque seuls les contenus illégaux sont purement et simplement exclus de la protection par la liberté d'expression⁴³.

³⁸ Décision-cadre 2008/913/JAI du Conseil du 28 novembre 2008 sur la lutte contre certaines formes et manifestations de racisme et de xénophobie au moyen du droit pénal, *J.O.U.E.*, n° L 328 du 6 décembre 2008, p. 55. La lutte contre les discours de haine illégaux pourrait prochainement évoluer au niveau du droit de l'Union. En effet, la Commission a dévoilé une communication « Une Europe plus inclusive et plus protectrice : extension de la liste des infractions de l'UE aux discours de haine et aux crimes de haine » dans laquelle elle souhaite aller plus loin que l'actuelle décision-cadre 2008/913/JAI et ainsi rajouter, au rang des discours de haine illégaux, d'autres mobiles discriminatoires que le racisme et la xénophobie comme, par exemple, le sexe, l'orientation sexuelle, l'âge et le handicap. Voy. Communication de la Commission au Parlement européen et au Conseil, « Une Europe plus inclusive et plus protectrice : extension de la liste des infractions de l'UE aux discours de haine et aux crimes de haine », 9 décembre 2021, COM(2021) 777 final.

³⁹ Directive (UE) 2017/541 du Parlement européen et du Conseil du 15 mars 2017 relative à la lutte contre le terrorisme et remplaçant la décision-cadre 2002/475/JAI du Conseil et modifiant la décision 2005/671/JAI du Conseil, *J.O.U.E.*, n° L 88 du 31 mars 2017, p. 6.

⁴⁰ Les contenus peuvent être protégés par différents droits de propriété intellectuelle (droit d'auteur, droit des marques, droit *sui generis* des bases de données...) et leur utilisation en ligne peut entraîner des violations de ces droits. Voy., parmi d'autres, directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information, *J.O.U.E.*, n° L 167 du 22 juin 2001, p. 10.

⁴¹ Directive 93/13/CEE du Conseil du 5 avril 1993 concernant les clauses abusives dans les contrats conclus avec les consommateurs, *J.O.U.E.*, n° L 95 du 21 avril 1993, p. 29; directive 2005/29/CE du Parlement européen et du Conseil du 11 mai 2005 relative aux pratiques commerciales déloyales des entreprises vis-à-vis des consommateurs dans le marché intérieur et modifiant la directive 84/450/CEE du Conseil et les directives 97/7/CE, 98/27/CE et 2002/65/CE du Parlement européen et du Conseil et le règlement (CE) n° 2006/2004 du Parlement européen et du Conseil, *J.O.U.E.*, n° L 149 du 11 juin 2005, p. 22.

⁴² A. DE STREEL *et al.*, *Online Platforms' Moderation of Illegal Content Online*, Study for the committee on Internal Market and Consumer Protection, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg, 2020, p. 16.

⁴³ La Commission européenne considère que les contenus illégaux et les contenus dommageables soulèvent des problématiques radicalement différentes et nécessitent alors des solutions,

→

L'objet de la présente contribution n'est pas d'expliciter l'ensemble des obligations de modération des contenus illégaux.

Contentons-nous de citer les principaux instruments verticaux au sein de l'Union.

Premièrement, pour les contenus pédopornographiques, la directive 2011/93/UE impose aux États membres de prendre les mesures nécessaires pour les retirer ou en bloquer l'accès efficacement⁴⁴.

Deuxièmement, la même exigence est prévue pour les contenus terroristes par la directive 2017/541⁴⁵. En 2021, le règlement 2021/784 a ajouté des obligations de modération spécifiques pour les contenus à caractère terroriste en ligne. Cet instrument détermine à la fois des mesures que les États membres doivent adopter pour combattre la diffusion de contenus terroristes (notamment, la possibilité d'émettre des injonctions de retrait), ainsi que des obligations de diligence à charge des fournisseurs de services d'hébergement. Sur ce second plan, les principales obligations imposées aux hébergeurs consistent, d'un côté, en la suppression ou le blocage d'accès aux contenus terroristes endéans un délai maximal d'une heure à partir de la réception d'une injonction et, de l'autre côté, en la transparence par rapport aux politiques de lutte contre les contenus terroristes et aux mesures de modération exécutées⁴⁶.

Troisièmement, pour les discours de haine racistes ou xénophobes, la décision-cadre 2008/913/JAI impose aux États membres de les poursuivre mais n'explique aucune mesure spécifique⁴⁷.

←

qu'elles soient légales ou techniques, très différentes. Voy. Communication de la Commission au Conseil, au Parlement européen, au Comité économique et social européen et au Comité des régions, «Contenu illégal et préjudiciable sur Internet», 16 octobre 1996, COM(96) 487 final, p. 10; Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, «Lutter contre la désinformation en ligne: une approche européenne», 26 avril 2018, COM(2018) 236 final, p. 4.

⁴⁴ Directive 2011/93/UE, préc., art. 3, 4 et 5.

⁴⁵ Directive (UE) 2017/541, préc.

⁴⁶ Règlement (UE) 2021/784, préc., art. 3, § 3, 7 et 11. Pour un commentaire, voy. E. DELHAISE, «Retrait des contenus terroristes en ligne: l'Union européenne lutte contre la propagande terroriste visuelle», *R.D.T.I.*, n° 85/2021, pp. 29-50.

⁴⁷ Décision-cadre 2008/913/JAI, préc. Le manque d'effectivité de cet instrument est partiellement pallié par des initiatives d'autorégulation, principalement par le Code de conduite visant à combattre les discours de haine illégaux en ligne qui liste des mesures spécifiques de modération. Nous verrons toutefois que certaines suscitent des craintes pour la liberté d'expression en promouvant à l'excès les mesures proactives. Voy. *infra*, point III.

Quatrièmement et enfin, le très débattu⁴⁸ article 17 de la directive 2019/790 impose en principe aux fournisseurs de services de partage de contenus en ligne d'obtenir l'autorisation des titulaires de droit pour les actes de communication au public qu'ils effectuent en donnant accès aux œuvres protégées téléversées par leurs utilisateurs sur leurs services⁴⁹. En l'absence d'une telle autorisation, ces acteurs engageront leur responsabilité pour la communication illégale d'œuvres protégées, sauf s'ils ont (i) fourni leurs meilleurs efforts pour obtenir une autorisation, (ii) fourni leurs meilleurs efforts pour garantir l'indisponibilité des contenus protégés pour lesquels les titulaires de droit ont apporté les précisions nécessaires et (iii) ont agi promptement pour supprimer ou pour bloquer l'accès aux contenus litigieux signalés par les titulaires de droit et ont fourni leurs meilleurs efforts pour empêcher que ces contenus illégaux ne soient remis en ligne dans le futur⁵⁰.

III. Un renversement de paradigme dangereux pour la liberté d'expression en ligne

Les récentes initiatives en matière de modération, qu'elles proviennent des législateurs ou du secteur lui-même, amènent la doctrine à percevoir un renversement de la position des plateformes par rapport aux contenus diffusés par leurs utilisateurs⁵¹. Bien que perçue par le législateur comme nécessaire pour lutter efficacement contre les contenus illégaux en ligne, la responsabilisation accrue des acteurs du numérique suscite néanmoins des critiques au regard de la protection et du bon respect de la liberté d'expression.

