

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

La cancel culture, la technologie et le rôle des plateformes à l'aune du principe de la liberté d'expression

Bontridder, Noémi; Pouillet, Yves

Published in:
Journal des Tribunaux

Publication date:
2022

Document Version
le PDF de l'éditeur

[Link to publication](#)

Citation for pulished version (HARVARD):

Bontridder, N & Pouillet, Y 2022, 'La cancel culture, la technologie et le rôle des plateformes à l'aune du principe de la liberté d'expression', *Journal des Tribunaux*, numéro 6915, pp. 657-670.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Doctrines

La *cancel culture*, la technologie et le rôle des plateformes à l'aune du principe de la liberté d'expression,
par N. Bontridder et Y. Pouillet 657

Jurisprudence

■ Procédure pénale - Police de la circulation routière Déchéance du droit de conduire - Jugement par défaut - Opposition dans le délai extraordinaire, déclarée recevable - Effets juridiques - Mise à néant rétroactive de la condamnation intervenue
C. const., 25 novembre 2021,
note de C.Y. 671

■ Mise en état judiciaire - Dérégulation conventionnelle - Écartement (non)
Cass., 18 novembre 2021 672

■ Application de la loi dans le temps - Loi de procédure - Application immédiate - Article III.26, § 2, CDE
Cass., 21 octobre 2021 672

■ Jugement par défaut - Exécution provisoire - Autorisation - Motivation spéciale (art. 1397, al. 2, C. jud.)
Civ. Hainaut, div. Charleroi, 3^e ch.,
11 mai 2022 673

Chronique

La vie du palais - Bibliographie - Coups de règle.

Bureau de dépôt : Louvain 1
Hebdomadaire, sauf juillet et août
ISSN 0021-812X
P301031

Journal des tribunaux

<https://jt.larcier.be>
5 novembre 2022 - 141^e année
35 - N° 6915
Georges-Albert Dal, rédacteur en chef

Doctrines

La *cancel culture*, la technologie et le rôle des plateformes à l'aune du principe de la liberté d'expression¹

« Je ne veux pas que ma maison soit murée de toutes parts, ni mes fenêtres bouchées, mais qu'y circule librement la brise que m'apportent les cultures de tous les pays »

Mahatma Gandhi, *Young India*, 1921

Expression d'origine américaine, la *cancel culture* désigne un mouvement multidimensionnel consistant en substance à ostraciser une personne considérée comme ayant tenu des propos ou posé des actes jugés offensants. Après avoir mis cette pratique en perspective, l'article propose, dans une première partie, quelques réflexions sur la façon dont la technologie et, en particulier, celle utilisée par les plateformes, peut accroître le phénomène et ses effets négatifs. Analysant dans une seconde partie la *cancel culture* à l'aune de la liberté d'expression et du numérique dans un État de droit, les auteurs s'interrogent sur la possibilité d'un changement du modèle économique de la toile et examinent les enjeux pour le droit de veiller à une démocratie respectueuse de la parole de chacun et balançant les intérêts poursuivis par les uns et les autres.

Introduction : du concept et de sa réalité au plan de la contribution

1. La « *cancel culture* »², en français « culture de l'annulation » ou « culture de l'effacement », peut être définie comme une « tentative collective de nuire à la réputation, aux moyens de subsistance et à la présence sociale de personnes, de produits, d'émissions de télévision, de célébrités, d'employés, etc., pour avoir violé une norme idéologique particulière »³. Ce phénomène, dont l'expression nous vient des États-Unis, s'est imposé ces dernières années en Europe dans le prolongement des mouvements #MeToo⁴ ou #BalanceTonPorc et Black Lives Matter, déclenchés en réaction aux violences sexuelles faites aux femmes et aux atteintes mortelles de personnes noires par des policiers blancs⁵. Ont ainsi été mises en évidence des atteintes aux droits élémentaires de certaines fractions de la population et s'exprime aujourd'hui une colère qui se manifeste en des actes publics de contestation (par exemple, la peinture rouge recouvrant les statues des colonialistes tel Léopold II, des manifestations contre le racisme ou contre les violences policières), et des revendications (par exemple, des campagnes et pétitions exigeant le déboulonnage des statues des

(1) La présente contribution fut présentée par les deux auteurs lors de la journée d'étude « La *cancel culture* face aux stéréotypes, à la mémoire nécessaire et à la liberté d'expression : l'exemple des dessins de presse » organisée par la Commission belge francophone et germanophone de l'UNESCO, le 4 octobre 2021. Elle a été mise à jour le 7 juillet 2022. Y sont repris quelques passages et idées développées par les auteurs dans des articles précédents. Voy. à cet égard, Y. POULLET et N. BONTRIDDER, « L'Union européenne et la régulation de la désinformation », *R.D.T.I.*, n° 81, Bruxelles, Larcier, 2021, pp. 29-52 ; et N. BONTRIDDER et Y. POULLET, « The role of artificial intelligence in disinformation », *Data & Policy*, 3, E32, 2021, <https://www.doi.org/10.1017/dap.2021.20>.

(2) Concernant l'apparition de cette expression, voy. A. ROMANO, « Why we can't stop fighting about cancel culture - Is cancel culture a mob mentality, or a long overdue way of speaking truth to power ? », *Vox*, 25 août 2020, <https://www.vox.com/culture/2019/12/30/20879720/what-is-cancel-culture-explained-history-debate>.

(3) M. SLICK, « What is Cancel Culture ? », *CARM*, 5 janvier 2021, <https://carm.org/social-justice/what-is-cancel-culture/> (traduction libre).

(4) Ce mouvement a débuté en 2007 et est particulièrement connu depuis octobre 2017 à la suite de l'affaire *Weinstein*.

(5) En particulier depuis la mort de George Floyd en mai 2020.



LES ALIMENTS

Nicole Gallus

Quentin Detienne pour le droit social

Nicole Gallus offre ici un commentaire approfondi de toutes les règles applicables à la matière des obligations alimentaires. Quentin Detienne propose, quant à lui, une pertinente analyse du droit à l'intégration sociale et à l'aide sociale.

> Tiré à part du Répertoire

Notarial

396 p. • 270,00 € • Édition 2022

orders@larcier.com

Lefebvre Sarrut Belgium SA

Boulevard Baudouin 1^{er}, 25 • B-1348 Louvain-la-Neuve

Tél. 0800/39 067 - Fax 0800/39 068



colonialistes). La presse et les caricaturistes⁶ se voient également dans le collimateur de la « cancel culture », de même que l'opéra⁷, le cinéma⁸ et la littérature⁹. Se multiplient aussi les appels à boycott de conférences¹⁰. Des personnalités publiques et professeurs d'université ou de l'enseignement secondaire¹¹ sont la cible de virulentes critiques pour des propos jugés non conformes. Une campagne sur les réseaux sociaux à l'encontre d'une personne peut ainsi réclamer et obtenir son licenciement, ou même son assassinat lorsque le radicalisme s'en mêle, comme ce fut le cas pour le défunt professeur Samuel Paty¹².

