



HAL
open science

À la recherche du juste rôle des plateformes exploitant des réseaux sociaux dans la lutte contre la "haine en ligne".

Emmanuel Netter

► To cite this version:

Emmanuel Netter. À la recherche du juste rôle des plateformes exploitant des réseaux sociaux dans la lutte contre la "haine en ligne". : Une comparaison des approches française, allemande et européenne.. Centre de droit privé et de sciences criminelles d'Amiens. Les plateformes d'intermédiation numérique, A paraître. hal-03650659

HAL Id: hal-03650659

<https://hal.science/hal-03650659>

Submitted on 25 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial | 4.0 International License

**À LA RECHERCHE DU JUSTE RÔLE DES PLATEFORMES
EXPLOITANT DES RÉSEAUX SOCIAUX
DANS LA LUTTE CONTRE LA « HAINE EN LIGNE ».
UNE COMPARAISON DES APPROCHES
FRANÇAISE, ALLEMANDE ET EUROPÉENNE.**

Emmanuel Netter
*professeur de droit privé à l'université d'Avignon
LBNC (EA 3788)*

contact @ enetter . fr

1. Nouveaux acteurs, nouvelles règles - Il n'existe pas de rédacteur en chef chez Facebook. L'idée semble évidente aujourd'hui : un réseau social n'est pas un média traditionnel. Le modèle classique de diffusion des idées et des discours était pyramidal : un petit nombre de concepteurs professionnels de contenus et d'amateurs éclairés s'adressait à un vaste auditoire par voie de presse, de radio ou de télévision. Il était alors facile d'exiger qu'un directeur de publication ou un responsable des programmes endosse la responsabilité d'une diffusion illicite¹. Lorsque chaque utilisateur de Twitter, d'Instagram ou de TikTok acquiert la capacité à s'exprimer aussi bien qu'à recevoir les discours d'autrui, on peut imaginer que la pyramide devient sphère. D'un point de vue technique, la plateforme est au centre : sans son infrastructure, la conversation collective cesse aussitôt. Mais déterminer la position qu'elle occupe intellectuellement, socialement, politiquement par rapport aux contenus qu'elle participe à diffuser est une affaire plus complexe.

2. Un ordonnancement automatisé des contenus - Les réseaux sociaux apparaissent comme objet de régulation nouveau, pour l'essentiel au cours des années 2000². À ce moment, le législateur européen a pris conscience de la nécessité de protéger les simples « prestataires intermédiaires », qui assurent le bon fonctionnement d'internet en se contentant de rendre des services purement techniques³. L'un des statuts bénéficiant d'un régime de faveur est celui d'hébergeur, « consistant à stocker des informations fournies par un destinataire du service », par opposition au rôle d'un éditeur, qui dispose d'un véritable pouvoir de décision sur les contenus qu'il affiche en ligne⁴. Les plateformes exploitant des réseaux sociaux se trouveraient à leur aise dans le statut d'hébergeur, dont il faut toutefois reconnaître qu'il a été conçu avant leur montée en puissance, en considération d'acteurs plus passifs : un hébergeur de site web traditionnel se contente de mettre à disposition un espace en ligne, que l'internaute utilise à sa guise, tandis que Facebook ou Twitter agencent, mettent en scène, valorisent certaines publications plutôt que d'autres, dans des « fils d'actualité » qui cessent rapidement d'être exhaustifs et purement chronologiques. Leurs interventions sont cependant automatisées. Elles ne reposent nullement sur des choix politiques ou éditoriaux opérés

1 V. ainsi l'art. 42 de la loi du 29 juillet 1881 sur la liberté de la presse : « Seront passibles, comme auteurs principaux des peines qui constituent la répression des crimes et délits commis par la voie de la presse, dans l'ordre ci-après, savoir : 1° Les directeurs de publications ou éditeurs, quelles que soient leurs professions ou leurs dénominations, et, dans les cas prévus au deuxième alinéa de l'article 6, de les codirecteurs de la publication ; 2° A leur défaut, les auteurs ; 3° A défaut des auteurs, les imprimeurs ; 4° A défaut des imprimeurs, les vendeurs, les distributeurs et afficheurs (...) ».

2 LinkedIn a été lancé en 2003, Facebook (pour sa version accessible hors de certains campus américains) et Twitter en 2006.

3 Section 4 de la 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 »).

4 Art. 14 de la directive précitée. La loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique, qui transpose la directive e-commerce en droit français, définit les hébergeurs en son article 6, I, 2 comme « Les personnes physiques ou morales qui assurent, même à titre gratuit, pour mise à disposition du public par des services de communication au public en ligne, le stockage de signaux, d'écrits, d'images, de sons ou de messages de toute nature fournis par des destinataires de ces services ».

par l'esprit humain : elles sont le fruit de traitements algorithmiques fondés sur le succès passé d'une publication et sur les goûts supposés de l'utilisateur, tels qu'ils sont révélés par son utilisation antérieure de l'outil. Les réseaux sociaux ne sont plus tout à fait des hébergeurs, mais ils ne sont pas pour autant des éditeurs.

3. Le régime de faveur de l'hébergement - En l'absence d'une troisième voie, les juges français et européens ont cependant dû trancher⁵. Se prononçant sur le statut du moteur de recherche de Google, la Cour de Justice de l'Union européenne a posé des critères susceptibles d'être transposés aux médias sociaux : « (le statut d'hébergeur) s'applique au prestataire d'un service de référencement sur Internet lorsque ce prestataire n'a pas joué **un rôle actif** de nature à lui confier une connaissance ou un contrôle des données stockées »⁶. Lorsqu'une requête donnée est adressée au moteur, il n'y a pas de réponse objective possible : retenir certains résultats plutôt que d'autres et les ordonner entre eux force à choisir, mais ces choix reposent généralement sur des automatismes. Considérer que la simple mise en œuvre de ces traitements de données, sans intervention humaine, doit être assimilée à un rôle actif, reviendrait à rendre tous les moteurs de recherche et réseaux sociaux comptables de l'intégralité des contenus qu'ils présentent à leurs utilisateurs, ce qui n'apparaît ni réaliste, ni souhaitable. Peu de temps après, la Cour de cassation a donc appliqué le régime de l'hébergement à un acteur qui « se bornait à structurer et classifier les informations mises à la disposition du public pour faciliter l'usage de son service »⁷.

Mais en quoi consiste précisément ce régime de faveur ? Aux termes de la loi française de 2004 ayant transposé le régime européen (Loi pour la confiance dans l'économie numérique ou LCEN), les personnes qualifiées d'hébergeurs ne peuvent engager ni leur responsabilité civile, ni leur responsabilité pénale du fait des informations stockées « (...) si elles n'avaient pas effectivement connaissance de leur caractère manifestement illicite ou de faits et circonstances faisant apparaître ce caractère ou si, dès le moment où elles en ont eu cette connaissance, elles ont agi promptement pour retirer ces données ou en rendre l'accès impossible »⁸. La situation la plus courante dans laquelle la plateforme acquiert cette connaissance est le signalement par un utilisateur du service.

4. La crainte d'une intervention excessive des plateformes - Il est essentiel de relever que l'adverbe « manifestement » ne figurait pas dans la rédaction d'origine. Il a été ajouté à la suite d'une réserve d'interprétation du Conseil constitutionnel, saisi du contrôle de constitutionnalité a priori de la LCEN. Le Conseil avait décidé que « ces dispositions ne sauraient avoir pour effet d'engager la responsabilité d'un hébergeur qui n'a pas retiré une information dénoncée comme illicite par un tiers si celle-ci ne présente pas **manifestement** un tel caractère ou si son retrait n'a pas été ordonné par un juge »⁹. Ainsi les Sages ont-ils décidé qu'il n'appartenait pas aux sociétés privées exploitant ces services de trancher lorsqu'il existe une incertitude sur la licéité d'un maintien en ligne. Pour le Conseil constitutionnel, leur rôle consiste exclusivement à expurger les espaces numériques des violences en ligne les plus grossières. Dès lors qu'existe un doute sur l'intention humoristique ou parodique du propos ; dès l'instant où, malgré son caractère agressif, il est susceptible de s'inscrire dans les limites normales d'une campagne électorale ou d'un vif débat de société ; sitôt que la censure pourrait porter une atteinte excessive au droit du public à l'information, Twitter, Facebook, Instagram ou Tiktok ne devraient pas intervenir de leur propre chef. Il faudrait que le retrait ait été « ordonné par un juge », seul à même de séparer, au sein des contenus « gris », le bon grain de l'ivraie et de fixer finement les limites de la liberté d'expression dans une société démocratique.

5 V. not. P. Le Tourneau (dir.), *Contrats du numérique*, Dalloz, 2021/2022, 422.31 et s. ; F. Chopin, « Cybercriminalité – Responsabilité pénale des intermédiaires techniques », in *Répertoire IP/IT et communication*, Dalloz, janvier 2020, n° 390 s.