⁴⁸ Selon certains, le mécanisme établi par cette disposition suscite des inquiétudes pour la réalisation d'une juste balance des droits fondamentaux en présence, à savoir les droits de propriété intellectuelle et la liberté d'expression et le droit à l'information des utilisateurs. Voy. notamment G. FROSIO et Chr. GEIGER, «Taking Fundamental Rights Seriously in the Digital Services Act's Platform Liability Regime», *European Law Journal*, 2022 (forthcoming), disponible sur https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3747756, p. 11 et pp. 22-23.

⁴⁹ Directive (UE) 2019/790, préc., art. 17, § 1^{er}.

⁵⁰ Directive (UE) 2019/790, préc., art. 17, § 4.

⁵¹ G. FROSIO et Chr. GEIGER, *op. cit.*, pp. 6-7; T. MCGONAGLE, *op. cit.*, pp. 468 et 481; M. L. MONTAGNANI, «A New Liability Regime for Illegal Content in the Digital Single Market Strategy», in G. Frosio (éd.), *The Oxford Handbook of Online Intermediary Liability*, Oxford University Press, Oxford, 2020, pp. 299 et 309. Dans le même sens, voy. E. DOUEK, «The Limits of International Law in Content Moderation», *UC Irvine Journal of International, Transnational and Comparative Law*, 2021, vol. 6, p. 42.

A. *Interdiction des obligations générales de surveillance et respect du principe de proportionnalité*

Tout d'abord, la multiplication des obligations de diligence imposées aux plateformes, tant dans les instruments horizontaux que verticaux, soulève des interrogations sur leur articulation avec le régime d'exemption de responsabilité des intermédiaires⁵². Ces obligations imposent davantage que par le passé aux acteurs du numérique d'intervenir sur les contenus ou, à tout le moins, de mettre en œuvre volontairement et proactivement des mesures périphériques qui pourraient leur faire perdre le bénéfice de la « neutralité » essentielle à l'application du régime d'exemption. Si certains doutes peuvent partiellement être levés par l'application de la règle *lex specialis derogat generali*, il n'empêche que les solutions demeurent parfois en tension avec le principe de proportionnalité et le droit fondamental à la liberté d'expression.

Avec l'article 17 de la directive 2019/790 qui impose une obligation de « notice and stay down »⁵³ aux fournisseurs de services de partage de contenus en ligne, la mise en œuvre pratique de la législation spécifique en matière de droit d'auteur pose des questions au regard de l'interdiction des obligations générales de surveillance. L'obligation de diligence imposée pour éviter la remise en ligne de contenus contrefaits par les utilisateurs ne semble pouvoir être respectée, à tout le moins pour les plus gros acteurs, que par le recours à un outil automatisé de filtrage *ex ante* ou de contrôle des contenus au moment même du téléversement pour empêcher leur diffusion. Dès lors, même si la disposition précise qu'elle n'implique aucune obligation générale de surveil-

⁵² G. FROSIO et Chr. GEIGER, *op. cit.*, p. 7; V. MOSCON, « Free Circulation of Information and Online Intermediaries: Replacing One 'Value Gap' with Another », *International Review of Intellectual Property and Competition Law*, 2020, vol. 51, n° 8, p. 980; M. L. MONTAGNANI, *op. cit.*, pp. 467-485. *A contrario*, pour d'autres, les législations sectorielles exigeant l'adoption par les acteurs du numérique de mesures proactives sont compatibles, ou à tout le moins « complémentaires », à l'interdiction des obligations générales de surveillance. Voy. J. BARATA, *op. cit.*, p. 12.

⁵³ Dans le cadre d'un mécanisme de « notice and stay down », l'intermédiaire reçoit à l'origine un signalement pour un contenu illégal. Il doit ensuite le supprimer ou en bloquer l'accès, mais aussi déployer les mesures nécessaires pour empêcher que le contenu litigieux ne soit reposté, que ce soit par le même utilisateur ou par un autre. Ce mécanisme implique donc une obligation continue dans le chef des intermédiaires de prévenir la réapparition de la même infraction dans le futur. Sur ce point, voy. A. KUCZERAWY, « From 'Notice and Take Down' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression », in G. Frosio (éd.), *The Oxford Handbook of Online Intermediary Liability*, Oxford University Press, Oxford, 2020, pp. 538-539. Les mécanismes de « notice and stay down » peuvent également s'étendre aux contenus similaires. À cet égard, voy. *infra*, C.J.U.E., arrêt *Eva Glawischnig-Piesczek c. Facebook Ireland Limited*, 3 octobre 2019, aff. C-18/18.

lance, l'on perçoit mal en pratique les autres moyens à la disposition de ces acteurs pour respecter cette exigence. En effet, « pour reconnaître efficacement un contenu illégal, il convient d'utiliser un outil technologique pour contrôler systématiquement l'intégralité des contenus téléversés par les utilisateurs »⁵⁴. Si le contrôle de l'ensemble des contenus, pour filtrer un contenu contrefaisant déterminé, se distingue certes d'une analyse minutieuse de tout contenu pour déceler toute illégalité, la difficulté de l'obligation de « notice and stay down » réside dans le fait qu'elle impose *de facto* le recours à des outils automatisés de filtrage à l'égard de l'ensemble des utilisateurs⁵⁵. Saisie d'un recours en annulation introduit par la Pologne qui dénonçait ces mêmes écueils, la Cour de justice vient de trancher, par un arrêt du 26 avril 2022, en faveur de la compatibilité du mécanisme établi par l'article 17, § 4, de la directive 2019/790 avec le droit fondamental à la liberté d'expression⁵⁶.

Toujours dans le cadre des mécanismes de « notice and stay down », la Cour de justice a aussi pu se prononcer, avec l'affaire *Eva Glawischnig-Piesczek*⁵⁷, sur la question de savoir si l'injonction de retrait d'un contenu illégal, émise par une autorité nationale à l'encontre d'un hébergeur, pouvait s'étendre aux

⁵⁴ A. KUCZERAWY, « From 'Notice and Take Down'... », *op. cit.*, p. 541. Traduction libre. Dans le même sens, voy. M. L. MONTAGNANI, *op. cit.*, p. 303. *Contra*, voy. E. ROSATI, « La législation sur les services numériques et l'application du droit d'auteur : le cas de l'article 17 de la Directive DSM », in M. Cappello (éd.), *Décrypter la législation sur les services numériques*, *op. cit.*, pp. 84 et 86. Selon l'auteure, l'article 17 prévoit comme garantie que les titulaires de droits fournissent les informations pertinentes et nécessaires pour permettre au fournisseur de services de partage de contenus de remplir son obligation de « notice and stay down » sans basculer dans une obligation générale de surveillance. Elle précise aussi que, bien qu'aucune obligation générale de surveillance ne puisse être imposée, les acteurs du numérique peuvent, de leur propre chef, les implémenter, à condition de respecter le principe de proportionnalité et la liberté d'expression. C'est, à notre sens, là que le bât blesse : l'on perçoit mal comment une surveillance généralisée de tous les contenus téléversés par les utilisateurs pourrait se conformer à ces principes...