Selon l'acteur LeVar Burton¹³, plutôt que de *cancel culture*, il faudrait parler de « conséquence culture ». En effet, ce phénomène est inextricablement lié à un appel à la responsabilité des personnes mises en cause. Sans avoir encore été labellisée « cancel culture », cette idée d'effacement voire d'annulation était à l'origine un outil permettant aux communautés marginalisées d'affirmer leurs valeurs contre les personnalités publiques qui conservaient le pouvoir et l'autorité même après avoir commis des actes répréhensibles. À présent, ces appels à la responsabilité, selon A. Romano, constituent pour leurs auteurs un moyen de défendre leurs valeurs dans une sphère sociale et publique dans laquelle les célébrités et politiciens continuent d'exercer une influence considérable malgré leurs propos ou leurs actes¹⁴.

2. Dans le milieu de la culture ainsi qu'à l'université, nombreux sont ceux qui s'inquiètent d'une censure allant à l'encontre de la liberté d'expression et de la liberté académique. Le grand coup de semonce des réactions face au phénomène fut la tribune publiée dans le magazine américain *Harper's* en juillet 2020¹⁵ et signée par 150 personnalités parmi lesquelles on retrouve des écrivains, des artistes et des journalistes. Celle-ci dénonce l'« intolérance à l'égard des opinions divergentes, [le] goût pour l'humiliation publique et l'ostracisme, [la] tendance à dissoudre des questions politiques complexes dans une certitude morale aveuglante », en faveur d'un « conformisme idéologique »¹⁶.

L'expression « cancel culture » a été récupérée par la droite américaine qui, selon certains, utilise cette expression péjorative pour pouvoir continuer à insulter et employer des termes racistes. En effet, l'expression est reprise en boucle par la droite conservatrice qui ne cesse de crier à la censure. En conséquence, les répliques s'enchaînent et on ne sait plus très bien qui est la victime et qui est le coupable.

Prenons le mouvement #MeToo, qui encourage les femmes à dénoncer publiquement sur les réseaux sociaux leurs agresseurs souvent nommément. En contrepartie, les hommes se sentent menacés et se disent à leur tour victimes de harcèlement féminin injuste et fantasment une action collective concertée. L'accusation de conspiration est

alors adressée aux féministes en général là où devrait plutôt être constaté le fait que nombre de femmes observent les mêmes choses et dénoncent les mêmes agissements¹⁷, se ralliant individuellement à un message porteur de leur souffrance.

3. — Il y a lieu de noter que les messages de *cancel culture* sont largement appréciés par les plateformes de réseaux sociaux (Twitter, Facebook, Instagram, YouTube...) dans la mesure où de tels messages se révèlent attractifs pour les internautes : il est constaté que les messages fondés sur la polarisation des opinions et sur l'indignation des internautes circulent six fois plus vite que les messages informationnels de nos médias traditionnels. Les promoteurs de la *cancel culture* à l'instar des auteurs de désinformation font souvent usage des moyens technologiques pour atteindre leur but. Feront ainsi l'objet de notre premier chapitre des réflexions sur la façon dont la technologie et, en particulier, celle utilisée consciemment par les plateformes, peut accroître le phénomène et ses effets négatifs.

Les réseaux sociaux donnent précisément aux auteurs de messages de *cancel culture* la possibilité de juger et de condamner au vu et au su de tous les internautes sans procès ni pitié. Et ces messages échangés sur les réseaux sociaux contribuent encore un peu plus au clivage de la société. En outre, la crainte d'une dénonciation publique à la suite d'un message émis, dont la teneur est parfois détournée de son contexte, ou qui n'est cité que partiellement, engendre un climat de peur et entraîne des risques d'autocensure. Et ce d'autant plus que certains observent un glissement d'un appel à la responsabilité vers une punition sans pardon, le public étant hermétique à toute éventuelle remise en question ou évolution venant de la personne critiquée¹⁸.

D'emblée, on le pressent, la *cancel culture* soulève pour le juriste un dilemme. D'une part, elle est l'émanation de la liberté d'expression des internautes. Restreindre la possibilité de véhiculer des messages y compris ceux contribuant à la *cancel culture* apparaît comme une limitation de la liberté d'expression, principe sacré de nos sociétés démocratiques. Cette possibilité de restriction serait d'autant plus critiquable qu'elle peut émaner des plateformes elles-mêmes voire de la technologie mise en place par ces plateformes et constitue dès lors le résultat d'une justice privée, en dehors de tout contrôle de la justice rendue par nos tribunaux. D'autre part, le phénomène de *cancel culture* peut disqualifier l'expression d'autrui, voire l'empêcher par crainte de représailles. Ne faut-il pas dès lors redouter une « normalisation » de la parole qui s'exprime sur les réseaux sociaux ? Dans quelle mesure cette réalité ne doit-elle pas également être prise en compte ? Faut-il dès lors limiter la liberté d'expression de certains au nom de la liberté d'expression du plus grand nombre et de chacun ? L'enjeu est pour le droit de veiller à une démocratie respectueuse de la

(6) Voy. la décision du *New York Times* de supprimer les dessins de presse dans son édition internationale de juin 2019, ainsi que les excuses présentées par la directrice de la rédaction du journal *Le Monde* en janvier 2021 pour avoir publié un dessin de Xavier Gorce (https://www.lemonde.fr/le-monde-et-vous/article/2021/01/19/a-nos-lecteurs_6066802_6065879.html).

(7) La fin de l'opéra *Carmen* de Bizet a ainsi été réécrite par le metteur en scène Leo Muscato (ce n'est plus Carmen qui meurt mais c'est elle qui tue Don José, son ancien amant jaloux). Sur la *cancel culture* dans le milieu de l'opéra, voy. S. FORT, « Cancel Culture & Confrontation des temps », *forumopera*, 18 octobre 2021, <https://www.forumopera.com/actu/cancel-culture-confrontation-des-temps>.

(8) Certaines œuvres doivent ainsi être restreintes ou remises en contexte sous peine d'être supprimées, parce qu'elles livreraient des propos jugés racistes, sexistes ou heurtants. C'est le cas du film « Autant en emporte le vent », mais aussi de films Disney tels « Les Aristochats », « Dumbo », « Peter Pan », « La Belle et le clochard », ou encore « Aladdin ».

(9) Le roman *Les dix petits nègres*

d'Agatha Christie paraît aujourd'hui dans une version révisée, sans le mot « nègre » et sous un nouveau titre : *ils étaient 10*.

(10) Voy. par exemple le communiqué de l'Union syndicale étudiante et du Cercle féministe de l'Université libre de Bruxelles (ULB), « Les réactionnaires de *Charlie Hebdo* invitée-s en grande pompe à l'ULB ! », 12 février 2020, <https://use.be/les-reactionnaires-de-charlie-hebdo-invite/>.

(11) L'enseignement secondaire désigne en France l'ensemble des cours enseignés au collège et au lycée.

(12) Voy. à ce sujet le communiqué de presse publié sur le site des Nations unies : « Experts de l'ONU : Honorez la mémoire de Samuel Paty, enseignant français assassiné, en défendant les droits et en confrontant le fondamentalisme », 15 octobre 2021, <https://www.ohchr.org/fr/press-releases/2021/10/honour-slain-french-teacher-samuel-paty-defending-rights-defying>.