6 CJUE 23 mars 2010, aff. jointes C-236/08 à C-238/08, *Google France SARL et Google Inc. c/ Louis Vuitton Malletier SA*. Souligné par nous.

7 Cass. 1ère civ., 17 février 2011, 09-13.202, *Fuzz*. V. aussi, le même jour, l'arrêt 09-67.896, *Dailymotion*.

8 Art. 6., I, 2 et 3 de la loi précitée du 21 juin 2004.

9 Décision n° 2004-496 DC du 10 juin 2004, considérant n° 9.

5. La crainte d'une intervention insuffisante des plateformes – Puis les années passèrent, le succès des réseaux sociaux grandit et c'est un discours inverse qui commença à s'imposer : ce qu'il fallait craindre n'était pas l'excès de zèle des plateformes, mais leur laxisme. Le PDG de Twitter n'avait-il pas lui-même déclaré, en 2015, que son entreprise était « nulle » lorsqu'il s'agissait de lutter contre le cyberharcèlement¹⁰ ? En 2017, le Ministère de la Justice allemand affirmait qu'« Actuellement, on constate un changement massif du discours social sur la toile et en particulier sur les réseaux sociaux. La culture du débat sur la toile est souvent agressive, blessante et bien souvent haineuse. Par le biais de la criminalité haineuse et d'autres contenus punissables, chacun peut être diffamé en raison de ses opinions, de la couleur de sa peau ou de son origine, de sa religion, de son sexe ou de sa sexualité. Les crimes de haine et autres contenus punissables qui ne peuvent pas être combattus et poursuivis efficacement représentent un grand danger pour la cohabitation pacifique d'une société libre, ouverte et démocratique »¹¹. Le Ministère avait réalisé une série de tests, en signalant aux grandes plateformes plusieurs centaines de contenus qu'il considérait comme manifestement illicites et en examinant leur réaction. Si YouTube avait opéré plus de 90% de suppressions, le chiffre n'était plus que de 39% pour Facebook, et de moins de 1% pour Twitter¹².

6. Les réactions allemande et française - Le Gouvernement allemand s'appuya sur ce constat pour provoquer l'adoption, le 30 juin 2017, de la loi controversée *Netzwerkdurchsetzungsgesetz* (« application du droit sur les réseaux »), qui sera présentée plus loin. Un an plus tard, en France, un rapport *Renforcer la lutte contre le racisme et l'antisémitisme sur Internet* était remis au Premier ministre par un groupe de travail dirigé par la députée Laetitia Avia¹³. Mettant régulièrement en avant « l'exemple allemand », il proposait notamment « de définir un statut particulier d'hébergeur, qui pourrait être dénommé "accélérateur de contenus" pour les réseaux sociaux et les moteurs de recherche les plus utilisés, assorti d'obligations renforcées ». Ainsi proposait-on ouvertement de créer cette « troisième voie » qui semblait manquer à la directive e-commerce... sans que le droit européen ait évolué entre-temps.

7. La réaction européenne - C'est donc fort logiquement que la Commission européenne mit en garde la France contre le risque, si elle adoptait un tel projet, de manquer à ses obligations issues du droit de l'Union¹⁴. Conciliant, l'exécutif européen achevait son courrier de notification en constatant que « La Commission partage l'objectif politique des autorités françaises concernant la lutte contre les contenus illégaux en ligne. Cependant, au vu de l'intention de la Commission et des travaux en cours concernant la proposition et l'adoption d'une législation européenne dans un avenir proche, il est suggéré aux États membres de reporter l'adoption d'initiatives nationales sur le sujet, comme c'est le cas du projet notifié ». Et il est vrai que les initiatives solitaires de la France et de l'Allemagne auraient trouvé une justification, politique à défaut d'être juridique, dans une éventuelle inertie des 27 sur la question. Tel n'est pas le cas : un important projet de règlement relatif à un marché intérieur des services numériques est en préparation, mieux connu sous son acronyme anglais de

10 R. Brunet, « Twitter est "nul" pour éviter le harcèlement, avoue son PDG », article france24.com du 5 février 2015.

11 Traduction libre de l'exposé de la notification du projet de loi NetzDG à la Commission européenne, partie A « Problème et objectifs ».

12 Le site internet destiné à valoriser cette étude n'est plus en ligne, mais les données restent accessibles, à la suite de l'exercice d'un droit d'accès à des documents administratifs, à l'adresse <http://fragdenstaat.de/anfrage/rohdaten-des-monitoring-berichts-von-jugendschutznet/>.

13 K. Amellal, L. Avia et G. Taïeb, rapport remis le 20 septembre 2018, consultable sur <http://www.gouvernement.fr/rapport-visant-a-renforcer-la-lutte-contre-le-racisme-et-l-antisemitisme-sur-internet>.

14 V. la synthèse de ces critiques par M. Rees, « Loi Avia contre la cyberhaine : les critiques de la Commission européenne », article nextinpact.com du 25 novembre 2019. Le document discuté a été mis en ligne deux jours plus tard par le journal : M. Rees, « Loi contre la Cyberhaine : télécharger la lettre incendiaire adressée par la Commission européenne », article nextinpact.com du 27 novembre 2019. On relèvera en particulier que « la Commission conclut qu'il existe un risque que le projet notifié viole les articles 3, 14 et 15, paragraphe 1, de la directive sur le commerce électronique ».

DSA pour *Digital Services Act*¹⁵. Nous ne traiterons pas ici du cas particulier des contenus à caractère terroriste, qui font l'objet d'un règlement spécifique déjà en vigueur¹⁶.

8. Plan – Les lois allemande et française ont en commun la mise en place d'incitations structurelles et puissantes en faveur d'une surcensure : dans le doute sur la licéité d'un contenu, la plateforme a nettement intérêt à le retirer plutôt qu'à prendre le risque de le laisser en ligne. De l'excès de laxisme, on passe ainsi à un excès d'interventionnisme, qui ne semble pas non plus souhaitable dans des sociétés fondées sur le principe d'une liberté d'expression (§1). L'approche européenne, en l'état du projet de directive sur les services numériques, semble davantage propice à la mise en place d'une modération équilibrée. De nombreuses incertitudes subsistent cependant sur les effets concrets que produirait l'entrée en vigueur de ce texte (§2).

§1 – Les excès de législations nationales favorables à une modération agressive

9. Plan – La loi NetzDG se propose d'agir sur les structures des plateformes exploitant des réseaux sociaux¹⁷ (A). La loi française se fonde sur le destin individuel des contenus signalés, mais non retirés promptement : plutôt que de se cantonner à une approche par les « moyens », elle se fonde sur des obligations « de résultat » (B).

A – L'approche allemande par les structures : la loi NetzDG

10. Champ d'application – Le texte s'applique aux sociétés « qui exploitent des plates-formes sur l'Internet dans l'intention de réaliser un profit, qui sont destinées à permettre aux utilisateurs de partager tout contenu avec d'autres utilisateurs ou de le rendre accessible au public¹⁸. Un réseau social est toutefois exempté des principales obligations nées du texte s'il compte moins de deux millions d'utilisateurs enregistrés en Allemagne¹⁹. La notion de contenus illicites est définie par renvoi à une liste d'articles du Code pénal²⁰.

11. Obligation de publier un rapport – Une partie importante de la loi NetzDG contraint les services exploitant des médias sociaux à rendre les processus de modération beaucoup plus transparents. La démarche est vertueuse dans son principe. Ainsi, aux termes de l'article 2.1, « Les fournisseurs de réseaux sociaux qui reçoivent plus de 100 plaintes concernant des contenus illicites au cours d'une année civile sont tenus d'établir tous les six mois un rapport en langue allemande », qui est publié au Journal officiel fédéral ainsi que sur la propre page d'accueil de la plateforme, où il

15 Le projet est consultable en français, dans une version datée du 15 décembre 2020, à l'adresse <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52020PC0825>.

16 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. La caractéristique principale de ce texte est d'imposer une obligation de retrait en une heure, sur injonction de « l'autorité compétente » (article 3.3). C'est l'ARCOM, présentée plus loin dans cette étude, qui devrait être désignée comme autorité compétente aux termes d'une proposition de loi portant diverses dispositions d'adaptation au droit de l'UE en matière de prévention de la diffusion de contenus à caractère terroriste en ligne, actuellement en cours d'adoption.

17 Toutes les citations de la loi NetzDG sont de libres traductions de la version en langue allemande, **dans sa rédaction initiale** telle qu'elle a inspiré le législateur français. Le texte a été modifié depuis.

18 Art. 1.1. Certains services sont cependant exclus : « Les plates-formes proposant des offres journalistiques et éditoriales, dont le prestataire de services est lui-même responsable, ne sont pas considérées comme des réseaux sociaux au sens de la présente loi. Il en va de même pour les plates-formes destinées à la communication individuelle ou à la diffusion de contenus spécifiques ».

19 Art. 1.2. Cela devait concerner au total moins de dix réseaux d'après la notification du projet à la Commission européenne, partie « E.2 Charge de travail pour l'économie ».