⁵⁵ La directive sur les services de médias audiovisuels pose moins d'inquiétudes puisqu'elle précise que les mesures que les plateformes de partage de vidéos doivent adopter pour lutter contre les contenus illégaux et dommageables ne doivent pas mener à des mesures de contrôle *ex ante* ni de filtrage des contenus au moment du téléversement. Voy. directive (UE) 2018/1808, préc., art. 28ter, § 3.

⁵⁶ C.J.U.E., Gde Ch., arrêt *République de Pologne c. Parlement européen et Conseil de l'Union européenne*, 26 avril 2022, aff. C-401/19, points 53-58, 78-80, 90 et 98. Après avoir reconnu que ce mécanisme imposait aux plateformes d'opérer un contrôle préalable des contenus téléversés à l'aide d'outils automatisés de filtrage restreignant la liberté d'expression, la Cour a considéré que des garanties appropriées étaient prévues (notamment, par les paragraphes 7 à 9).

⁵⁷ C.J.U.E., *Eva Glawischnig-Piesczek c. Facebook Ireland Limited*, préc. Dans une affaire antérieure concernant des droits de propriété intellectuelle, la Cour de justice avait estimé que

→

contenus identiques ainsi qu'aux contenus équivalents à celui pour lequel l'illicéité a été auparavant établie. La Cour a répondu par l'affirmative en divisant son raisonnement selon le type de contenus. Pour les contenus identiques, elle considère que l'injonction de retrait peut s'y étendre sans condition, peu importe que l'auteur du contenu identique diffère de l'auteur du contenu ayant fait l'objet d'un constat d'illicéité⁵⁸. Pour les contenus équivalents, elle estime que l'injonction peut les viser uniquement lorsque (i) le contenu demeure substantiellement inchangé par rapport au contenu ayant fait l'objet d'un constat d'illicéité, (ii) les différences de formulation n'amènent pas l'hébergeur à entreprendre une appréciation autonome du contenu et (iii) le contenu équivalent comporte des éléments spécifiés dans l'injonction (le nom de la personne concernée par la violation, les circonstances entourant la déclaration d'illicéité et un contenu équivalent)⁵⁹. Selon la Cour, ces conditions évitent d'imposer une charge excessive sur les épaules des hébergeurs et leur permettent de développer des outils de recherche automatisés. Si un tel arrêt offre une protection renforcée aux victimes de contenus illégaux, il provoque des risques conséquents pour les droits fondamentaux des émetteurs. D'une part, l'extension pure et simple de l'injonction à tout contenu identique, sans prise en compte des circonstances de l'espèce et du contexte particulier, pourrait résulter en la suppression d'un contenu parfaitement légal, bien qu'identique à un contenu ayant précédemment fait l'objet d'un constat d'illicéité. En effet, en matière de liberté d'expression, les facteurs contextuels possèdent une importance capitale dans la détermination des éventuelles restrictions. Pensons, par exemple, au retrait d'un article dans lequel un journaliste informe le public à propos d'un contenu illégal ou d'une activité illicite, pour la simple raison qu'il reprend un contenu ayant précédemment fait l'objet d'un constat d'illicéité ou encore à l'application d'une exception au droit d'auteur qui pourrait renverser la qualification d'un contenu soi-disant contrefaisant. D'autre part, l'extension de l'injonction aux contenus équivalents sous prétexte que les hébergeurs peuvent utiliser des outils de recherche automatisés est susceptible de porter dispropor-

←

l'injonction faite à un hébergeur de mettre en œuvre une mesure de filtrage (i) généralisée (c'est-à-dire, visant indistinctement tous les utilisateurs), (ii) des informations stockées sur ses serveurs par ses utilisateurs, (iii) à titre préventif, (iv) illimitée dans le temps et (v) à ses frais exclusifs constituait une obligation générale de surveillance. Voy. C.J.U.E., arrêt *Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) c. Netlog NV*, 16 février 2012, aff. C-360/10.

⁵⁸ C.J.U.E., *Eva Glawischnig-Piesczek c. Facebook Ireland Limited*, préc., points 37 et 53.

⁵⁹ *Ibid.*, points 45, 46 et 53.

tionnellement atteinte à la liberté d'entreprise⁶⁰. Il convient, à notre sens, de ne pas négliger la circonstance que cet arrêt a été rendu à l'égard de *Facebook* et que tous les hébergeurs n'ont pas les mêmes moyens et la même puissance. À cet égard, soulignons que le DSA adopte des obligations et une responsabilisation graduées en fonction du type d'acteur. En tout état de cause, ces mécanismes de « notice and stay down » soulèvent, selon nous, toujours de sérieuses interrogations au regard de leur articulation avec l'interdiction des obligations générales de surveillance.

B. *Pratiques de suppression ou de blocage excessifs*

Un autre danger pour la stabilité du régime d'exemption des intermédiaires provient de la nouvelle protection des « bons Samaritains » instituée par le DSA. Son objectif est d'empêcher que les intermédiaires qui entreprennent, de leur propre initiative, des actions pour détecter, recenser ou combattre les contenus et les activités illicites ne soient privés du bénéfice des exonérations de responsabilité, à condition d'agir de bonne foi et avec diligence⁶¹. L'on comprend rapidement que l'ambition réelle du législateur de l'Union est de renforcer la mise en place de mesures proactives par les acteurs du numérique dans le cadre de sa politique de lutte contre les contenus illégaux, tendance que l'on retrouve déjà du côté du secteur lui-même dans les instruments d'autorégulation⁶². Si cette protection rassurera certainement les intermédiaires et les rendra plus enclins à développer des outils de détection des contenus illégaux sur leurs services, elle renforcera nécessairement les pratiques d'« overcompliance », c'est-à-dire de suppression ou de blocage d'expressions parfaitement licites, alors que la *ratio legis* du régime d'exemption de responsabilité des intermé-

⁶⁰ Pour certains acteurs, le coût économique du développement des mesures techniques et organisationnelles nécessaires au respect des exigences de filtrage peut attenter à la concurrence et à l'innovation et ainsi constituer une sérieuse violation de la liberté d'entreprise. Voy. G. FROSIO et Chr. GEIGER, *op. cit.*, pp. 30-31 et 44.

⁶¹ Proposition de règlement DSA, préc., art. 6 et cons. 25.

⁶² Voy. Code européen de bonnes pratiques contre la désinformation, adopté en septembre 2018; Code de conduite de l'UE visant à combattre les discours de haine illégaux en ligne, 30 juin 2016. Pour la désinformation, les signataires s'engagent à intensifier les mesures proactives qu'ils adoptent, dont le développement d'outils automatisés, avec les risques que l'on connaît pour la liberté d'expression et le manque de prise en compte d'éléments contextuels importants pour la proportionnalité de toute mesure de modération. Pour les discours de haine illégaux, l'on retrouve également des engagements critiquables au regard de la liberté d'expression, tels que le traitement de la majorité des signalements soumis en moins de 24 heures, multipliant ainsi les risques d'analyses bâclées ne tenant pas compte d'éléments contextuels.

diaires était justement d’y pallier... Comme le décrit très justement Aleksandra Kuczerawy, avec une telle disposition, les hébergeurs qui, par le passé, étaient réticents à entreprendre des mesures proactives par crainte d’endosser un rôle actif qui leur ferait perdre le bénéfice de l’exonération pourraient désormais agir avec moins de raison, ce qui induirait davantage de suppression de contenus⁶³. La protection des « bons Samaritains » mène par ailleurs à une situation contraire à l’objectif clé du DSA d’assurer une meilleure protection des droits fondamentaux, en ce sens qu’il devient plus prudent pour les hébergeurs « de retirer plus, plutôt que moins, afin d’éviter toute responsabilité »⁶⁴.