(13) Voy. son tweet (publié le 26 avril 2021) : https://twitter.com/levarburton/status/1386709523407130624?ref_src=twsrc%5Etfw%7Ctw-camp%5Etfw%7Ctw-term%5E1386709523407130624%7C

7Ctwgr%5E%7Ctwcon%5E-1_&ref_url=https%3A%2F%2Ffillreporter.com%2Fwatch-levar-burton-explains-that-consequence-culture-is-for-everyone-99206.

(14) A. ROMANO, « The second wave of "cancel culture" : How the concept has evolved to mean different things to different people », *Vox*, 5 mai 2021, <https://www.vox.com/22384308/cancel-culture-free-speech-accountability-debate>. En ce sens, l'historienne Marie Peltier indique que « [l]a *cancel culture* est née dans un contexte américain. Depuis une vingtaine d'années, on assiste à une expression émergente de la voix des minorités. Qu'elles soient d'orientations sexuelle, de genre, culturelle, elles prennent de plus en plus de place et revendiquent une place dans le débat public. Elles s'attaquent à la parole des dominants, des privilégiés ». M. PELTIER, *Obsession, dans les coulisses du récit complotiste* (2018). Voy. <https://www.rtf.be/article/la-cancel-culture-quand-le-harceleur-devient-victime-10561604>.

(15) *A Letter on Justice and Open Debate*, *Haper's* magazine, 7 juillet 2020, <https://harpers.org/a-letter-on-justice-and-open-debate/>.

(16) Selon les auteurs de la tribune,

« [o]n renvoie des rédacteurs en chef pour avoir publié des articles controversés ; on retire des livres sous le prétexte d'un manque d'authenticité ; on empêche des journalistes d'écrire sur certains sujets ; on enquête sur des professeurs à cause des œuvres littéraires qu'ils citent en classe ; un chercheur est renvoyé pour avoir fait circuler un article scientifique dûment examiné par des pairs ; et on limogé des dirigeants d'organisation pour des erreurs qui ne sont parfois que des maladroites ». (traduction publiée dans le journal *Le Monde* : https://www.lemonde.fr/idees/article/2020/07/08/mark-lilla-margaret-atwood-wynton-marsalis-notre-resistance-a-donald-trump-ne-doit-pas-conduire-a-dogmatisme-ou-a-la-coercition_6045547_3232.html).

(17) M. PELTIER, *Obsession, dans les coulisses du récit complotiste* (2018). Voy. <https://www.rtf.be/article/la-cancel-culture-quand-le-harceleur-devient-victime-10561604>.

(18) A. ROMANO, « The second wave of "cancel culture" : How the concept has evolved to mean different things to different people », *Vox*, 5 mai 2021, <https://www.vox.com/22384308/cancel-culture-free-speech-accountability-debate>.

parole de chacun et balançant dès lors les intérêts poursuivis par les uns et les autres. Face à ce dilemme, que peut dire le droit ? Ce sera l'objet du second chapitre.

1 Le rôle amplificateur de la technologie dans le phénomène de *cancel culture*

A. Réflexion liminaire : *Cancel culture*, désinformation et désinterprétation

4. Récemment, le phénomène de désinformation sur les réseaux sociaux a été épinglé par de nombreux médias et fait l'objet d'initiatives de régulation. Ces initiatives ont été prises tant par les pouvoirs publics que par les opérateurs privés actifs sur la toile comme plateformes d'information. Même si notre propos se focalise sur le phénomène de *cancel culture*, il importe de noter que la *cancel culture* peut consister en un fait de désinformation telle que cette notion est définie par les textes européens : « On entend par désinformation, les informations dont on peut vérifier qu'elles sont fausses ou trompeuses, qui sont créées, présentées et diffusées dans un but lucratif ou dans l'intention délibérée de tromper le public et qui sont susceptibles de causer un préjudice public. Par préjudice public on entend les menaces aux processus politiques et d'élaboration des politiques démocratiques et aux biens publics, tels que la protection de la santé des citoyens de l'Union, l'environnement ou la sécurité. La désinformation ne comprend pas les erreurs de citation, la satire, la parodie, ni les informations et commentaires partisans clairement identifiés »¹⁹. La désinformation est ainsi à distinguer de la mésinformation, à savoir l'information fautive, inexacte ou trompeuse partagée par des personnes qui ne la reconnaissent pas comme telle²⁰. L'étude commanditée par la Commission LIBE du Parlement européen et publiée en 2019²¹ retient quatre éléments dans la définition du concept de « désinformation » : premièrement, l'utilisation d'un contenu soit faux, soit prêtant à confusion (*misleading*), soit utilisant des techniques non éthiques de persuasion ; deuxièmement, l'intention de tromper par de tels messages à des fins de profit ou d'intérêt politique ; troisièmement, l'impact sur des sujets d'intérêt général (élections, environnement, santé publique...) troublant le fonctionnement de la société ; et, quatrièmement, l'utilisation stratégique de la technologie (y compris de systèmes d'intelligence artificielle) comme mode de dissémination des messages.

La *cancel culture*, nous dira-t-on, porte sur des faits « publics » décrits ou qualifiés de réels et pas nécessairement sur des faits erronés et n'a pas pour but de calomnier. Cependant, si le constat part de faits réels ou supposés tels, c'est l'éclairage apporté à ces faits et/ou des rapprochements contextuels avec d'autres faits, qui posent problème et peuvent amener à une « misleading information » (élément constitutif de la désinformation) et dans certains cas peuvent troubler le fonctionnement de nos sociétés (autre élément de la désinformation), comme quand la *cancel culture* a pour but de discréditer un homme politique ou l'action gouvernementale en matière de lutte contre une pandémie (voy. le mouvement antivaccin).

5. Tout comme la Commission européenne distingue la mésinformation de la désinformation²², le journaliste Z. Aleem propose de distinguer la mésinterprétation (interprétation incorrecte du message) de la désinterprétation (interprétation incorrecte du message due au manque d'intention de le comprendre)²³. Il met en avant que sur *Twitter*, les internautes désinterprètent les messages de manière contradictoire, antisociale et en vue de les exploiter afin de renforcer leurs propos. En effet, selon lui, les gens n'interprètent pas complètement les messages mais les rangent sans réflexion dans l'une ou l'autre case : bon/mauvais ; socialement acceptable/problématique ; avisé/hors de propos ; mon équipe/l'ennemi²⁴. Ceci témoigne d'une polarisation de la société et d'une transformation du débat public en un champ de bataille ou un dialogue de sourds.

Précisément, les plateformes ne sont pas adaptées aux débats ni aux échanges de grandes idées. Le lecteur d'un message ne peut pas prédire le flux d'une conversation ni extrapoler les opinions de l'auteur à partir de sa seule déclaration. Sur les réseaux sociaux, les internautes se trouvent ou se tiennent dans une posture très affirmative où l'objectif est de convaincre des spectateurs ou lecteurs dans un jeu de validation sociale. Cette attitude aboutit à la confrontation et exclut le dialogue. C'est pourquoi tout discours est subordonné à l'activisme, « ce qui élimine la possibilité d'ambiguïté publique, d'ambivalence, d'idiosyncrasie, d'auto-interrogation »²⁵.