20 Art. 1.3 : « Les contenus illicites sont des contenus au sens du paragraphe 1 qui répondent aux critères des articles 86, 86a, 86a, 89a, 91, 100a, 111, 126, 129 à 129b, 130, 131, 140, 166, 184b en liaison avec 184d, 185 à 187, 201a, 241 ou 269 du code pénal et ne sont pas justifiés ».

« doit être facilement reconnaissable, immédiatement accessible et disponible en permanence ». Ce rapport comprend de nombreuses informations, parmi lesquelles : une description générale des diligences de la plateforme pour prévenir la commission de délits sur ses services ; une description des mécanismes permettant de signaler un contenu illicite et des critères employés par les équipes de modération pour prendre leur décision ; le nombre de signalements reçus, avec un classement en fonction du motif ; la description des moyens dévolus aux équipes de modération ; des informations quant au sort réservé aux signalements ; quant au délai ayant séparé le signalement et la décision de retrait du contenu ; une description des mesures prises pour informer tant l'utilisateur ayant signalé le contenu que celui qui est à l'origine de sa publication²¹.

12. Obligation d'efficacité dans le traitement des plaintes – L'article 3 de la loi allemande doit être considéré avec bien davantage de prudence. Certains aspects méritent une approbation sans réserve, comme l'obligation faite aux plateformes d'offrir un mécanisme de signalement clair et facilement identifiable²², de continuer à stocker les contenus retirés pendant un certain délai afin qu'il puisse leur être réservé un traitement pénal²³, ou d'informer correctement les utilisateurs du sens et des motifs des décisions prises²⁴. En revanche, apparaît problématique la forte pression exercée sur les plateformes en matière d'efficacité des retraits : elle n'est pas compensée par une injonction à la prudence, qui viserait à éviter les censures excessives. Au contraire : non seulement un délai relativement bref de 24 heures, quel que soit le contexte, est fixé pour retirer le « manifestement illicite »²⁵, mais on exige même un retrait de l'illicite non manifeste, moyennant un délai simplement allongé à 7 jours²⁶, susceptible d'être encore prorogé dans certaines circonstances particulières²⁷. Ce qui est recherché, c'est « l'efficacité » de la modération et non sa justesse. À cet égard, il est révélateur que la campagne de « testing » du Ministère allemand de la Justice, évoquée plus haut, n'aie consisté qu'à vérifier que des contenus dont l'illicéité était postulée étaient bien retirés. La démarche inverse, consistant à signaler des contenus dérangeants, mais licites pour s'assurer qu'ils étaient bien maintenus en ligne, n'a pas été envisagée. La capacité à ne pas censurer n'est considérée à aucun moment.

13. Sanctions – La loi NetzDG a frappé les esprits par la sévérité des sanctions encourues en cas de manquements aux obligations qu'elle édicte. Sont notamment punis à l'article 4 le défaut de publication du rapport semestriel, sa publication incomplète ou tardive, l'absence ou la mise en place incorrecte de l'interface de signalement des contenus illicites, la gestion incorrecte des signalements ou les défauts structurels d'organisation dans les activités de modération²⁸. L'amende administrative susceptible d'être infligée atteint 50 millions d'euros²⁹.

14. Critiques – Les réticences que suscite ce texte ont été efficacement formulées par le rapporteur spécial des Nations Unies sur la promotion et la protection du droit à la liberté d'opinion et

21 Art. 2.2.

22 Art. 3.1.

23 Art. 3.2.4.

24 Art. 3.2.5.

25 Art. 3.2.2. Il est toutefois précisé : « ceci ne s'applique pas si le réseau social a convenu d'un délai plus long avec l'autorité répressive compétente pour la suppression ou le blocage du contenu manifestement illicite ».

26 Art. 3.2.3.

27 Même texte. « Le délai de sept jours peut être dépassé si : a) la décision sur l'illégalité du contenu dépend de la véracité d'une allégation factuelle ou d'autres circonstances factuelles reconnaissables ; dans de tels cas, le réseau social peut donner à l'utilisateur la possibilité de commenter la plainte avant de prendre une décision ; b) le réseau social confie la décision d'illégalité à un organisme d'autorégulation réglementé reconnu en vertu des paragraphes 6 à 8 dans les sept jours suivant la réception de la plainte et se soumet à la décision dudit organisme ».

28 Art. 4.1.

29 L'art. 4.2 évoque un maximum de 5 millions d'euros, mais le rapport précité *Renforcer la lutte contre le racisme et l'antisémitisme sur Internet* précise que « la législation allemande prévoit une multiplication par 10, dès lors qu'une "infraction administrative" fait l'objet de dispositions légales, ce qui est le cas ici » (note 29).

d'expression de l'époque, M. David Kaye³⁰. Après avoir rappelé la liste pour le moins hétéroclite des « contenus violents » tels qu'ils sont définis par NetzDG³¹, il relève à juste titre : « Une interdiction de la diffusion d'informations fondée sur **des critères vagues et ambigus**, tels que l'"insulte" ou la "diffamation", est incompatible avec l'article 19 du Pacte international relatif aux Droits civils et politiques. La liste des violations est large, et inclut des violations qui n'exigent pas le même niveau de protection. En outre, bon nombre des violations couvertes par le projet de loi **dépendent fortement du contexte**, contexte que les plateformes ne sont pas en mesure d'évaluer »³². Le texte repose en effet sur une conception viciée de ce que l'on peut et de ce que l'on doit attendre des sociétés exploitant des réseaux sociaux. La philosophie sous-jacente est à l'exact inverse de ce qu'exigeait, en France, le Conseil constitutionnel dans sa décision de 2004. Pour le Conseil, il n'était acceptable de s'en remettre à leur appréciation qu'en présence d'une qualification absolument évidente. Le Gouvernement allemand choisit au contraire de leur confier – de force – des qualifications juridiques particulièrement subtiles.

Le rapporteur spécial ajoute : « Le risque semble encore plus élevé si l'on considère les délais stricts de 24 heures et de 7 jours selon lesquels les réseaux sociaux doivent évaluer et supprimer les contenus en violation du droit national. La brièveté des délais, associée à la sévérité des sanctions susmentionnées, pourrait conduire les réseaux sociaux à surréglementer l'expression - en particulier, à supprimer une expression légitime, non susceptible d'être restreinte en vertu du droit des droits de l'homme, par précaution pour éviter les sanctions. Une telle censure préventive porterait atteinte au droit de chercher, de recevoir et de diffuser des informations de toute nature sur l'internet ». Ainsi, non seulement les plateformes sont placées dans un espace de décision trop vaste, qui leur octroie un pouvoir politique qu'elles n'ont aucune légitimité à exercer ; non seulement ces sociétés commerciales sont dénuées de la neutralité et de la compétence technique du magistrat ; mais, si cela ne suffisait pas, elles sont encore l'objet de pressions violentes et univoques en faveur d'un retrait des contenus signalés.

La loi NetzDG pouvait donc difficilement être considérée comme un modèle pour une législation française, tant elle semblait incompatible avec notre bloc de constitutionnalité. Elle a pourtant été non seulement reprise, mais ses travers ont à certains égards été accentués, et ses dangers rendus plus considérables encore.

B – L'approche française par les résultats : la loi contre les contenus haineux sur internet

15. Obligations de moyens – Madame Laetitia Avia, dont on se souvient qu'elle était l'un des coauteurs du rapport « Renforcer la lutte contre le racisme et l'antisémitisme sur Internet » de 2018

30 Lettre au Gouvernement allemand du 1^{er} juin 2017, consultable sur <http://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf>. Les passages cités infra sont librement traduits de l'anglais.

31 « L'avant-projet de loi définit les "contenus violents" comme les contenus visés aux articles 86 (diffusion de matériel de propagande d'organisations anticonstitutionnelles), 86a (utilisation de symboles d'organisations anticonstitutionnelles), 89a (préparation d'une infraction violente grave mettant en danger l'État), 90 (diffamation du président), 90a (diffamation de l'État et de ses symboles), 90b (diffamation anticonstitutionnelle des organes constitutionnels), 91 (encouragement à la commission d'une infraction violente grave mettant en danger l'État), 100a (falsification par trahison), 111 (incitation publique au crime), 126 (atteinte à la paix publique par menace de commettre des infractions), 129 à 129b (formation d'organisations criminelles et terroristes, dans le pays et à l'étranger), 130 (incitation à la haine), 131 (diffusion de représentations de la violence), 140 (récompenser et approuver des infractions), 166 (diffamation des religions, des associations religieuses et idéologiques), 184b (distribution, acquisition, 184b (distribution, acquisition et détention de pornographie infantine), 184d (distribution de spectacles pornographiques par des services de radiodiffusion, de médias ou de télécommunications), 185 à 187 (injure et diffamation), 241 (mise en danger de poursuites pénales par des informations sur une personne) et 269 (falsification de données destinées à fournir des preuves) du code pénal allemand ».