Les pratiques de suppression ou de blocage excessifs sont par ailleurs renforcées par les instruments qui imposent aux acteurs du numérique des obligations d’agir dans des délais très serrés et, davantage encore, lorsque le non-respect de telles obligations est assorti de lourdes sanctions financières⁶⁵. Ces dispositions légales « agressives » les poussent, par crainte de voir leur responsabilité engagée, à préférer, dans le doute, retirer ou bloquer un contenu. Ce risque est encore exacerbé lorsque les instruments sanctionnent le non-respect des obligations de diligence de modération des contenus sans sanctionner les pratiques d’*overcompliance*⁶⁶.

La Cour européenne des droits de l’homme s’est souvent prononcée sur des affaires relatives à des retraits ou à des blocages d’accès excessifs. À son estime, une ordonnance de blocage d’accès émise sans qu’un tribunal ne se soit préalablement prononcé sur l’illégalité d’un contenu constitue une restriction préalable à la liberté d’expression. Or, une telle restriction ne se justifie, au titre de la Convention, que sur la base de circonstances exceptionnelles et lorsqu’il existe un cadre juridique clair permettant un contrôle juridictionnel effectif et approfondi des mesures adoptées⁶⁷. Récemment, à plusieurs reprises, la Cour a d’ailleurs conclu à des violations de l’article 10 de la Convention dans des affaires de blocage de sites web entiers sans spécification des pages

⁶³ A. KUCZERAWY, « The Good Samaritan that wasn’t: voluntary monitoring under the (draft) Digital Services Act », 14 janvier 2021, disponible sur www.law.kuleuven.be/citip/blog/the-good-samaritan-that-wasnt/. Dans le même sens, voy. J. BARATA, *op. cit.*, p. 18.

⁶⁴ A. KUCZERAWY, *ibid.* Traduction libre.

⁶⁵ Lorsque ces obligations existent, elles devraient se limiter, pour des raisons de proportionnalité, à viser les contenus dont le caractère illicite est évident et qui sont susceptibles de causer un dommage imminent, tels que les contenus à caractère pédopornographique. Sur ce point, voy. G. FROSIO et Chr. GEIGER, *op. cit.*, p. 38.

⁶⁶ M. HANOT et A. MICHEL, *op. cit.*, p. 189.

⁶⁷ Cour eur. dr. h., arrêt *Kablis c. Russie*, 20 avril 2019, §§ 90-92; arrêt *OOO Flavia e.a. c. Russie*, 23 juin 2020, § 39.

web problématiques⁶⁸ ou en raison d'un seul contenu problématique⁶⁹, de blocage collatéral de sites web partageant la même adresse IP qu'un site web diffusant des contenus illégaux⁷⁰ et de menaces de blocage total d'un site web en l'absence de retrait de certains contenus⁷¹. Dans tous ces cas, la condition de nécessité et l'absence de garanties appropriées posaient problème. L'existence de garanties appropriées et efficaces pour protéger la liberté d'expression est un élément clé pour évaluer la conformité des mesures de retrait et de blocage avec l'article 10, notamment en ce qui concerne les conditions de légalité et de nécessité. Comme le déclare la Cour de Strasbourg, «il existe une incompatibilité avec l'État de droit lorsque le cadre juridique n'établit pas de garanties capables de protéger les individus contre les effets excessifs et arbitraires des mesures de blocage»⁷².

C. Adoption d'approches indifférenciées

Ensuite, il faut également se méfier des approches indifférenciées dans la régulation de l'expression. L'approche est dite «différenciée» lorsque les solutions et les mesures qui sont recommandées prennent en compte tant le type de contenu que l'objectif de sa diffusion. En effet, en matière de liberté d'expression, le contexte est un élément primordial dans l'analyse de la nécessité d'une ingérence.

Dans un premier temps, en raison de leur nature intrinsèquement différente, les contenus illégaux et les contenus dommageables ne peuvent pas faire l'objet des mêmes mesures de modération⁷³. Un contenu illégal ne se conforme pas aux dispositions légales en vigueur tandis qu'un contenu dommageable ne viole pas de disposition légale mais, bien que «légal», peut porter atteinte aux sentiments de certaines personnes (généralement, les plus vulnérables comme les mineurs), aux valeurs (démocratiques, culturelles ou religieuses) ou, parfois, à l'ensemble d'une société donnée. Un contenu dommageable peut également s'avérer problématique en raison d'un contexte particulier (climat politique ou

⁶⁸ *OOO Flavus e.a. c. Russie*, préc., §§ 37-38. Pour la Cour, le blocage total d'un site web est une mesure extrême qui ignore la distinction entre les contenus légaux et illégaux qui sont hébergés. Cette mesure rend de nombreux contenus légaux indisponibles et doit être spécifiquement justifiée.

⁶⁹ Cour eur. dr. h., arrêt *Bulgakov c. Russie*, 23 juin 2020.

⁷⁰ Cour eur. dr. h., arrêt *Vladimir Kharitonov c. Russie*, 23 juin 2020.

⁷¹ Cour eur. dr. h., arrêt *Engels c. Russie*, 23 juin 2020.

⁷² *Vladimir Kharitonov c. Russie*, préc., § 46. Traduction libre. Voy. également *Engels c. Russie*, préc., § 34.

⁷³ Dans le même sens, voy. G. FROSIO et Chr. GEIGER, *op. cit.*, pp. 4 et 37.

social tendu). Il s'agit typiquement de la désinformation qui, au niveau du droit de l'Union⁷⁴, est un contenu qualifié de «légal mais dommageable»⁷⁵. Alors que l'on considère que les contenus illégaux doivent être traités à la source et que leur pure suppression se justifie, les contenus dommageables, pour des raisons de proportionnalité, sont soumis à des moyens plus souples, tels que l'autonomisation des utilisateurs, la sensibilisation, l'éducation aux médias, la transparence, le signalement, le contre-discours, la démonétisation ou encore la priorisation ou la dégradation de la visibilité selon des paramètres choisis par l'utilisateur⁷⁶. Par ailleurs, au sein même des contenus illégaux, certaines catégories de contenus doivent, en raison du contexte, faire l'objet de mesures plus fortes que d'autres. Si l'illégalité d'un contenu pédopornographique peut apparaître évidente au premier regard, la détermination du caractère illicite d'autres contenus peut nécessiter un examen plus approfondi⁷⁷. C'est notamment le cas des atteintes au droit d'auteur où l'application d'une exception exige une appréciation autonome du contenu. Le DSA ne tient pas compte de cette nécessité de renforcer l'étape d'appréciation et de contrôle du contenu en cas de doute sur son caractère légal ou illégal, ce qui est à notre sens regrettable.