Ces éléments psychologiques expliquent partiellement du moins le phénomène qualifié de « *cancel culture* », qui entend effacer ou censurer les propos rangés dans la case « problématique ». Pour poursuivre le propos, nous nous interrogeons sur une autre facette de l'explication de ce clivage : en quoi le fonctionnement des plateformes de réseaux sociaux joue-t-il un rôle dans la polarisation des débats et dans le phénomène de *cancel culture* ?

(19) Commission européenne, « Lutter contre la désinformation en ligne : une approche européenne », (Communication) COM(2018) 236 final.

(20) Voy. le rapport du « High Level Group of Experts on Disinformation », publié par la Commission européenne (Directorate-Ge-

neral for Communication Networks, Content and Technology) : *A multi-dimensional approach to disinformation : Report of the independent High level Group on fake news and online disinformation*, Luxembourg, Publications Office of the European Union, 2018, disponible en ligne sur <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>, p. 10.

La Commission LIBE du Parlement européen (*Disinformation and propaganda - impact on the functioning of the rule of law in the EU and its Member States*, février 2019, <http://www.europarl.europa.eu/supporting-analyses>, p. 26) reprend la classification suivante proposée par les auteurs Claire Wardle et Hossein Derakhshan (*Information disorder : Toward an interdisciplinary framework for research and policy making*, Council of Europe report DGI(2017)09) :

www.europarl.europa.eu/supporting-analyses, p. 26) reprend la classification suivante proposée par les auteurs Claire Wardle et Hossein Derakhshan (*Information disorder : Toward an interdisciplinary framework for research and policy making*, Council of Europe report DGI(2017)09) :

Types of information disorders	Definition	Example
Misinformation	When false information is shared, but no harm is meant	During the 2016 US presidential elections, a tweet about a « rigged » voting machine in Philadelphia was shared more than 11.000 times. It was later established that the original tweet was a mistake made by a voter who had failed to follow the instructions exhibited on the voting machine.
Disinformation	When false information is knowingly shared to cause harm	During the 2017 French presidential elections, a duplicate version of the Belgian newspaper <i>Le Soir</i> was created, with a false article claiming that Emmanuel Macron was being funded by Saudi Arabia.
Mal-information	When genuine information is shared to cause harm	Examples include intentional leakage of a politician's private emails, as happened during the presidential elections in France.

(21) Commission LIBE du Parlement européen, *Disinformation and propaganda - impact on the functioning of the rule of law in the EU and its Member States*, op cit. (note 20), pp. 14 et s.

(22) Pour rappel, il s'agit d'information fautive, inexacte ou trompeuse

conçue, présentée et promue pour causer intentionnellement un préjudice public ou dans un but lucratif.

(23) Ze. ALEEM, « The things we do and do not say - Notes on the impossibility of talking online and rise of disinformation », *What's Left*, 2019, <https://zeeshanaleem.substack.com/>

p/twitter-bad-faith-disinterpretation.

(24) Ainsi, sur Twitter, « commentators are constantly being characterized as believing things they don't believe, and entire intellectual positions are stigmatized based on associations with ideas that they don't have any substantive affiliation

with, often just because they don't appear to fit into classic left-right or liberal-left binaries ».

(25) Z. ALEEM, « The things we do and do not say - Notes on the impossibility of talking online and rise of disinformation », op cit. (note 23).

B. L'expression publique agressive et attentatoire : un phénomène ancien mais décuplé par la technologie

6. Tout comme le phénomène de désinformation, la volonté individuelle ou collective de combattre telle ou telle opinion n'est pas un phénomène nouveau. Nothomb, homme politique belge, écrivait il y a presque 150 ans : « Quel est l'homme public ou de notoriété, souvent même le citoyen le plus obscur, qui n'ait souffert des coups de ces sycophantes modernes et n'ait été blessé dans ses sentiments les plus intimes ou dans les êtres qui lui sont plus chers que lui-même ? Qui alors n'éprouve contre ces plumes indignes et leurs odieuses diatribes un mouvement de révolte intérieure, et cédant à un légitime ressentiment, ne se sent porté à demander la réparation la plus sûre et la plus prompte ? »²⁶.

N'est par conséquent pas récente la peur d'exprimer des opinions non conformes à des normes sociales prédominantes. En 1974, la sociologue allemande Elisabeth Noelle-Neumann indiquait²⁷ que « [p]our ne pas se retrouver isolé, un individu peut renoncer à son propre jugement. C'est là une condition de la vie dans une société humaine ; s'il en allait autrement, l'intégration serait impossible. Cette peur de l'isolement (non seulement la peur qu'a l'individu d'être mis à l'écart, mais aussi le doute sur sa propre capacité de jugement) fait, selon nous, partie intégrante de tous les processus d'opinion publique. Là est le point de vulnérabilité de l'individu ; c'est là que les groupes sociaux peuvent le punir de ne pas avoir su se conformer. Il y a un lien étroit entre les concepts d'opinion publique, de sanction, et de punition ». L'autrice décrit dès lors un modèle interactionniste du processus de formation de l'opinion publique : « L'opinion partagée s'affirme toujours plus fréquemment, et avec plus d'assurance ; on entend l'autre de moins en moins. Les individus perçoivent ces tendances, et adaptent leurs convictions en conséquence. L'un des deux camps en présence accroît son avance pendant que l'autre recule. La tendance à s'exprimer dans un cas et à garder le silence dans l'autre, engendre un processus en spirale qui installe graduellement une opinion dominante ». C'est ce mécanisme psychosocial qu'elle nomme « la spirale du silence ».

Ainsi, l'opinion publique évolue dans le temps comme suit : « une minorité convaincue de sa domination future et, par suite, disposée à s'exprimer, verra son opinion devenir dominante, si elle est confrontée à une majorité doutant que ses vues prévalent encore dans le futur, et donc moins disposée à les défendre en public. L'opinion de cette minorité devient une opinion qu'on ne peut désormais contredire sans courir le risque de quelque sanction. Elle passe ainsi du statut de simple opinion d'une faction à celui d'opinion publique »²⁸.

7. Suivant l'analyse par la politologue P. Norris des résultats d'enquêtes menées auprès des personnes concernées²⁹, dans la plupart des pays « en développement », où les valeurs sociales conservatrices continuent de prévaloir comme opinion majoritaire, les professeurs d'université et étudiants ayant des opinions progressistes peuvent ressentir des pressions pour se conformer aux normes et valeurs sociales prédominantes. À l'inverse, dans de nombreuses sociétés postindustrielles en Europe occidentale et en Amérique du Nord, l'équilibre de l'opinion publique sur un large éventail de questions a progressivement évolué dans une direction plus libérale et progressiste sur le plan social, pour devenir l'opinion publique majoritaire. Ainsi, les partisans des valeurs socialement conservatrices perdent leur statut hégémonique sur les campus, voire dans la société. Par conséquent, pour éviter l'isolement social, les universitaires de droite ayant des croyances et des attitudes conservatrices sont susceptibles de ressentir des pressions croissantes pour se conformer à l'évolution des valeurs sociales.