32 Souligné par nous.

a déposé en mars 2019 une proposition de loi ouvertement inspirée de l'exemple allemand³³, qui deviendra la loi du 24 juin 2020 visant à lutter contre les contenus haineux sur internet³⁴. Il n'est donc pas surprenant qu'elle se soit fondée en partie sur une approche par les structures et les procédures. Plusieurs obligations imaginées par les Allemands étaient reprises sous une forme similaire, voire à l'identique. Ainsi prévoyait-on l'implémentation d'un dispositif de notification uniforme ; un accusé de réception délivré à l'auteur d'une notification, suivie d'une information sur le sort réservé à son signalement ; la mise en œuvre de procédures et moyens de modération, humains et technologiques, proportionnés aux besoins ; des dispositifs de recours internes³⁵. Des obligations de transparence étaient édictées, relatives notamment aux recours internes, mais aussi juridictionnels dont disposent les utilisateurs, aux moyens alloués à la modération³⁶. Une personne physique devait représenter les opérateurs étrangers sur le territoire français³⁷. Les conditions générales d'utilisation devaient être formulées « en termes précis, aisément compréhensibles, objectifs et non discriminatoires »³⁸.

16. Le rôle du CSA - Le respect de ces obligations de moyens devait être contrôlé par une autorité administrative indépendante, le Conseil supérieur de l'audiovisuel³⁹. Le choix de confier la supervision des plateformes numériques au régulateur historique des médias traditionnels n'allait pas de soi. La radio ou la télévision ont certes été contraintes d'adapter leurs modes de diffusion à l'ère numérique et d'assurer une présence en ligne de leurs contenus, en direct ou en différé. Par capillarité, le CSA a donc été contraint de s'intéresser à internet, mais nous avons ouvert cette étude en soulignant la profonde différence à la fois technique et philosophique qui sépare une diffusion pyramidale classique de l'information du rôle d'une plateforme opérant un réseau social. C'est donc une compétence entièrement nouvelle qui devait être confiée au CSA. Cela n'a évidemment rien d'impossible, mais laisser penser que cette autorité pouvait capitaliser sur son expérience antérieure était susceptible de créer des malentendus et même des contresens. Toujours est-il que le CSA pouvait mettre en demeure une plateforme insuffisamment diligente quant à sa transparence, ses procédures ou l'adéquation de ses moyens de modération à ses besoins effectifs. Arrivait alors une disposition destinée à marquer les esprits : « Lorsque l'opérateur faisant l'objet de la mise en demeure ne se conforme pas à celle-ci, le Conseil supérieur de l'audiovisuel peut, dans les conditions prévues à l'article 42-7 de la présente loi, prononcer une sanction pécuniaire dont le montant prend en considération la gravité des manquements ainsi que, le cas échéant, leur caractère réitéré, sans pouvoir excéder 20 millions d'euros ou 4 % du chiffre d'affaires annuel mondial total de l'exercice précédent, le montant le plus élevé étant retenu »⁴⁰. Le montant de cette amende administrative s'inspire manifestement des sanctions susceptibles d'être prononcées par les autorités de protection des données dans le cadre du RGPD⁴¹.

17. Obligations de résultat – En droit allemand, l'obligation de retirer rapidement les messages haineux était incluse dans les « obligations de moyens » et était appréciée de façon globale. Tel n'était pas le cas dans la loi Avia. Dans un chapitre trompeusement intitulé « Simplification des dispositifs de notification de contenus haineux en ligne », figure une obligation de retrait en 24h de très nombreux types de contenus⁴². Contrairement au dispositif issu de NetzDG, on vise ici uniquement le « manifestement » illicite : on rappelle qu'il s'agissait, en France, d'une exigence

33 En ce sens, la mention de NetzDG dès les premiers paragraphes de l'exposé des motifs du texte déposé le 20 mars 2019 à l'Assemblée nationale.

34 Nous nous référerons ici à la version définitivement adoptée par l'Assemblée nationale le 13 mai 2020, en précisant d'emblée que certaines dispositions citées sont absentes du texte actuellement en vigueur, en raison de leur censure par le Conseil constitutionnel.

35 Art. 4.

36 Art. 5.

37 *Ibid.*

38 *Ibid.*

39 Art. 7.

40 *Ibid.*

41 Art. 83 du Règlement 2016/679.

constitutionnelle. Par ailleurs, Madame Avia n'a pas repris la proposition, issue de son rapport *Renforcer la lutte contre le racisme et l'antisémitisme sur Internet*, consistant à « mettre en quarantaine » les « contenus gris » - ceux dont la licéité est douteuse – dans l'attente d'une décision de qualification⁴³. Jusqu'ici plus prudente que le texte allemand, la loi française bascule cependant dans des excès spectaculaires au stade de la sanction. Il n'est plus question ici d'une amende administrative sanctionnant une attitude générale de la plateforme, après dialogue puis mise en demeure : **chaque** contenu non retiré dans le délai de 24h en dépit de son appartenance aux catégories listées par le législateur fait encourir une amende de nature pénale et d'un montant de 250.000 euros. En revanche, en cas d'excès de retrait, aucune amende individuelle n'est prévue. Comme dans NetzDG, la pression est donc exercée à sens unique, mais son intensité est plus élevée encore dans le texte français.

18. La décision du Conseil constitutionnel - Dès lors, une censure par le Conseil constitutionnel était largement attendue, et n'a pas manqué d'intervenir⁴⁴. Dans leur décision, les Sages relèvent en premier lieu que n'importe quel individu isolé a le pouvoir de déclencher un signalement qui pourra aboutir, en cas d'inertie, au prononcé de l'amende⁴⁵. Le risque de submerger l'opérateur est évident. Il est relevé, en deuxième lieu, que l'obligation faite aux plateformes de cantonner leur action à l'illicite manifeste entre en flagrante contradiction avec la subtilité de l'examen qu'il leur est demandé d'exercer : « (...) le législateur a retenu de multiples qualifications pénales justifiant le retrait de ces contenus. En outre, son examen ne doit pas se limiter au motif indiqué dans le signalement. Il revient en conséquence à l'opérateur d'examiner les contenus signalés au regard de l'ensemble de ces infractions, alors même que les éléments constitutifs de certaines d'entre elles peuvent présenter une technicité juridique ou, s'agissant notamment de délits de presse, appeler une

42 Article 1^{er} de la loi. L'étude fastidieuse des nombreux renvois peut être avantageusement remplacée par la lecture du considérant 12 de la décision n° 2020-801 DC du Conseil constitutionnel : « Il s'agit des infractions d'apologie à la commission de certains crimes ; de provocation à la discrimination, à la haine ou à la violence à l'égard d'une personne ou d'un groupe de personnes à raison de leur origine ou de leur appartenance ou de leur non-appartenance à une ethnie, une nation, une race ou une religion déterminée ou à raison de leur sexe, de leur orientation sexuelle ou identité de genre ou de leur handicap ou de provocation à la discrimination à l'égard de ces dernières personnes ; de contestation d'un crime contre l'humanité tels qu'ils sont définis par l'article 6 du statut du tribunal militaire international annexé à l'accord de Londres du 8 août 1945 et qui ont été commis soit par les membres d'une organisation déclarée criminelle en application de l'article 9 dudit statut, soit par une personne reconnue coupable de tels crimes par une juridiction française ou internationale ; de négation, de minoration ou de banalisation de façon outrancière de l'existence d'un crime de génocide, d'un autre crime contre l'humanité que ceux précités, d'un crime de réduction en esclavage ou d'exploitation d'une personne réduite en esclavage ou d'un crime de guerre lorsque ce crime a donné lieu à une condamnation prononcée par une juridiction française ou internationale ; d'injure commise envers une personne ou un groupe de personnes à raison de leur origine ou de leur appartenance ou de leur non-appartenance à une ethnie, une nation, une race ou une religion déterminée ou envers une personne ou un groupe de personnes à raison de leur sexe, de leur orientation sexuelle ou identité de genre ou de leur handicap ; de harcèlement sexuel ; de transmission d'une image ou d'une représentation d'un mineur lorsque cette image ou cette représentation présente un caractère pornographique ; de provocation directe à des actes de terrorisme ou d'apologie de ces actes ; de diffusion d'un message à caractère pornographique susceptible d'être vu ou perçu par un mineur ».

43 P. 16 du rapport précité. La mise en quarantaine était une solution temporaire, dans l'attente d'une qualification non pas par un juge, mais par « une autorité de régulation » de type CSA. Nous reviendrons dans le II, A sur le point de savoir s'il doit appartenir à une autorité administrative d'opérer des qualifications pénales individuelles de contenus.

44 Décision n° 2020-801 DC du 18 juin 2020 : *Revue française de droit constitutionnel*, 2021, n° 126, p. 207, note C. Denizeau ; *Daloz IP/IT*, 2020, p. 577, obs. B. Bertrand et J. Sirinelli ; *Dr. pénal*, sept. 2020, p. 12, note F. Safi ; *AJ Pénal*, sept. 2020, p. 407, note N. Droin ; *D.*, 2020, p. 1448, note C. Bigot ; *Gaz. Pal.*, 7 juill. 2020, p. 16, note M. Quémener. Adde L. Castex, K. Favro et C. Zolynski, « La lutte contre la haine en ligne : de l'appel du 18 juin au discours de la méthode », *D.*, 2021, p. 246.