Dans un second temps, dans la détermination des mesures de modération, il convient de prêter attention à l'objectif poursuivi par la personne qui diffuse un contenu. Le règlement 2021/784 a tenu compte de cet élément et mentionne que les contenus diffusés à des fins éducatives, journalistiques, artistiques, de recherche, de prévention ou de lutte contre le terrorisme ne constituent pas des contenus à caractère terroriste⁷⁸. Ces contenus sont donc exclus des mesures de modération prévues et le législateur insiste sur la nécessité d'une analyse pour déterminer le «véritable objectif» de la diffusion. Le DSA ne prévoit aucune exclusion de ce genre du champ d'application. Ce choix du législateur, qui a d'ailleurs fait couler beaucoup d'encre⁷⁹, est hautement critiquable, surtout au regard de la liberté de la presse. Il apparaît en effet à notre sens inconcevable

⁷⁴ Dans certains États membres, la désinformation est considérée comme illégale en vertu du droit national. Voy. notamment *supra*, note n° 6, concernant les contenus de désinformation en période électorale en France.

⁷⁵ Les contenus dommageables peuvent également inclure les contenus pornographiques, les avances sexuelles non désirées et les expressions sur des sujets sensibles.

⁷⁶ Voy. Plan d'action de la Commission européenne visant à promouvoir une utilisation sûre d'Internet, Bruxelles, 26 novembre 1997, IP/97/1041. Voy. également G. FROSIO et Chr. GEIGER, *op. cit.*, p. 4 et pp. 37-38.

⁷⁷ Sur ce point, voy. G. FROSIO et Chr. GEIGER, *ibid.*, pp. 4 et 37.

⁷⁸ Règlement (UE) 2021/784, préc., art. 1^{er}, § 3.

⁷⁹ Lors des discussions, des amendements ont été déposés afin d'introduire une exemption pour les contenus journalistiques. Aucun n'a été retenu. Voy. L. BERTUZZI, «L'exemption pour

→

que les mesures de modération listées dans le DSA puissent s'appliquer aux activités journalistiques exercées sur les plateformes, alors que le contenu journalistique jouit en principe d'une protection renforcée⁸⁰ par rapport aux autres contenus.

D. *Définition imprécise du champ d'application matériel*

La définition du champ d'application matériel des instruments de modération des contenus constitue un autre point d'attention pour la liberté d'expression. Un champ d'application matériel composé de définitions imprécises peut mener au retrait ou au blocage de contenus légaux et ainsi violer la liberté d'expression et le droit à l'information des internautes.

En matière de lutte contre les contenus terroristes en ligne, la doctrine a dénoncé le caractère large de la notion de «contenu terroriste», englobant de nombreux comportements, qui lui confère un caractère flou et vague, au préjudice des principes de légalité et de proportionnalité⁸¹. Une critique semblable peut être adressée à l'égard du DSA qui ne définit pas exhaustivement⁸² les catégories de contenus illicites visées mais qui fait dépendre la notion d'illicéité tant du droit de l'Union que de droit national. Cette approche confère une acception extrêmement large au «contenu illicite» lui permettant de recouvrir tout type d'«informations relatives aux contenus, produits, services et activités illicites»⁸³. Le champ d'application matériel du DSA dépasse dès lors large-

←

les médias a été exclue des négociations du DSA, mais pourrait revenir», 24 novembre 2021, disponible sur www.euractiv.fr/section/medias/news/lexemption-pour-les-medias-a-ete-exclue-des-negociations-du-dsa-mais-pourrait-revenir/.

⁸⁰ Sur la protection renforcée de la liberté de la presse, voy. Cour eur. dr. h., arrêt *Tillack c. Belgique*, 27 novembre 2007, § 55; Gde Ch., arrêt *Stoll c. Suisse*, 10 décembre 2007, §§ 105 et 106; arrêt *Ressiot e.a. c. France*, 28 juin 2012, § 101.

⁸¹ Voy. entre autres, E. DELHAISE, *op. cit.*, pp. 49-50.

⁸² Giancarlo Frosio et Christophe Geiger plaident pour une définition exhaustive des contenus illégaux soumis au DSA (G. FROSIO et Chr. GEIGER, *op. cit.*, pp. 4, 20 et 37-38). De son côté, David Kaye souligne que, en vertu du principe de légalité, les restrictions à l'expression doivent établir avec une précision suffisante la distinction entre expressions licite et illicite. De même, les restrictions contractuelles imposées par les acteurs du numérique à leurs utilisateurs doivent permettre de déterminer avec un degré suffisant de prévisibilité les contenus inadmissibles qui pourraient faire l'objet d'une restriction. Voy. Rapport du Rapporteur spécial des Nations Unies, préc., §§ 7 et 46.

⁸³ Proposition de règlement DSA, préc., cons. 12. Voy. l'article 2, (g), pour la notion de contenu illicite.

ment les catégories de contenus illégaux harmonisées au niveau de l'Union, mais englobe aussi tout autre contenu qui serait illégal selon le droit national.

Une autre imprécision du champ d'application matériel du DSA se rapporte aux contenus dommageables. Alors que le champ d'application matériel est circonscrit aux contenus illicites, l'on constate que certaines obligations imposées par le DSA s'étendent pourtant aux contenus dommageables. Il en va ainsi des articles 26 et 27 qui imposent aux très grandes plateformes en ligne d'évaluer et d'atténuer les risques systémiques suscités par l'utilisation de leurs services. Ces risques systémiques couvrent à la fois la diffusion de contenus illégaux mais aussi tout effet négatif pour l'exercice des droits fondamentaux ainsi que toute manipulation intentionnelle de leurs services pouvant impacter la santé publique, les mineurs, le discours civique, les élections ou la sécurité publique⁸⁴. Il est indubitable que le législateur entend notamment ici viser la désinformation qui est pourtant un contenu dommageable au niveau du droit de l'Union ainsi que dans la grande majorité des États membres. Nous nous rallions à cet égard à l'avis de Giancarlo Frosio et Christophe Geiger selon lequel les contenus dommageables devraient pourtant demeurer en dehors du champ d'application du DSA⁸⁵. En effet, comme nous l'avons précédemment relevé, ces contenus doivent faire l'objet de mesures spécifiques.

IV. Vers une intégration des droits humains dans les pratiques de modération ?

En raison de la pression régulatoire qui pèse sur les acteurs du numérique et des risques qui en découlent pour l'expression en ligne, des voix s'élèvent pour replacer les garanties offertes par les standards internationaux des droits humains au cœur de la modération des contenus. Toutefois, comme nous allons le voir, la réflexion autour de la conformité des pratiques des plateformes avec les droits fondamentaux n'en est qu'à ses balbutiements et de nombreuses zones d'ombre doivent encore être éclaircies pour passer de la théorie à la pratique⁸⁶.

⁸⁴ Voy. Proposition de règlement DSA, préc., art. 26 et 27 et cons. 53 à 59.

⁸⁵ G. FROSIO et Chr. GEIGER, *op. cit.*, pp. 4 et 37.

⁸⁶ ARTICLE 19, «Social Media Councils: One piece in the puzzle of content moderation», 12 octobre 2021, disponible sur www.article19.org/wp-content/uploads/2021/10/A19-SMC.pdf, pp. 6 et 22.