Le passage d'un statut hégémonique vers un statut minoritaire expliquerait dès lors en partie du moins pourquoi la droite conservatrice ressent de manière accrue un phénomène qu'elle dénonce et nomme « cancel culture » — ce sentiment d'effacement de son opinion auparavant dominante. Ce phénomène aurait tout aussi bien pu être dénoncé il y a quelques décennies par la gauche minoritaire d'alors, mais n'ayant pas encore eu voix au chapitre, elle ne s'est pas insurgée d'une perte de pouvoir, ou n'avait tout simplement pas encore les moyens de se faire entendre.

8. Bien que le conformisme social soit, comme nous venons de le démontrer, loin d'être un phénomène nouveau, il y a lieu à notre époque de prendre en compte une évolution toute récente qui bouleverse à bien des égards la formation de l'opinion des individus et la conduite des débats, à savoir les technologies de l'information et de la communication (TIC). En effet, les avancées technologiques et l'omniprésence grandissante des TIC permettent la diffusion des contenus agressifs ou attentatoires à une vitesse et une ampleur jamais égalées, souvent à un public ciblé comme nous le montrons. Les répercussions sont d'autant plus conséquentes que le message circule rapidement. Ce message, qui circule au gré des destinataires devenant souvent eux-mêmes émetteurs, se décontextualise, se déforme et s'amplifie. Ci-après, nous analysons deux éléments propres à ces avancées technologiques qui contribuent au phénomène de *cancel culture* : (1) la pérennité des messages et des sources ; et (2) la structure des plateformes.

1. La pérennité des messages et des sources

9. L'accessibilité aisée, la persistance des messages publiés sur la *web*, la puissance et la convivialité des modes d'interrogation et de recherche sur la toile permettent de retrouver des contenus anciens ou non qui, sortis de leur contexte y compris culturel, sont alors sujets à une virulente critique pouvant mener à la *cancel culture*. En effet, des messages peuvent être retrouvés pour nourrir un argumentaire de *cancel culture*, et les moteurs de recherche jouent un rôle prépondérant dans l'accessibilité d'anciennes informations sans reflet de leur relative virulence temporelle et culturelle.

Cette absence de relativité culturelle va jusqu'à inciter Disney à censurer certaines scènes jugées offensantes eu égard au contexte culturel actuel. Or, « [r]éécrire les classiques en coupant ou modifiant des passages pointés du doigt ne reviendrait pas seulement à les dénaturer, ce serait céder à une véritable négation de l'histoire. Cacher ces offenses que l'on ne saurait voir conduirait à occulter les progrès accomplis pour instaurer *de facto* une utopie postmoderne proche du cauchemar de George Orwell »³⁰.

C'est également du fait de la pérennité des messages publiés sur la toile que d'anciens *tweets* homophobes de l'humoriste Kevin Hart ont refait surface et suscité de vives réactions lorsqu'il fut annoncé comme l'hôte de la cérémonie des Oscars 2019, à la suite de quoi il décida de ne pas animer la cérémonie³¹. De même, des *tweets* racistes et homophobes publiés dix ans auparavant par la journaliste politique américaine Alexia McCammond lui ont coûté son poste d'éditrice en chef de *Teen Vogue*³². Tous deux avaient auparavant déjà adressé leurs excuses et exprimaient avoir évolué depuis lors. C'est pourquoi certains parlent d'une évolution allant d'un appel à la responsabilité à une punition sans pardon³³.

2. La structure des plateformes³⁴

10. Suivant l'analyse du professeur de littérature comparée G. Shullenberger, les algorithmes de *Facebook* et *Twitter*, qui encouragent et récompensent des comportements particuliers, jouent sur nos instincts

(26) Nothomb, Rapport fait au nom de la section centrale à propos de propositions de lois sur « Dommages et intérêts et visites domiciliaires en matière de presse », 6 juin 1879, *Doc. parl.*, Chambre, n° 162 (1879).

(27) E. NOELLE-NEUMANN, « The spiral of silence : a theory of public opinion », *Journal of Communication*, 24 : 43-54, 1974. Traduit par Gilles Achaiche, Dorine Bregman, Daniel Dayan.

(28) *Ibid.*

(29) P. NORRIS, « Closed Minds ? Is a

“Cancel Culture” Stifling Academic Freedom and Intellectual Debate in Political Science ? » *HKS Faculty Research Working Paper Series R.W.P20-025*, août 2020, <https://www.hks.harvard.edu/publications/closed-minds-cancel-culture-stifling-academic-freedom-and-intellectual-debate>.

(30) C. CHELEBOURG, « Racisme, stéréotypes, appropriation culturelle : faut-il censurer les classiques Disney ? », *Le journal du dimanche*, 23 octobre 2020, <https://www.le-jdd.fr/Societe/racisme-stereotypes-appropriation-culturelle-faut-il-censurer-les-classiques-disney-4000711>.

(31) E. ST. JAMES, « Kevin Hart is out as Oscars host amid controversy over homophobic jokes », *Vox*, 7 décembre 2018, <https://www.vox.com/culture/2018/12/7/18130239/kevin-hart-oscars-host-2019-jokes-tweets-homophobic>.

(32) K. ROBERTSON, « Teen Vogue Editor Resigns After Fury Over Racist Tweets », *The New York Times*, 10 mai 2021, <https://www.nytimes.com/2021/03/18/business/media/teen-vogue-editor-alexi-mccammond.html>.

(33) Voy. les réflexions introductives, *supra* n° 3.

(34) Voy. G. SHULLENBERGER, « Human Sacrifice and the Digital Business Model », *Tablet*, 20 juillet 2020, <https://www.tabletmag.com/sections/science/articles/sacrificial-games-cancel-culture>.

les plus primaires et notre désir de créer des spectacles de violence symbolique et de sacrifice. Il explique cela de la manière suivante.

Sur les réseaux sociaux, les internautes qui partagent une publication cherchent généralement à recevoir des partages, « likes », « followers », etc., parce que la plateforme nous incite à percevoir leur accumulation comme une preuve de l'aura de la personne. Notamment, règne une apparence d'égalité sur les réseaux sociaux puisque chacun peut accumuler ces « points » sans égard à son statut dans la vie réelle³⁵. Cependant, cette apparence d'égalité entre en contradiction avec l'inégalité de fait qui y réside : tandis que sur Twitter et Instagram, le nombre moyen de « followers » par compte est respectivement 707 et 150, les comptes des élites en amassent des millions. En conséquence, la personne *lambda* perd au jeu et nourrit un sentiment de frustration et d'envie. Ce sentiment peut générer des conflits, d'autant plus que la récompense (les partages, *likes* et *followers*) est attribuée non aux compétences ou activités de la personne mais directement à celle-ci ou à son profil qui n'est rien d'autre que sa représentation numérique. Les récompenses ou le manque de récompenses semblent dès lors valider ou invalider la personne concernée.