45 Considérant 14. La loi Avia prévoyait certes que « Le fait, pour toute personne, de présenter aux opérateurs mentionnés aux premier et deuxième alinéas du I du présent article un contenu ou une activité comme étant illicite au sens du même I dans le but d'en obtenir le retrait ou d'en faire cesser la diffusion alors qu'elle sait cette information inexacte est puni d'un an d'emprisonnement et de 15 000 euros d'amende » (art. 1^{er}). Mais la démonstration extrêmement difficile du caractère intentionnel d'un tel comportement constitue un obstacle évident à l'efficacité d'une telle incrimination.

appréciation au regard du contexte d'énonciation ou de diffusion des contenus en cause ». En troisième lieu, le délai de 24 est jugé trop bref lorsqu'il est placé en relation avec les deux constats précédents : à savoir, d'une part, avec son déclenchement possible par n'importe quel utilisateur ordinaire et, d'autre part, avec la subtilité de l'examen attendu⁴⁶. En quatrième lieu, le Conseil relève qu'une avalanche de signalements dans un court laps de temps ne constitue pas une cause d'exonération⁴⁷. En dernier lieu est critiquée la sanction de 250.000 euros, « encourue pour chaque défaut de retrait et non en considération de leur répétition »⁴⁸. La conclusion est la suivante : « Il résulte de ce qui précède que, compte tenu des difficultés d'appréciation du caractère manifestement illicite des contenus signalés dans le délai imparti, de la peine encourue dès le premier manquement et de l'absence de cause spécifique d'exonération de responsabilité, les dispositions contestées **ne peuvent qu'inciter les opérateurs de plateforme en ligne à retirer les contenus qui leur sont signalés**, qu'ils soient ou non manifestement illicites »⁴⁹.

La décision mérite approbation en tous points. Sa leçon principale est la suivante : il ne suffit pas au législateur d'affirmer que l'intervention des plateformes se cantonnera à l'illicéité grossière. Il n'est pas suffisant d'apposer l'adverbe « manifestement » dans le texte comme on rédige une clause de style dans un contrat pour faire advenir, comme par magie, un examen équilibré des contenus. Tel n'était pas le cas en l'espèce, les incitations écrasantes en faveur d'un retrait n'étant pas compensées par des injonctions à la prudence et au maintien. Qu'une démocratie moderne puisse occasionnellement chercher à faire taire ceux qui abusent de leur liberté d'expression se conçoit ; qu'elle ne se préoccupe pas de laisser se développer, tout le reste du temps, de libres discours, est préoccupant.

La loi Avia a donc été censurée non seulement s'agissant de l'obligation de retrait en 24h, mais, la majeure partie des dispositions du texte étant emmêlée dans un écheveau de renvois, c'est la quasi-totalité de la loi contre les contenus haineux sur internet qui a été renvoyée aux oubliettes de l'histoire. L'espoir, pour qui espère une approche plus mesurée, vient à présent du niveau européen.

§2 – L'espoir d'une législation européenne favorable à une modération équilibrée

19. Le DSA – L'échelon européen est souvent le plus naturel et le plus efficace lorsqu'il s'agit de réguler des technologies numériques qui, par nature, posent des questions identiques dans tous les pays de l'Union. À cet argument d'opportunité s'ajoute une considération technique : ce que l'Europe avait fait, avec le statut des intermédiaires techniques issus de la directive e-commerce de 2000, elle est le mieux placée pour le refaire. Tel est l'objectif de la proposition de règlement relatif à un marché intérieur des services numériques (DSA). Elle dépasse de loin la question des plateformes de réseaux sociaux, mais les englobe : elle propose une série de qualifications juridiques de plus en plus précises, qui s'emboîtent les unes dans les autres comme des poupées gigognes, chacune étant assortie d'un régime propre. Ainsi certaines règles s'appliquent-elles aux « services intermédiaires », d'autres plus spécifiquement aux « fournisseurs d'hébergement, y compris les plateformes en ligne », d'autres encore aux seules « plateformes en ligne », d'autres enfin aux « très grandes plateformes en ligne ». Les réseaux sociaux, quelle que soit leur taille, relèvent normalement des trois premières catégories. Les mastodontes relèvent de surcroît de la quatrième.

46 Considérant 16.

47 Considérant 17. Le Conseil y critique par ailleurs la rédaction confuse de la seule cause d'exonération prévue par le texte, selon laquelle : « Le caractère intentionnel de l'infraction (...) peut résulter de l'absence d'examen proportionné et nécessaire du contenu notifié ».

48 Considérant 18. Souligné par nous.

49 Considérant 19.

20. Plan - En ce qu'elle repose essentiellement sur des obligations de moyens, une approche par les structures, la philosophie du DSA peut sembler proche de la loi NetzDG, mais l'analogie est trompeuse : cette fois-ci, on vise la mise en place de procédures bien plus saines d'examen des contenus (A). Mais ces progrès ne doivent pas occulter les graves incertitudes qui persistent lorsqu'il s'agit d'identifier précisément les rôles respectifs des plateformes, des régulateurs et des juges (B).

A- Les avancées relatives aux procédures d'examen des contenus

21. Le maintien d'une immunité conditionnée des hébergeurs – Le projet DSA ne remet pas en cause le cœur du statut des intermédiaires techniques tel qu'il résultait déjà de la directive sur le commerce électronique de 2000 : il ne doit peser sur eux aucune « obligation générale de surveiller les informations qu'ils transmettent ou stockent, ou de rechercher activement des faits ou des circonstances révélant des activités illicites »⁵⁰. Une précision nouvelle est cependant apportée, connue sous le nom de « clause de bon samaritain » : afin de ne pas dissuader les opérateurs de prendre, de leur propre chef, des initiatives visant à détecter l'illicite avant qu'il soit porté à leur connaissance par un tiers, il est précisé que cela n'aura jamais pour effet de leur faire perdre leurs exemptions de responsabilité⁵¹. Précisément, les exemptions dont bénéficie un hébergeur sont équivalentes à celles que l'on connaît actuellement. Il n'est « pas responsable des informations stockées à la demande d'un bénéficiaire du service », à la double condition de ne pas avoir « effectivement connaissance de l'activité ou du contenu illicite » et, dès lors qu'il acquiert cette connaissance, d'agir « promptement pour retenir le contenu illicite ou rendre l'accès à celui-ci impossible »⁵². Le délai chiffré promu en Allemagne aussi bien qu'en France n'a donc pas été retenu à ce stade, et c'est la notion de prompt réaction qui prévaudra. Elle n'est pas nécessairement plus laxiste : face à un afflux massif de signalement, la plateforme pourrait bénéficier d'une certaine mansuétude qu'un délai fixe de 24h ne permet pas, ce qui répond aux objections du Conseil constitutionnel. À l'inverse, dans certains cas de contenus troublant gravement l'ordre public et dont l'illicéité est patente, on peut imaginer que 24h soit un délai beaucoup trop long et qu'une plus grande réactivité soit attendue du professionnel. Cette souplesse est donc une force, mais elle octroie corrélativement un important pouvoir à l'autorité en charge de la régulation des plateformes, ce que le DSA appelle « le coordinateur pour les services numériques ».

22. L'illicite qualifié par les pouvoirs publics – Le projet DSA vise ensuite à armer les pouvoirs publics, lorsqu'ils cherchent à obtenir des plateformes d'une part qu'elles agissent contre un contenu illicite, d'autre part lorsqu'elles cherchent à obtenir « des informations », par exemple l'adresse IP utilisée par un utilisateur pour se connecter au service, qui permettra par la suite d'obtenir son identité civile complète en s'adressant au fournisseur d'accès à internet concerné. Désobéir à de telles injonctions serait passible des très lourdes sanctions susceptibles d'être prononcées en cas de méconnaissance du règlement. Ce pouvoir d'injonction est reconnu de la même façon aux « autorités judiciaires ou administratives nationales pertinentes » : il n'est donc pas réservé aux juges.

23. L'illicite qualifié par la plateforme - Laissons de côté l'hypothèse dans laquelle un contenu a été qualifié d'illicite par un magistrat ou une autorité administrative, car, dans de tels cas, la plateforme n'a plus qu'à obéir et sa capacité d'initiative est nulle. Les cas les plus sensibles, et de très loin les plus nombreux statistiquement, sont ceux dans lesquels des messages, images, vidéos auront été signalés par les utilisateurs du service. Il appartiendra alors au réseau social de prendre une décision par lui-même. La procédure d'examen doit dans ce cas être finement encadrée. C'est ici

50 Art. 7.

51 Art. 6.

52 Art. 5.

que le projet DSA se distingue nettement des lois allemande et française. Les deux législations nationales étaient entièrement bâties autour du postulat d'une tendance au laxisme, qui rendait nécessaires des incitations en direction d'une modération plus musclée. Le DSA semble quant à lui beaucoup moins préjuger du dénouement souhaitable de l'examen. La possibilité qu'un contenu ait été signalé à tort et que son maintien en ligne constitue la meilleure solution n'est pas occultée. Plutôt qu'un examen « efficace », au sens de la loi NetzDG, on cherche à favoriser un examen de qualité.