A. *Le consensus autour des standards internationaux des droits humains*

Ces dernières années, les appels se sont multipliés au sein des acteurs de la société civile et du milieu académique pour alerter les législateurs et le secteur sur l'importance des droits humains dans les pratiques de modération des contenus en ligne⁸⁷. À cet égard, soucieux des dérives de la délégation de prérogatives publiques à des entreprises privées⁸⁸, David Kaye propose un cadre pour la modération des contenus qui offre aux standards internationaux des droits humains une place centrale⁸⁹. Il invite ainsi les plateformes à intégrer directement, dans leurs conditions générales, les normes clés des droits fondamentaux qui impliquent que toute mesure restreignant un contenu respecte les principes de légalité, de légitimité et de nécessité⁹⁰.

Le consensus autour du rôle des standards internationaux des droits humains tend à se renforcer au fil du temps (à tout le moins, en Occident), notamment grâce à leur aspiration à l'universalité et à leur légitimité⁹¹. Ils

⁸⁷ Voy., parmi d'autres, ARTICLE 19, «Side-stepping rights: Regulating speech by contract», Policy brief, 2018, disponible sur www.article19.org/wp-content/uploads/2018/06/Regulating-speech-by-contract-WEB-v2.pdf; ARTICLE 19, «Social Media Councils...», *op. cit.*; E. DOUEK, *op. cit.*, pp. 37-38; G. FROSIO et Chr. GEIGER, *op. cit.*, pp. 3, 6, 36 et 37. Le Parlement européen invite également la Commission à intégrer les standards internationaux des droits humains dans la révision de la directive sur le commerce électronique (donc, dans l'adoption du DSA). Voy. Résolution du Parlement européen, «La législation sur les services numériques et les questions liées aux droits fondamentaux», 20 octobre 2020, (2020/2022(INI)), point 19.

⁸⁸ Niva Elkin-Koren et Maayan Perel résumant ces risques pour la démocratie et l'État de droit sous trois axes majeurs: (i) elle met à mal la distinction entre les responsabilités publiques et les intérêts privés, (ii) elle délègue le pouvoir décisionnel sur la légitimité des contenus à des outils algorithmes opaques et (iii) elle contourne la garantie constitutionnelle de la séparation des pouvoirs (N. ELKIN-KOREN et M. PEREL, *op. cit.*, p. 670).

⁸⁹ Rapport du Rapporteur spécial des Nations Unies, préc., § 2. Voy. aussi Recommandation CM/Rec(2018)2 du Comité des ministres aux États membres sur les rôles et les responsabilités des intermédiaires d'Internet, adoptée par le Comité des ministres le 7 mars 2018, lors de la 1309^e réunion des Délégués des ministres, point 11.

⁹⁰ Rapport du Rapporteur spécial des Nations Unies, préc., §§ 45 et 70.

⁹¹ M. HANOT (éd.), A. VIDAL (coord.) et al., *The Media Councils Debates: Facing the Challenges of the Digital Age*, coll. CDJ Recherches & Enjeux, AADJ/CDJ, Bruxelles, janvier 2022, p. 105. Les standards internationaux des droits humains permettent également de débattre des pratiques de modération avec un vocabulaire commun à tout un chacun, de bénéficier d'arguments solides pour repousser les demandes arbitraires de censure formulées par des gouvernements autoritaires et d'offrir de réelles garanties procédurales pour atténuer le pouvoir des plateformes. Sur les avantages offerts par les standards internationaux des droits humains, voy. E. DOUEK, *op. cit.*, pp. 44-49. Voy. aussi le Rapport du Rapporteur spécial des Nations Unies, préc., spéc. §§ 42 et 70.

offrent de réelles garanties procédurales aux internautes – telles que la transparence des pratiques et le renforcement des voies de recours pour contester les décisions – qui limitent le risque d'imposition de restrictions disproportionnées à la libre expression et le pouvoir décisionnel des entreprises privées. Comme l'a formulé Evelyn Douek, la force des standards internationaux des droits humains « ne vient pas du fait qu'ils obligent les plateformes à prendre les bonnes décisions, mais qu'ils les guident pour qu'elles prennent leurs décisions de la bonne manière »⁹². Du côté du secteur lui-même, pourtant réticent au départ, l'on perçoit également un changement de mentalité par rapport au respect des droits fondamentaux⁹³. En effet, les acteurs du numérique semblent davantage sensibilisés à la question que par le passé⁹⁴. Néanmoins, en raison du rôle prédominant qu'ils jouent à l'échelle mondiale sur le débat public et sur le contrôle du contenu, il s'impose de s'assurer que leurs déclarations ne s'apparentent en réalité pas à une forme de « human rights washing »⁹⁵, c'est-à-dire à des stratégies d'entreprise trompeuses sur la conformité des pratiques de modération des contenus avec les standards de protection des droits humains dans le but d'embellir l'image de l'entreprise. Pour qu'il ne s'agisse pas de « simples paroles en l'air », ces engagements des géants du numérique doivent entraîner une responsabilisation ainsi que de réelles contraintes⁹⁶.

À notre estime, le consensus autour des droits humains doit s'étendre tant aux standards internationaux que régionaux. Puisque la portée de la liberté d'expression est éminemment contextuelle, les instruments régionaux ainsi que l'interprétation qui en est donnée par les autorités judiciaires locales doivent indispensablement être pris en compte dans les pratiques de modération des contenus⁹⁷.

⁹² E. DOUEK, *op. cit.*, p. 63. Traduction libre.

⁹³ *Ibid.*, p. 50.

⁹⁴ Les géants du numérique adoptent des déclarations dans lesquelles ils s'engagent à respecter les standards des droits humains. Par exemple, Facebook s'engage à respecter les instruments internationaux de protection des droits humains selon les approches développées dans les principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme. Voy. Facebook, « Corporate Human Rights Policy », disponible sur <https://about.fb.com/wp-content/uploads/2021/04/Facebooks-Corporate-Human-Rights-Policy.pdf>.

⁹⁵ M. HANOT (éd.), A. VIDAL (coord.) *et al.*, *op. cit.*, p. 105.

⁹⁶ E. DOUEK, *op. cit.*, p. 40.

⁹⁷ Dans le même sens, voy. *ibid.*, pp. 53 à 55.

B. *Les limites des droits humains comme guide des pratiques de modération*

Il est indéniable que les standards internationaux (et régionaux) des droits humains offrent des garanties élevées pour la protection de la liberté d'expression. Néanmoins, il convient de ne pas négliger les limites de leur application aux pratiques de modération des contenus par les acteurs du numérique. Dans un article paru en 2021, Evelyn Douek se livre, à cet égard, à ce qu'elle appelle une « critique amicale » en mettant en lumière les limites de l'application de tels standards. À son estime, adopter une approche idéaliste au regard de l'impact concret des standards internationaux des droits humains en fermant les yeux sur leurs limites ne fera que renforcer la position des géants du numérique au détriment des intérêts de leurs utilisateurs⁹⁸.