Il existe néanmoins un autre moyen de satisfaction que cette compétition individuelle. Il s'agit de l'effusion collective, souvent provoquée par l'outrage. Une telle effusion peut consister tant en une lutte collective contre une injustice (comme avec le mouvement des « black squares »³⁶), qu'en une hostilité commune envers une victime semblant le mériter (ayant fait quelque chose pouvant être vu comme blessant, gênant ou jugé divergent). Ce dernier point est conforme au phénomène du bouc émissaire³⁷, selon lequel les groupes en proie à la compétition et à la rivalité se tournent vers une victime (le bouc émissaire) comme forme de pacification. Selon R. Girard, l'unité sociale était maintenue via des rituels tels que le sacrifice dans les sociétés traditionnelles, avant que le système judiciaire moderne entende résoudre les conflits internes via des mécanismes organisés de médiation. Dans un espace tel que le *web*, où un système de pacification institutionnalisé fait défaut, le phénomène du bouc émissaire refait donc surface.

De surcroît, le fonctionnement des plateformes de réseaux sociaux encourage de telles agressions en ligne puisque l'utilisateur qui lance « la première pierre » peut se voir hautement récompensé par une flopée de partages, de *likes* et de *followers*.

Shullenberger met en avant que l'équivalent des prêtres des sociétés religieuses traditionnelles, qui canalisait la violence en des expériences communes qui unifiaient la collectivité, sont actuellement les ingénieurs. Et le but de ces derniers n'est pas la cohésion sociale ni le bien-être de la communauté, mais bien le profit. Or, précisément, les spectacles de victimisation accroissent l'audience et donc le revenu des plateformes³⁸. Il en conclut que les plateformes n'encouragent en rien le conformisme social mais bien au contraire une symbiose entre déviance et conformité permettant de perpétuer de tels spectacles d'agressivité qui augmentent l'engagement des internautes. Ainsi, les techniques développées par les plateformes pour accroître l'engagement des internautes et donc les revenus générés, participent à la polarisation des débats et au phénomène qualifié de « cancel culture »³⁹.

C. De l'économie de l'attention aux technologies utilisées par les plateformes comme outils décuplant la portée et la diffusion des messages de *cancel culture* et de désinformation

1. L'économie de l'attention et le « business model » des plateformes

11. L'économie de l'attention est, selon les termes repris sur Wikipédia, « une branche des sciences économiques et de gestion qui traite de l'attention et de son contrôle comme d'une ressource rare en prenant appui sur les théories économiques afin de problématiser le fonctionnement de marchés dans lesquels l'offre est abondante (et donc économiquement dévalorisée) et la ressource rare devient le temps et l'attention des consommateurs ». Cette économie se base sur les théories de H. Simon, depuis prix Nobel d'économie, exprimées pour la première fois en 1971 comme suit : « [D]ans un monde riche en informations, l'abondance d'informations entraîne la pénurie d'une autre ressource : la rareté devient ce que consomme l'information. Ce que l'information consomme est assez évident : c'est l'attention de ses receveurs. Donc une abondance d'informations crée une rareté de l'attention et le besoin de répartir efficacement cette attention parmi la surabondance des sources d'information qui peuvent la consommer »⁴⁰.

12. Cette théorie trouve une application particulièrement intéressante dans le monde du *web* et des plateformes d'information et de communication. En effet, le monde du *web* est un monde de surinformation dans lequel les algorithmes utilisés par la plateforme serviront de guide. Dans la mesure où les revenus des plateformes sont, eu égard à la gratuité (apparente) des services offerts aux internautes, générés par la publicité, il est important d'attirer et de retenir l'attention de ces derniers par des messages adaptés à leur personnalité préalablement « profilée » (créant ainsi des « bulles de filtres »⁴¹), qu'ils seront tentés de répercuter auprès de leurs « amis » ou de leur propre audience (créant des « chambres d'écho »). Pour ce « profilage », les plateformes ont développé des systèmes d'intelligence artificielle qui opèrent à partir des traces⁴² laissées par l'internaute (vues, liens, *likes*, traces [*logs*]), son environnement soit propre aux services de la plateforme (mesures d'audience, *PageRank*, nombre d'amis, recommandation), soit déduit de l'utilisation de services d'entreprises liées à la plateforme (par exemple, pour *Facebook*, des données venant de l'utilisation de *WhatsApp*, filiale comme elle de *Meta* ou d'entreprises ayant accepté de partager ses données avec *Facebook*) voire de données externes anonymes ou non (par exemple le revenu ou le niveau d'éducation moyen des habitants de la localité de l'internaute).

Ces méthodes de traçage et de profilage se basent donc sur des technologies d'IA pour pouvoir cibler les messages publicitaires à adresser à des internautes en particulier, dans un souci d'efficacité et de rentabilité du service offert par les plateformes. Elles permettent de profiler de manière de plus en plus précise les internautes et de les appâter via ce qu'il est convenu d'appeler des *nudges*⁴³ afin de maximiser leur consommation des services de la plateforme et donc des revenus publicitaires. Il s'agit de mécanismes propres à des modèles d'affaire dits de « capitalisme libertarien », qui sans jamais imposer, recommandent et conseillent le client « libre » de choisir la voie proposée, c'est-à-dire la plus profitable à l'opérateur.

(35) Par opposition à la vie virtuelle.

(36) <https://sosoier.lesoir.be/blackout-tuesday-le-mouvement-qui-envahit-les-reseaux-sociaux-de-photos-noires>.

(37) Cette théorie a été développée par l'anthropologue et philosophe René Girard : <https://iep.utm.edu/girard/>

?_ga=2.105364739.1050555066.1656585497-25803953.1656585497#H3.

(38) « The gamified mechanisms that precipitate us toward indignation against enemies also drive our continued use of the platforms. The more of us that are transfixed by spectacles of victimization, the greater the revenue the platform brings in. Like a bloodthirsty god, the platform bu-

siness feeds off of sacrifice ».

(39) En ce sens, de Coorebyter indique que « [l]e problème ne réside pas, comme le disent certains idéologues de droite, dans un monopole de la bien-pensance qui régnerait sur le Net, mais plutôt dans le fait que la chasse électronique aux idées qui choquent percolent en dehors du Net et contribue à assécher le débat public non virtuel, qui prend de plus en plus la forme d'un dialogue de sourds, caractérisé par une violence croissante ». V. DE COOREBYTER, « L'Internet : démocratie ou démagogie », in Y. POULLET (éd.), *Vie privée, liberté d'expression et démocratie dans la société numérique*, coll. Cahier du CRIDS, n° 47, Bruxelles, Larcier, 2020, p. 249.

(40) H. SIMON, « Designing Organizations for an Information-Rich World », in M. GRENBERGER (éd.), *Computer, communications and the public interest*, Baltimore MD : The John Hopkins Press, 1971, pp. 37-72. Plus récemment, lire l'ouvrage de Y. Citton, *Pour une écologie de l'attention*, Paris, éd. du Seuil, dl 2014, cop. 2014, 312 p.