24. Le déclenchement d'un examen par les utilisateurs – Passons rapidement sur les dispositions visant à rendre les boutons de signalement « faciles d'accès et d'utilisation » ainsi qu'à normer le contenu des notifications afin d'en améliorer la prise en charge : on les retrouvait sous des formes proches dans les textes français et allemand⁵³. Le dispositif des « signaleurs de confiance » mérite en revanche une pleine attention. Un signaleur de confiance est une personne disposant « d'une expertise et de compétences particulières aux fins de la détection, de l'identification et de la notification des contenus illicites », qui « représente des intérêts collectifs et est indépendante de toute plateforme en ligne » et « s'acquitte de ses tâches aux fins de la soumission des notifications en temps voulu, de manière diligente et objective »⁵⁴. Ces qualités sont reconnues par un le coordinateur pour les services numériques, qui délivre un agrément. Dès cet instant, les notifications soumises par le signaleur de confiance font l'objet d'un examen et d'une décision « prioritaires »⁵⁵. Ce mécanisme est d'ores et déjà pratiqué par certaines plateformes, dont les équipes de modération considèrent avec une grande célérité les signalements opérés par certaines associations, ou par des utilisateurs individuels dont disposant d'un historique substantiel de notifications fiables⁵⁶. La démarche est particulièrement saine. Les signalements d'utilisateurs ordinaires présentent le risque d'être globalement peu fiables : il est fréquent, dans des moments de tension entre communautés tels qu'une campagne électorale, d'assister à des signalements massifs de propos qui n'ont d'autre défaut que d'avoir été tenus par des adversaires, sans présenter de caractère manifestement illicite voire sans illicéité aucune. S'arc-bouter sur le devoir des plateformes d'examiner ces océans de notifications médiocres dans un délai rigide, comme le faisait la loi Avia, était un non-sens. Mettre en place un circuit de signalements de qualité permettra vraisemblablement des retraits en bien moins de 24h.

25. Les garanties entourant l'examen – Une fois la publication portée à l'attention de la plateforme, le plus difficile reste à faire. Des conflits de droits vont être tranchés : d'un côté, les conditions générales de la plateforme ainsi que les lois et règlements sanctionnant ce qu'il est convenu de regrouper sous l'appellation molle de « contenus haineux » ; de l'autre, le droit fondamental à la liberté d'expression. Le DSA édicte un certain nombre de garanties procédurales, susceptibles d'augmenter la qualité du travail accompli par l'opérateur. Ici encore transparaît la volonté de ne favoriser aucun dénouement plutôt qu'un autre : le retrait n'est intrinsèquement pas plus souhaitable que le maintien. Il convient de prendre la bonne décision.

L'exigence d'un « exposé des motifs » joue un rôle important à cet égard⁵⁷. Il s'agit d'éviter les décisions fondées par exemple, de manière elliptique, sur « la violation de standards de la communauté ». Ainsi, « lorsque la décision concerne des contenus prétendument illicites, [elle doit contenir] une référence au fondement juridique sous-jacent et des explications des motifs pour lesquels ces informations sont considérées comme des contenus illicites sur cette base ». De même, « lorsque la décision se fonde sur la prétendue incompatibilité des informations avec les conditions générales du fournisseur, [elle doit contenir] une référence aux clauses contractuelles sous-jacentes et des explications des raisons pour lesquels ces informations sont considérées comme

53 Art. 14.

54 Art. 19.2.

55 Art. 19.1.

56 V. ainsi le « programme YouTube Trusted Flagger » : <https://support.google.com/youtube/answer/7554338?hl=fr>.

57 Art. 15.

incompatibles avec ces clauses ». À propos de ces conditions générales d'utilisation, relevons qu'un autre article du projet rappelle qu'elles doivent être conformes « aux droits fondamentaux applicables des bénéficiaires du service »⁵⁸. Il ne devrait ainsi plus être possible à Facebook, par exemple, de censurer la publication du tableau « L'origine du monde » de Gustave Courbet ou des campagnes de lutte contre le cancer du sein, au motif que la société serait souveraine pour interpréter ce que signifie l'interdiction des contenus représentant la nudité telle qu'elle résulte de ses conditions d'utilisation⁵⁹. C'est bien la marque de ce que le projet de règlement se soucie autant du maintien en ligne de contenus licites que du retrait des contenus prohibés. Dans ce sens, on relèvera encore que « l'exposé des motifs » doit fournir, « le cas échéant, des informations relatives à l'utilisation de moyens automatisés pour prendre la décision, y compris lorsque cette dernière concerne des contenus détectés ou repérés par des moyens automatisés »⁶⁰. En effet, le recours à des outils de modération automatisés de mauvaise qualité est à l'origine de nombreux retraits abusifs de contenus licites⁶¹.

Une fois la décision motivée rendue, entre en jeu une autre garantie procédurale de première importance : le droit reconnu aux utilisateurs de former des recours. Les plateformes doivent tout d'abord prévoir un « système interne de traitement des réclamations », « permettant d'introduire, par voie électronique et gratuitement, des réclamations contre les décisions [...] adoptées par la plateforme en ligne au motif que les informations fournies par les bénéficiaires constituent un contenu illicite ou sont incompatibles avec ses conditions générales »⁶². L'examen devra avoir lieu « en temps opportun, de manière diligente et objective »⁶³. Les réclamations internes doivent fournir à l'utilisateur censuré une opportunité de s'expliquer. Cette possibilité semble impliquer à elle seule que le réexamen sera conduit par un modérateur humain, mais le projet de règlement le précise de surcroît explicitement : il ne peut pas être le fait exclusif de moyens automatisés⁶⁴.

À ce droit de former un recours interne à la plateforme s'ajoute un mécanisme de « règlement extrajudiciaire des litiges »⁶⁵. Peut demander à être certifié par le coordinateur des services numériques tout organe « impartial et indépendant », disposant d'une expertise s'agissant d'un ou plusieurs types de contenus illicites ou s'agissant de l'application de conditions générales d'utilisation, proposant un règlement des litiges en ligne, rapide, efficace et économique, suivant une procédure claire et équitable⁶⁶. On l'aura compris, le recours à ce tiers n'est pas gratuit, mais si le litige est tranché en faveur de l'utilisateur, la plateforme devra prendre en charge les frais exposés par l'utilisateur⁶⁷. Ainsi sort-on, pour le meilleur ou pour le pire, de l'alternative entre « police privée par les plateformes » et « intervention d'un acteur public (administratif ou judiciaire) », en introduisant une troisième voie : l'arbitrage par une entité publique indépendante des opérateurs de réseaux sociaux. L'*Oversight Board* de Facebook préfigure cette tendance⁶⁸.

26. Les sanctions – Les garanties qui viennent d'être évoquées ne valent pas uniquement en cas de retrait pur et simple des contenus en ligne. Elles sont en effet associées au concept de « modération des contenus », qui couvre « les mesures prises qui ont une incidence sur la disponibilité, la

58 Art. 12.1.

59 P. Signoret, « Censure de L'Origine du monde : une faute de Facebook reconnue, mais pas sur le fond », article lemonde.fr du 15 mars 2018 ; E. Scappaticci, « Facebook censure une campagne contre le cancer du sein », article lefigaro.fr du 28 septembre 2016.

60 Art. 15.2.

61 V. par ex. « "Comment faire pour que les hommes cessent de violer" : sur Twitter, une question qui semble taboue », article franceinter.fr du 25 janvier 2021.

62 Art. 17.1.

63 Art. 17.3.

64 Art. 17.4.

65 Art. 18.

66 Art. 18.2.

67 Art. 18.3.

68 www.oversightboard.com.

visibilité et l'accessibilité de ces contenus illicites ou informations, telles que leur rétrogradation, leur retrait ou le fait de les rendre inaccessibles, ou sur la capacité du bénéficiaire à fournir ces informations, telles que la suppression ou la suspension du compte d'un utilisateur »⁶⁹. Cette approche englobante doit être approuvée. Dans un domaine distinct, mais à plusieurs égards voisins de celui de la haine en ligne, à savoir la diffusion d'informations trompeuses, les plateformes ont pris des mesures de « ralentissement » des contenus, qui doivent être traitées pour ce qu'elles sont : une censure « soft », dont l'emploi peut se révéler justifié ou injustifié, et doit être en toute hypothèse encadré⁷⁰. À l'inverse, la suspension de compte est une mesure plus énergique que le simple retrait de contenu, puisqu'elle empêche l'utilisateur d'accéder au service « pendant une période de temps raisonnable » : elle fait l'objet d'un encadrement spécifique, reposant notamment sur la nécessité d'une répétition des fautes et sur un avertissement préalable⁷¹. Relevons que le même texte prévoit la possibilité de suspendre provisoirement un utilisateur de son droit à signaler des contenus s'il s'est livré « fréquemment » à des notifications « manifestement infondées » : une fois encore, le DSA est à la recherche d'une approche équilibrée, qui protège les utilisateurs des excès de censure aussi bien que de la modération velléitaire⁷².