Au-delà des controverses existantes sur l'applicabilité et le caractère contraignant des droits fondamentaux à l'égard d'entreprises privées⁹⁹, l'auteure dénonce plusieurs limites à l'utilisation des standards internationaux des droits humains dans les pratiques de modération des acteurs du numérique. Tout d'abord, elle relève que le consensus autour de ces standards n'est pas si universel qu'il y paraît, surtout en matière de liberté d'expression en raison des différences culturelles, sociétales et communautaires¹⁰⁰. Ensuite, ces standards ne forment pas un corps de règles uniforme : il existe en réalité des incohérences et des divergences d'interprétation, qui s'expliquent par la variété d'instruments (internationaux et régionaux) et de cours et tribunaux amenés à les appliquer¹⁰¹. Par ailleurs, elle note les zones d'ombre qui persistent dans le champ des droits fondamentaux, laissant aux plateformes un large pouvoir de discrétion dans les cas les plus difficiles¹⁰². Toutefois, malgré le fait que ces dernières jouissent d'un considérable pouvoir décisionnel sur les aspects substantiels des normes, rappelons que les droits humains offrent des garanties procédurales non négligeables aux internautes. D'ailleurs, la Cour de Strasbourg a déjà eu l'occasion de mettre en lumière les garanties appropriées qui doivent nécessai-

⁹⁸ *Ibid.*, pp. 39 et 41.

⁹⁹ Contentons-nous de relever que plusieurs interprétations s'opposent : alors que d'aucuns considèrent que les acteurs privés ne sont nullement contraints de se conformer aux droits humains, d'autres estiment en revanche que tel est le cas. Sur la conformité des pratiques des entreprises aux droits fondamentaux, voy. Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme. Ces principes responsabilisent les entreprises privées et leur fournissent des indications pour se conformer aux exigences des droits fondamentaux.

¹⁰⁰ E. DOUEK, *op. cit.*, pp. 51-52.

¹⁰¹ *Ibid.*, pp. 52-56.

¹⁰² *Ibid.*, pp. 56-58.

rement entourer les mesures de modération¹⁰³. Les enseignements de la Cour en la matière peuvent être résumés comme suit : (1) l'auteur ainsi que toute autre personne dont les droits et les intérêts peuvent être affectés doivent être informés à l'avance de la mesure afin de leur permettre d'intervenir dans la procédure, (2) la mesure et sa justification doivent être communiquées à l'auteur, (3) les autorités doivent évaluer soigneusement l'impact de la mesure avant de la mettre en œuvre ou justifier les raisons de son application immédiate, (4) l'auteur doit se voir offrir la possibilité de retirer ou de bloquer le contenu illégal avant l'application de la mesure, (5) les parties concernées doivent pouvoir contester la mesure et être entendues par un tribunal ou un organe indépendant, et (6) tous les utilisateurs doivent pouvoir se rendre compte qu'une mesure de modération est intervenue, être informés des recours disponibles et avoir accès à une série d'informations sur la mesure (base juridique, organe émetteur, date, numéro, texte et motivation de la décision, etc.). La limite suivante décrite par Evelyn Douek découle de la précédente : la souplesse et la formulation vague des normes permettent aux acteurs privés de s'«emparer» du langage et de la légitimité des standards internationaux des droits humains, plutôt que d'être réellement contraints dans leurs activités par ces standards¹⁰⁴. En outre, elle considère que, même lorsqu'elles agissent de bonne foi, les plateformes ne disposent en réalité ni de la légitimité, ni de l'expertise et des compétences adéquates pour se livrer à des évaluations délicates et à des mises en balance d'intérêts en présence¹⁰⁵. Par exemple, les acteurs privés sont-ils réellement les plus à même de se prononcer sur la légitimité d'une mesure au regard de la sécurité nationale ou de l'ordre public ou peut-on réellement attendre d'un acteur privé qu'il ne prenne pas en compte ses propres intérêts commerciaux et sa liberté d'entreprendre dans sa prise de décision ? Enfin, l'auteure avance une dernière limite qui se révèle moins convaincante dès lors que l'on accorde de l'importance aux facteurs contextuels dans l'appréciation des restrictions à la liberté d'expression. Elle considère que les standards des droits humains ne permettent pas encore de faire face au bouleversement de l'écosystème numérique qui génère des quantités de contenus astronomiques, nécessitant alors

¹⁰³ Voy., entre autres, Cour eur. dr. h., arrêt *Mariya Alekhina e.a. c. Russie*, 17 juillet 2018, §§ 265-267 ; *Bulgakov c. Russie*, préc., §§ 35-37 ; *Vladimir Kharitonov c. Russie*, préc., §§ 43-44 ; *Engels c. Russie*, préc., § 31 ; *OOO Flavus e.a. c. Russie*, préc., §§ 40-43. Sur ces affaires, voy. A. MICHEL et Q. VAN ENIS, « Médias, liberté d'expression et nouvelles technologies [Chronique de jurisprudence 2018-2020] », *R.D.T.I.*, n° 82-83/2021, pp. 111-113.

¹⁰⁴ E. DOUEK, *op. cit.*, pp. 58-60.

¹⁰⁵ *Ibid.*, pp. 60-63. Dans le même sens, voy. A. KUCZERAWY, « From 'Notice and Take Down'... », *op. cit.*, p. 527 ; T. MCGONAGLE, *op. cit.*, p. 484. Voy. aussi le Rapport du Rapporteur spécial des Nations Unies, préc., § 17.

des approches systémiques de la liberté d'expression et non plus individualistes comme tel est le cas actuellement¹⁰⁶.

Au vu de ces limites, l'on comprend que l'intégration des standards internationaux des droits humains dans les pratiques de modération est loin d'être une mince affaire. Pour guider les États dans la définition des meilleures pratiques pour l'établissement de cadres juridiques et procéduraux efficaces pour les mécanismes d'autorégulation et de corégulation de la modération des contenus, le Conseil de l'Europe a élaboré une note d'orientation dans laquelle il invite ces derniers à : (1) identifier précisément le problème à traiter, (2) assurer la prévisibilité des mesures mises en œuvre, (3) éviter tout risque de conformité excessive ou de mise en œuvre discriminatoire, (4) énumérer les droits fondamentaux en jeu dans chaque cas concret et déterminer les mesures nécessaires pour les protéger, (5) définir de manière adéquate la nature de l'approche réglementaire en précisant le degré d'implication de l'État, (6) s'appuyer sur les réalisations de systèmes similaires d'autorégulation ou de corégulation dans un souci d'efficacité et de minimisation des effets négatifs, et (7) veiller à ce que toutes les données nécessaires à la transparence soient collectées et rendues publiques¹⁰⁷.

C. *L'initiative Social Media Councils d'Article 19, un exemple de bonne pratique*

Comme nous l'avons déjà mentionné, les acteurs du numérique jouissent actuellement d'un pouvoir et d'un contrôle considérables sur le débat public. En raison du flot immense de contenus qui circule au quotidien en ligne, l'on pourrait reconnaître, d'un point de vue pragmatique, qu'ils pourraient effectivement sembler les mieux à même d'agir. La pression s'intensifie toutefois autour des législateurs et des politiques pour ne plus laisser aux géants du numérique la mainmise sur l'expression des internautes et pour qu'un réel contrôle assorti de garanties appropriées entoure leurs pratiques¹⁰⁸.

¹⁰⁶ E. DOUEK, *op. cit.*, pp. 64-66.

¹⁰⁷ Conseil de l'Europe, « Modération de contenu : Meilleures pratiques en vue de la mise en place de cadres juridiques et procéduraux efficaces pour les mécanismes d'autorégulation et de corégulation de la modération de contenu », note d'orientation adoptée par le Comité directeur sur les médias et la société de l'information, 19-21 mai 2021, point 14, disponible sur <https://rm.coe.int/content-moderation-fr/1680a40294>.