(41) « La "bulle de filtres" ou "bulle de filtrage" (de l'anglais : *filter bubble*) est un concept développé par le militant d'Internet Eli Pariser. Selon Pariser, la « bulle de filtres » désigne à la fois le filtrage de l'information qui parvient à l'internaute par différents filtres ; et l'état d'« isolement intellectuel » et culturel dans lequel il se retrouve quand

les informations qu'il recherche sur Internet résultent d'une personnalisation mise en place à son insu » (définition Wikipédia).

(42) Pour E. KESSOUS (*L'attention au monde : sociologie des données personnelles à l'ère numérique*, Paris, Armand Colin, 2012, 315 p.), nous sommes passés avec le *web* d'un marketing de la segmentation à un marketing des traces. Il ajoute : « L'économie des plateformes Internet repose en quelque sorte sur un échange implicite entre un service gratuit et personnalisé et des dépôts d'attention » (Wikipédia).

(43) Le fait de donner un petit coup de coude, d'attirer l'attention.

2. Les technologies utilisées

13. Sans vouloir être exhaustifs, citons deux technologies utilisées, l'une par les plateformes à l'appui de leur modèle d'affaire : le micro-ciblage et, l'autre, par des opérateurs du web connus ou inconnus pour assurer la dissémination de messages ayant une portée polémique : les *bots* sociaux. Nous ajouterons les *deepfakes*, manipulation par les technologies du numérique d'images, de sons ou de textes, qui peuvent avoir pour but de créer de fausses informations.

a. Du micro-ciblage au *real-time bidding*

14. Le microciblage qui permettra le profilage des internautes exige la collecte d'informations mais surtout la possibilité de corréler les informations collectées auprès d'une ou plusieurs sources. Pour opérer de tels rapprochements, diverses technologies peuvent être utilisées. Ainsi, le suivi peut être fondé sur les *cookies* du navigateur (*Browser cookie based tracking*) et permettre, via des hyperliens invisibles, l'envoi de *cookies* (témoins de connexion) par des tiers ou dans le cadre de conventions d'échange entre opérateurs des données générées par les *cookies* des tiers (*third party tracking*) : « les entreprises intègrent du contenu — comme des réseaux publicitaires, des widgets de médias sociaux et des scripts d'analyse de sites web — dans les sites de la première partie, que les utilisateurs visitent directement. Cette relation d'intégration permet à la tierce partie de réidentifier les utilisateurs sur différents sites web et de créer un historique de navigation »⁴⁴. On connaît également les méthodes de suivi reposant non sur les *cookies* mais utilisant l'empreinte numérique du navigateur (*Browser fingerprinting*). Cette méthode « s'appuie sur des fonctionnalités et des spécificités du navigateur et du système d'exploitation pour obtenir une empreinte numérique du navigateur d'un utilisateur qui peut être utilisée pour corréler les visites de l'utilisateur sur les différents sites web »⁴⁵.

Les données collectées — messages échangés y compris les *like*, pratiques de *surfing*, choix opérés d'informations ou de services et tout

autre critère pouvant caractériser l'internaute — seront assemblées et structurées dans des mégadonnées (des *big data*). Les systèmes d'IA de la plateforme opéreront des corrélations au sein de ces *big data* et, en tenant compte du rapprochement entre les cas présents dans la base de données et le cas analysé, pourront prédire les préférences ou le comportement pouvant être attendu(s) d'un internaute. En d'autres termes, pourront ainsi être sélectionnés les individus à qui montrer le message, qu'il soit publicitaire ou d'opinion, et espérer d'autant plus qu'il y porte attention et le partage.

15. Récemment, certaines sociétés ont développé des techniques dites de *real time bidding* (système d'enchères en temps réel). Ces systèmes permettent à des plateformes (*Interactive Advertising Bureau (IAB)*, *Google Portal*) en temps réel de croiser des données venant de lieux multiples, afin d'offrir des publicités ou des messages encore mieux ciblés à la personne concernée. Il s'agit, en temps réel, d'utiliser les informations collectées par les sites membres de la plateforme à propos de l'utilisateur d'un de ces sites pour évaluer l'intérêt de lui présenter une certaine publicité, et ainsi valoriser au mieux l'emplacement publicitaire pour cet utilisateur particulier. En 100 millisecondes, les informations ainsi produites sont proposées à de nombreux services publicitaires via un système d'enchères fournissant des informations sur l'utilisateur (catégorie d'âge, profil, centres d'intérêt) afin de permettre à ces services de faire une offre éclairée⁴⁶.

16. Les méthodes de traçage, originellement développées afin de diffuser les publicités de manière plus efficace aux internautes, sont maintenant largement utilisées pour la diffusion ciblée de tout contenu en ligne. En effet, plus les internautes passent de temps sur un site ou sur une plateforme, plus les compagnies publicitaires rémunèrent ce diffuseur de contenu⁴⁷. Ainsi, les algorithmes des plateformes que sont *Facebook*, *Twitter* et *YouTube*, affectent les résultats que les internautes cherchent activement ainsi que les contenus auxquels ils sont exposés de manière passive. Or, cet écosystème a pour effet de directement amplifier la diffusion d'informations dommageables⁴⁸, et participe

(44) J. AKERS e.a., *Technology-Enabled Disinformation : Summary, Lessons, and Recommendations*, Technical Report UW-CSE, 2 décembre 2018, <https://arxiv.org/abs/1812.09383> (traduction libre) « companies embed content — like advertising networks, social media widgets, and website analytics scripts — in the first party sites that users visit directly. This embedding rela-

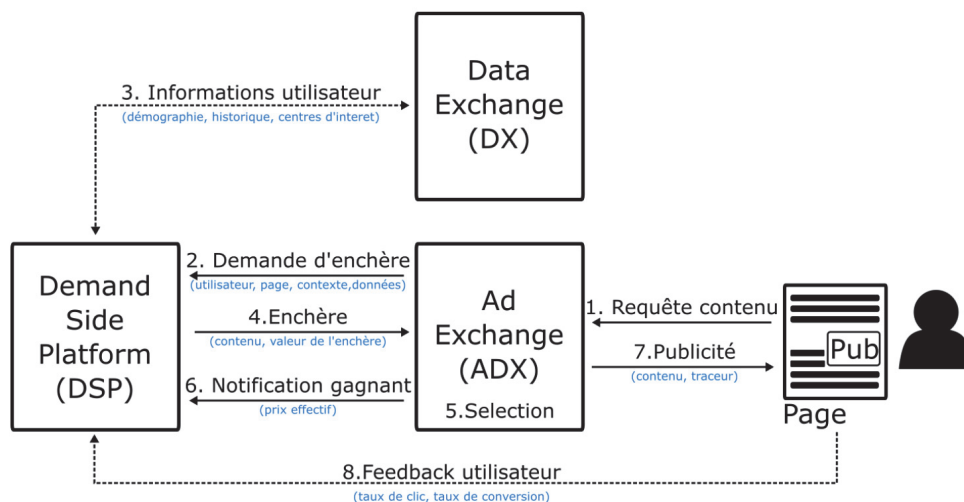
tionship allows the third party to re-identify users across different websites and build a browsing history ».