27. La redevabilité – Comme le RGPD avant lui et à l'instar d'autres projets européens comme le règlement sur l'intelligence artificielle, le DSA est teinté d'une approche générale souvent désignée sous le terme anglais « d'accountability », régulièrement traduit en français par le mot un peu vague de « responsabilisation » ou, de manière plus rare, quoique plus précise et plus intéressante, de « redevabilité »⁷³. L'idée peut être résumée comme suit : l'entreprise « redevable » jouit d'une certaine liberté d'action – qui s'est notamment traduite, en droit des données personnelles, par l'abandon des « formalités préalables » : déclaration des fichiers ou demande d'autorisation à la CNIL. En contrepartie, elle doit s'interroger avec sérieux sur la conformité de ses pratiques à la réglementation, procéder à des auto-diagnostics et se soumettre à des regards externes, documenter ses choix et procédures internes. Tout ceci permet d'identifier les risques qui pèsent sur son activité, puis de maîtriser ces risques par l'adoption de mesures techniques et organisationnelles appropriées.

Dans le DSA, cette approche se manifeste d'abord par l'obligation d'adopter et de publier régulièrement un rapport de transparence extrêmement complet. Il comprend notamment des données quantitatives sur les injonctions reçues des autorités administratives et juridictions et sur les signalements reçus des utilisateurs⁷⁴ ; sur le fonctionnement des organes de règlement extrajudiciaire, les suspensions de compte, le recours à des moyens de modération automatisés⁷⁵.

La logique de redevabilité est accentuée dans le cas des « très grandes plateformes ». Celles-ci doivent ainsi procéder au moins une fois par an à une identification de « tout risque systémique important trouvant son origine dans le fonctionnement et l'utilisation faite de leurs services au sein de l'Union »⁷⁶. Celle-ci porte certes sur « la diffusion de contenus illicites par l'intermédiaire de leurs services », mais aussi sur « tout effet négatif pour l'exercice des droits fondamentaux relatifs

69 Art. 2, p).

70 Par ex. « Facebook veut réduire la visibilité des fausses informations sur les vaccins », article lemonde.fr du 8 mars 2019 ; K. Yurieff, « Twitter won't let you retweet, like or reply to election tweets with warnings on them », article cnn.com du 13 octobre 2020.

71 Art. 20.1.

72 Art. 20.2.

73 En ce sens, L. Castex, K. Favro et C. Zolynski, « La lutte contre la haine en ligne : de l'appel du 18 juin au discours de la méthode », art. préc. : « L'Union européenne étend à ce titre la logique du RGPD des données aux contenus, et entend réguler en transposant l'approche par la compliance à la modération par les plateformes ». V. aussi K. Favro et C. Zolynski, « De la régulation des contenus haineux à la régulation des contenus (illicites) », *Légipresse*, 2019, p. 461.

74 Art. 13.

75 Article 23. Pour les très grandes plateformes, l'article 33 est également applicable, et oblige notamment à aborder dans le rapport les évaluations de risque et audits dont il va être à présent question.

76 Art. 26.1.

au respect de la vie privée et familiale, à la liberté d'expression et d'information »⁷⁷. Une fois de plus, l'approche est équilibrée : dans ce texte, l'insuffisance de censure constitue un risque pour la société, mais l'excès de censure tout autant. Une fois les risques identifiés, obligation est faite aux très grandes plateformes de mettre en place « des mesures d'atténuation raisonnables, proportionnées et efficaces ».

Relève encore à l'évidence de la logique de redevabilité l'obligation, une fois encore réservée aux très grandes plateformes, de procéder tous les ans à un audit indépendant destiné à vérifier le respect de leurs obligations⁷⁸. Les opérateurs devront justifier des mesures prises en réaction à toute critique figurant dans un rapport d'audit.

Signalons enfin une mesure puissante et inédite dans le projet DSA : les très grandes plateformes devront « ouvrir leurs données » aux services de la Commission aux fins de contrôler le respect du règlement. Elles devront également les fournir « à des chercheurs agréés », « à seule fin de procéder à des recherches contribuant à l'identification et à la compréhension des risques systémiques »⁷⁹. Une telle évolution contribuerait à une meilleure compréhension des phénomènes de propagation des contenus, des facteurs de viralité, des effets des contenus illicites sur le débat public, ce qui se révélerait précieux aussi bien en matière de haine en ligne que de diffusion d'informations trompeuses.

28. Une approche équilibrée par les structures – La philosophie du DSA est donc nettement distincte de celle des lois NetzDG et lutte contre les contenus haineux. La logique allemande, maladroitement répliquée par la France, consistait à agir sur l'organisation et les procédures des plateformes, mais avec un unique objectif assez nettement assumé : durcir les pratiques de modération et augmenter significativement les retraits de contenus. Si le texte européen repose lui aussi sur des « obligations de moyens », celles-ci sont orientées beaucoup plus prudemment vers un examen juste et équilibré des publications, qui peut aboutir aussi bien à une censure qu'à un maintien en ligne.

Le constat, jusqu'ici positif, doit à présent être nuancé. Dans les non-dits de ce projet de règlement se cachent d'importantes questions qu'il serait dangereux d'éluder.

B – Les incertitudes relatives aux rôles respectifs des plateformes, des régulateurs et des juges

29. Quel rôle pour les plateformes ? - Les développements qui précèdent ont pu donner au lecteur l'impression que le législateur européen est parfaitement au clair s'agissant des responsabilités que doivent assumer les plateformes dans la censure des contenus en ligne. Il nous semble pourtant que tel n'est pas le cas⁸⁰. Les plateformes sont saisies plus intelligemment grâce à l'intervention de signaleurs de confiance. Leurs décisions sont motivées, susceptibles de recours, prises par des équipes de modération dont les effectifs sont connus, en application de conditions générales clarifiées, à l'aide d'outils automatisés précisément décrits et dont l'usage est circonscrit. Mais quelle est, fondamentalement, la nature de leur contrôle ? Pour le Conseil constitutionnel français, il se limite au manifestement illicite. Les contenus dont la licéité mérite débat sont donc soit retirés à l'issue d'une intervention judiciaire, soit laissés en ligne, sans préjudice d'actions en responsabilité civile ou pénale contre leurs auteurs. Mais le législateur européen, quant à lui, ne dit rien de tel. L'illicite « manifeste » apparaît uniquement comme l'une des exigences s'imposant aux opérateurs de réseaux sociaux souhaitant suspendre des comptes d'utilisateur⁸¹. Pour le surplus, le texte fait

77 *Ibid.*

78 Art. 28.

79 Art. 31.

80 En ce sens, P. Auriel, « La liberté d'expression et la modération des réseaux sociaux dans la proposition de *Digital Services Act* », *Rev. UE*, 2021, 413.

81 Art. 20.1.

référence aux contenus simplement « illicites », dont la définition est renvoyée au droit de l'Union, et surtout aux droits nationaux⁸². Mais pourrait-on considérer avoir harmonisé la matière, si persistent côte-à-côte une conception allemande, dans laquelle l'illicite « discutable » est bien l'affaire des plateformes, moyennant un simple délai de réaction allongé, et la conception constitutionnelle française qui refuse de leur confier les contenus « gris » ? Ce sont deux philosophies radicalement différentes, pour ne pas dire incompatibles, qui cohabiteraient alors. On aurait alors défini avec beaucoup de soin des garanties procédurales, qui encadreraient des examens totalement divergents sur le fond.

On peut par ailleurs se demander si un obstacle technique n'empêchera pas le Conseil constitutionnel de maintenir ses exigences. Sa jurisprudence actuelle, on le sait, résulte de la LCEN, instrument de droit interne transposant la directive européenne sur le commerce électronique. Or, le DSA sera un règlement, en principe d'application directe. L'expérience du RGPD montre cependant que l'adoption d'un règlement n'est pas exclusive d'adaptations du droit interne, qui donnent alors prise au contrôle de constitutionnalité⁸³. Mais il faut ici relever que la France a d'ores et déjà procédé à une « prétransposition » du DSA⁸⁴. Un pays bien connu dans toute l'Union pour sa lenteur à transposer des directives, qui lui a régulièrement valu des condamnations par la CJUE, a tout à coup décidé d'intégrer dans son droit national les prétendus acquis d'un instrument européen dont la négociation n'est pas terminée, au risque d'en compliquer les négociations⁸⁵. Toujours est-il que cela aurait pu constituer une occasion de mesurer la réaction du Conseil constitutionnel, mais, d'une part la réforme française ne constitue qu'un DSA miniature encore éloigné de ce que sera le texte complet⁸⁶, d'autre part le contrôle de constitutionnalité *a priori* dont a bénéficié la loi véhiculant ces dispositions n'a pas porté sur ce proto-DSA, faute de saisine sur ce point du texte⁸⁷. La question reste donc entière. Relevons simplement qu'en 2004, le Conseil constitutionnel s'était saisi du régime juridique des hébergeurs pour lui appliquer la réserve d'interprétation selon laquelle l'intermédiaire ne doit réagir à un « contenu illicite » que si cet illicite est manifeste. Or, cette fois-ci, la règle sera posée directement par l'article 5 du règlement, qui reviendra au critère du contenu (simplement) « illicite ». Elle n'aura techniquement pas besoin d'être reprise en droit interne.