¹⁰⁸ Ce climat a justifié l'adoption par le législateur de l'Union du DSA. Sur ce point, voy. *supra*, point II, B, 1.

L'on vient aussi de relever le fait que l'implémentation des principes des droits humains dans les pratiques de modération des contenus suscite son lot d'interrogations. Néanmoins, comme l'a très justement déclaré Pierre François Docquir lors d'un séminaire dédié à l'apport de l'autorégulation journalistique à la régulation de l'expression, «souhaitons-nous que les plateformes en ligne utilisent le langage des droits humains selon leurs propres préférences ou faut-il mettre en place une réelle conversation où tous les acteurs de la société pourraient avoir leur mot à dire sur ce que signifie l'application des droits humains à la modération des contenus?»¹⁰⁹

À cet égard, l'initiative «Social Media Councils»¹¹⁰ (ci-après : «SMC») lancée par l'ONG Article 19 nous apparaît particulièrement prometteuse. L'objectif consiste à développer un mécanisme de contrôle externe des pratiques de modération des acteurs du numérique. Au sein de l'Union, les instruments qui contiennent les règles applicables à la modération promeuvent largement, à l'appui de l'approche régulatoire, les mécanismes d'autorégulation et de corégulation ainsi que les systèmes de règlement extrajudiciaire des litiges. Une telle initiative pourrait donc parfaitement s'insérer dans le paysage européen. En s'inspirant de l'expérience des conseils de presse, le SMC entend créer un forum indépendant, volontaire, inclusif, transparent et responsable où toutes les parties prenantes peuvent avoir leur mot à dire sur la modération des contenus, dans le respect des droits humains. Ils regrouperaient à la fois les acteurs du numérique, les régulateurs des médias, les conseils de presse, les médias et les journalistes, l'industrie publicitaire, les ONG et autres acteurs de la société civile ainsi que le milieu académique et de la recherche. Le SMC aurait à la fois la compétence pour examiner des plaintes individuelles après épuisement des voies de recours internes à la plateforme, mais aussi pour émettre des recommandations sur les pratiques de modération¹¹¹. Dans un objectif de compréhension et de prise en compte adéquate des spécificités contextuelles et commu-

¹⁰⁹ M. HANOT (éd.), A. VIDAL (coord.) *et al.*, *op. cit.*, p. 105. Traduction libre.

¹¹⁰ À l'origine de l'initiative, l'on retrouve des critiques quant aux pratiques des plateformes. Elles concernent le manque de transparence, de responsabilisation, de sécurité juridique, de voies de recours effectives, de garanties appropriées pour assurer la liberté d'expression ainsi que le fait que les plateformes basent intégralement leurs pratiques de modération sur leurs conditions générales. À ce propos, voy. ARTICLE 19, «Side-stepping rights...», *op. cit.*, pp. 14-18; P. Fr. DOCQUIR, «The Social Media Council: Bringing Human Rights Standards to Content Moderation on Social Media», 28 octobre 2019, disponible sur www.cigionline.org/articles/social-media-council-bringing-human-rights-standards-content-moderation-social-media/. L'initiative est d'ailleurs saluée par David Kaye. Voy. Rapport du Rapporteur spécial des Nations Unies, préc., § 58.

¹¹¹ ARTICLE 19, «Social Media Councils...», *op. cit.*

nautaires, le SMC s'établira au niveau local¹¹² avec une large représentation au sein de ses membres. Pour apprécier adéquatement une éventuelle restriction au regard des standards internationaux des droits humains, il convient en effet de prendre en compte tous les éléments pertinents du contexte, qu'ils soient politiques, culturels, linguistiques, sociologiques, religieux, philosophiques, historiques ou encore patrimoniaux. Une prise de décision respectueuse des droits fondamentaux exige ainsi de prendre en compte les différentes perspectives et expertises ainsi que de donner la parole à tous les «sentiments» qui composent une société, d'autant plus lorsque l'on entend comprendre et résoudre des questions sociétales complexes telles que le discours de haine ou la désinformation¹¹³.

Conclusion

Plus que jamais, Internet constitue un outil clé pour l'échange d'idées et d'informations. Il est donc essentiel de garantir la liberté d'expression dans toute démarche qui entend réguler les contenus en ligne.

Le législateur européen multiplie les instruments, qu'ils soient horizontaux ou verticaux, pour cadrer le pouvoir que les géants du numérique exercent sur le débat public en leur imposant de nombreuses obligations de diligence au regard des contenus diffusés par leurs utilisateurs. En la matière, la complexité des équilibres entre les droits et les intérêts en présence nécessite l'adoption de garanties adéquates ainsi que de prendre garde à ne pas créer de nouveaux dangers pour la liberté d'expression. Afin d'éviter la censure de contenus légitimes, il faudrait, plutôt que d'augmenter la pression sur les acteurs du numérique pour modérer rapidement plus de contenus, les responsabiliser davantage quant à leurs propres pratiques et les inciter à modérer mieux¹¹⁴. Pour emprunter les termes de Tarlach McGonagle, la régulation «en faveur» de la liberté d'expression ne doit en effet pas devenir la régulation «de» la liberté d'expression¹¹⁵...

¹¹² Un projet pilote de SMC en Irlande est actuellement en cours. De nombreuses discussions ont eu lieu entre les parties prenantes courant 2021 et un groupe de travail va prochainement être convoqué. Sa principale tâche sera d'étudier l'application des standards internationaux des droits humains aux pratiques de modération et de rédiger un «Code des principes des droits humains pour la modération des contenus». Voy. *ibid.*, pp. 13-27.

¹¹³ *Ibid.*, pp. 22-24.

¹¹⁴ N. ELKIN-KOREN et M. PEREL, *op. cit.*, pp. 670, 674 et 678.

¹¹⁵ T. MCGONAGLE, *op. cit.*, p. 467. L'auteur formulait toutefois ces propos sous forme interrogative.

Toutefois, comme nous l'avons relevé, la problématique de la modération des contenus présente de nombreuses zones d'ombre qui doivent encore être éclaircies pour parvenir à mettre adéquatement en œuvre les normes des droits fondamentaux dans les pratiques des acteurs du numérique. Un élément capital à cet égard concerne la compréhension et la prise en compte du contexte dans lequel s'inscrit la diffusion d'un contenu pour apprécier la nécessité d'une éventuelle restriction et ainsi parvenir à une décision conforme aux standards internationaux et régionaux des droits humains. L'on saisit d'emblée à cet égard les lacunes des outils automatisés de modération des contenus qui fleurissent sur l'ensemble des services offerts par les acteurs du numérique. Les technologies ne peuvent pas remplacer l'humain sur tous les fronts, et certainement pas pour les appréciations contextuelles des restrictions à la liberté d'expression. Ni, d'ailleurs, pour assumer la responsabilité d'une mauvaise prise de décision.

Reconnaître les faiblesses et les défis de l'application des standards internationaux (et régionaux) des droits humains permettra indéniablement l'élaboration de solutions appropriées. À notre estime, la résolution de cette problématique complexe ne peut être laissée aux seules mains du secteur et du régulateur. Dans une approche conforme aux droits fondamentaux, il importe d'impliquer et de responsabiliser l'ensemble des parties prenantes: de l'internaute aux experts, en passant par le milieu des médias ainsi que les différentes ONG.