30. Quels rôle pour l'Arcom et pour les juges ? - La loi Avia entendait confier la régulation de la « haine en ligne » au CSA. Celui-ci a depuis fusionné avec la Hadopi pour devenir l'Autorité de régulation de la communication audiovisuelle et numérique (Arcom), qui justifie d'un vertigineux portefeuille de compétences, dont les contenus haineux font bien partie, de même que la lutte contre la manipulation de l'information⁸⁸. Réalisons une brève incursion dans ce domaine voisin que

82 Art. 2, g) : « «contenu illicite», toute information qui, en soi ou de par sa référence à une activité, y compris la vente de produits ou la prestation de services, n'est pas conforme au droit de l'Union ou au droit d'un État membre, quel qu'en soit l'objet précis ou la nature précise ».

83 Loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles.

84 Article 42 de la loi n° 2021-1109 du 24 août 2021 confortant le respect des principes de la République. Sur ce texte, V. not. M. Rees, « Pré-transcription du DSA : une réforme critiquée, mais adoptée en commission au Sénat », article nextinpact.com du 24 mars 2021 (l'auteur soulignait que le Gouvernement avait agi par voie d'amendement, sans étude d'impact ni avis du Conseil d'État) ; E. Dreyer, « Lutte contre les discours de haine et les contenus illicites en ligne dans la loi du 24 août 2021 : de nouvelles obligations pour les plateformes sous le contrôle du CSA », *Légipresse*, 2021, p. 470 ; X. Bioy, « La loi confortant le respect des principes de la République et la liberté d'expression », *AJDA*, 2021, p. 2084.

85 Pour comprendre comment cette initiative solitaire de la France a pu aboutir sans opposition apparente : M. Rees, « Le DSA à la française sauvé par des coups de fil de l'Elysée à Prague », article nextinpact.com du 7 septembre 2021. La République tchèque avait en effet transmis un « avis circonstancié » à la Commission européenne, estimant à juste titre que la réforme française était contraire au principe du pays d'origine, comme l'avait révélé contexte.com dans son briefing numérique du 2 septembre 2021. Cet avis circonstancié a été retiré à la suite des pressions françaises.

86 Les obligations de moyen ainsi introduite figurent dans un nouvel article 6-4 de la LCEN.

87 Décision n° 2021-823 DC du 13 août 2021. Aucune des différentes saisines n'aborde l'article 42.

88 La fusion des deux autorités résulte de la loi n° 2021-1382 du 25 octobre 2021 relative à la régulation et à la protection de l'accès aux œuvres culturelles à l'ère numérique.

constituent les nouvelles trompeuses, pour constater qu'en cette matière, le législateur français a déjà demandé aux plateformes de faire preuve d'une transparence accrue, dont le régulateur se fait ensuite publiquement le relais⁸⁹. Le bilan dressé par l'autorité en juillet 2020 se félicite de la suspension de millions de comptes et de millions de contenus par les plateformes, agissant tantôt par l'intermédiaire d'automatismes, tantôt par l'intermédiaire d'équipes humaines de modération⁹⁰. Or, il est beaucoup plus rare qu'en matière de haine en ligne que les fausses nouvelles relèvent de l'illicéité manifeste⁹¹. Elles reposent bien souvent sur une base factuelle dont découlent ensuite des interprétations qui sont, elles, fallacieuses. Il peut être nécessaire de disposer de compétences spécialisées pour démontrer leur fausseté. Enfin, les inexactitudes qu'elles comportent peuvent être commises de bonne foi, ce qui exclut l'intention de manipuler autrui. Il est donc rare que de tels contenus relèvent d'un illicite « qui crève les yeux ». Il faut en conclure que d'ores et déjà, en France, dans un domaine distinct de celui de la haine en ligne, mais voisin, il est implicitement demandé aux plateformes de procéder à des retraits de contenus « gris », sous les encouragements de l'Arcom.

Par ailleurs, il semble difficile de considérer que le contrôle exigé par le DSA de la part du « coordinateur pour les services numériques » pourra se cantonner, en France comme ailleurs, au respect d'obligations globales de moyens, sans jamais dégénérer en une qualification de contenus précis. Comment pourrait-on en effet juger de la pertinence des procédures d'examen des opérateurs, ou du sérieux avec lequel ils ont identifié et limité des risques systémiques, sans adosser tôt ou tard ce jugement général à des paniers de contenus « témoins » dont on affirmera qu'ils ont été correctement ou incorrectement traités par la plateforme ? Or, cela suppose de les qualifier. Cette vraisemblable prérogative de la future ARCOM était explicitement imaginée dans le rapport Avia. Pour les auteurs de cette étude, l'AAI serait chargée « d'aider les opérateurs à qualifier les contenus haineux douteux (« gris), notamment en réponseur à leurs demandes par des avis »⁹². La procédure était comparée à un « rescrit fiscal »⁹³. L'analogie nous paraît trompeuse : par ses interprétations, l'administration fiscale ne met en cause que la capacité de l'État, dont elle est une émanation, à recouvrer l'impôt, tandis que les « rescrits » de l'ARCOM compromettraient les droits des tiers, auteurs de contenus ou victimes de la « haine en ligne ». Remarquons que dans sa décision de 2004, le Conseil constitutionnel n'envisageait pas la possibilité d'une qualification des contenus par l'administration, la seule alternative présentée résidant dans une illicéité manifeste ou dans le retrait « ordonné par un juge ». Certes, les décisions de l'ARCOM seront susceptibles de recours devant le Conseil d'État. Mais on relèvera d'abord que ce n'est pas la même chose que de confier une qualification juridique à l'administration sous le contrôle du juge et de la confier directement à un juge. On ajoutera ensuite que cela revient à confier au juge administratif des questions ressortissant initialement à la compétence de l'ordre judiciaire et du juge pénal. Une telle modification s'inscrirait, il est vrai, dans un mouvement bien plus vaste par lequel la matière numérique a été progressivement placée dans les mains des autorités administratives indépendantes et, par leur intermédiaire, dans celles du Conseil d'État : le droit des données personnelles aura constitué un modèle qui ne cesse d'être suivi depuis lors.

Un mécanisme pourrait être imaginé, par lequel l'ARCOM a la possibilité d'interroger le juge judiciaire lorsqu'il s'agit *in fine* de déterminer si une infraction pénale a été commise. Une alternative consisterait à considérer qu'il n'est pas utile d'interroger le juge le plus compétent et le plus légitime si la question qui est posée est simplement celle, grossière et à laquelle il est donc plus

89 Art. 11 de la loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information .

90 CSA, Lutte contre la diffusion de fausses informations sur les plateformes en ligne. Bilan de l'application et de l'effectivement des mesures mises en œuvres (sic) par les opérateurs en 2019, juillet 2020, disponible sur arcom.fr.

91 Sur l'intéressante question consistant à savoir si la diffamation doit automatiquement être considérée comme « non manifestement illicite » : T. Amico et P.-H. Baert, « L'office du juge confronté au refus des hébergeurs de retirer des contenus manifestement diffamatoires », AJ Pénal, 2021, p. 510.

92 Rapport précité, p. 26.

93 Note 41 du rapport.

aisé de répondre, de savoir si une infraction pénale a été « manifestement » commise. Cela nous ramène au débat fondamental : sera-ce bien la nature de l'obligation des plateformes ?

À supposer que la régulation des opérateurs achève de devenir une matière purement administrative, et que le juge pénal n'occupe pas même la position d'interlocuteur de l'ARCOM, observons pour finir qu'il est seul compétent et légitime lorsqu'il s'agit de sanctionner les utilisateurs. La régulation de la haine en ligne ne peut en effet se cantonner à escamoter les maux sans jamais en traiter les causes. Soustraire une publication (manifestement) haineuse au regard du public n'est évidemment pas inutile pour maintenir l'ordre social. Mais encore faut-il mettre fin au sentiment d'impunité qui habite souvent les internautes et qui s'explique à titre principal, non par le fragile pseudonymat à l'abri duquel ils croient s'exprimer, mais par l'indigence des moyens de police et de justice consacrés à cette forme de délinquance. Des moyens humains, matériels et des circulaires de politique pénale seront nécessaires. Ils sont les indispensables compléments d'une régulation équilibrée des plateformes.