(45) *Ibid.* (traduction libre) « relies on browser-specific and OS-specific features and quirks to get a fingerprint of a user's browser that can be used to correlate the user's visits across websites ».

(46) Le principe général du *real time bidding* est le suivant (c) Benjamin

Poilvé, « Les enchères en temps réel (RTB), un système complexe », Laboratoire d'Innovation Numérique de la CNIL (LINC), 14 janvier 2020, <https://linc.cnil.fr/fr/les-encheres-en-temps-reel-rtb-un-systeme-complexe>. À propos de la compatibilité du *real time bidding* et du RGPD, lire la décision de l'Autorité de protection des données belge (chambre contentieuse) du 2 février 2022 (décision sur

le fond 21/2022 accessible sur le site de l'APD : <https://www.autoriteprotectiondonnees.be/citoyen/publications/decisions>). Cette décision est actuellement en appel devant la Cour des marchés, qui a tout récemment, le 8 septembre 2022, envoyé quelques questions préjudicielles à la Cour de justice de l'Union européenne.



(47) Les revenus publicitaires de *Facebook* en 2020 atteignaient ainsi 84,2 milliards de dollars. Voy. <https://www.statista.com/statistics/267031/facebook-annual-revenue-by-segment/>.

(48) Pour illustration, ce qui fut nommé « le cercle vicieux de YouTube »,

peut être défini comme suit : 1. les contenus semant la discorde performant mieux ; 2. le système d'IA les promeut afin d'optimiser le temps passé sur la plateforme et ; 3. puisque ce type de contenu est plus regardé, les créateurs de contenus en créent davantage. Voy. *The Toxic Potential of*

YouTube's Feedback Loop, CADE Tech Policy Workshop, 17 novembre 2019, <https://www.youtube.com/watch?v=Et2n0J0OeQ8&feature=youtu.be>. En ce qui concerne *Facebook*, Mark Zuckerberg a illustré les effets néfastes de la stratégie d'engagement de l'entreprise avec le gra-

phique suivant ((c) note publiée par Mark Zuckerberg : <https://www.facebook.com/notes/mark-zuckerberg/blueprint-for-content-governance-and-enforcement/> 10156443129621634/)

donc activement à l'ambiance incendiaire qui règne sur les réseaux sociaux.

b. Les bots sociaux

17. Les *bots*⁴⁹ sociaux (*social bots*) contribuent également à la diffusion ciblée et à grande échelle des messages semant la discorde. Il s'agit de comptes utilisateurs entièrement ou partiellement automatisés qui opèrent sur les réseaux sociaux et qui sont programmés pour communiquer et imiter le comportement humain. Ces *artificial speakers* peuvent être programmés pour publier du contenu adapté à la communauté dans laquelle ils sont censés s'intégrer. Une fois que leur profil est crédible et semble digne de confiance, ils sont utilisés pour diffuser des messages afin de manipuler l'opinion de manière efficace et à grande échelle. Ainsi, comme le disent deux auteurs américains, « nous vivons dans un monde avec des orateurs artificiels avec un réel impact. Les *bots* fomentent des conflits politiques, faussent le discours en ligne et manipulent le marché »⁵⁰.

c. Les deepfakes

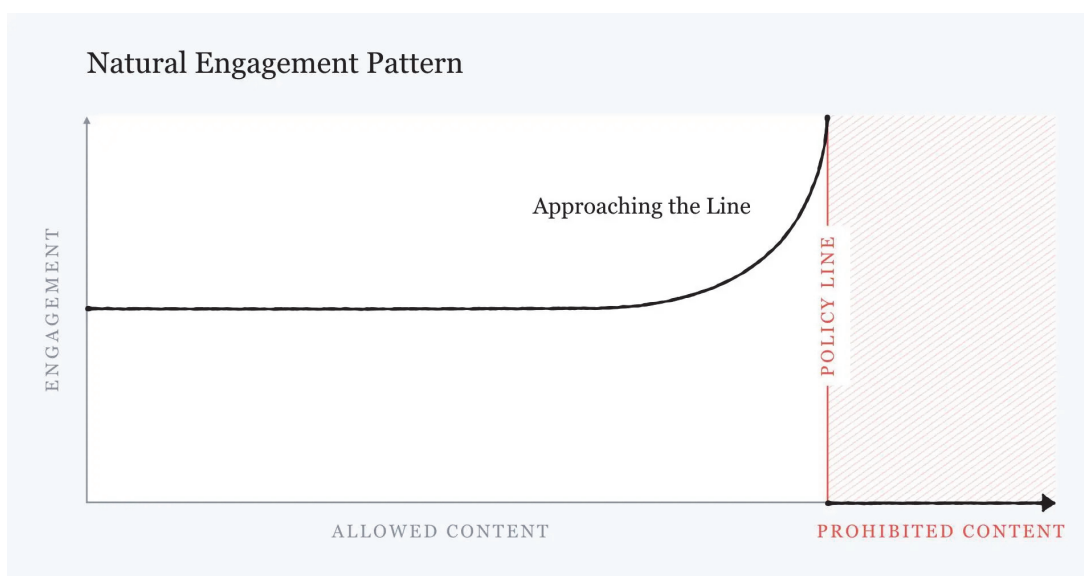
18. La création de faux contenus à des fins de manipulation de la réalité s'opérait assurément déjà longtemps avant l'avènement du numérique⁵¹, mais ce dernier a rendu la manipulation de textes particulièrement aisée, et les images peuvent être manipulées de manière convaincante depuis un certain temps via des logiciels tels que *Photoshop*. Plus récemment encore, des systèmes d'IA peuvent être programmés non seulement pour générer des textes de manière automatique, mais aussi pour créer des images, des enregistrements audios ou encore des vidéos de manière de plus en plus réaliste. La manipulation de textes et d'images était, comme nous le mentionnions, déjà aisée avec des technologies moins sophistiquées, mais les systèmes

d'IA rendent également la création d'enregistrements audios et de vidéos particulièrement accessibles⁵². Or, si ces *deepfakes*, comme il est convenu de les appeler, sont souvent générés à des fins de divertissement, leur présence dans les médias et sur les réseaux sociaux peut engendrer des effets singulièrement néfastes pour la société. Ainsi pourra-t-on créer artificiellement les images d'une manifestation, la rencontre de deux personnalités politiques, prêter à un orateur des propos qu'il n'a jamais tenus, etc.

3. Les conséquences : le web et le risque d'une société de plus en plus divisée

19. Nos développements le démontrent amplement : l'écosystème actuel du *web*, et en l'occurrence plus particulièrement celui présent sur les réseaux sociaux, joue un rôle prépondérant dans la polarisation de la société et encourage les internautes à créer des spectacles de violence. Si l'expression d'opinions divergentes a de tout temps été limitée par la crainte naturelle des humains de se trouver isolés du groupe ou même lynchés publiquement pour leurs propos, le *web* a évolué de manière à encourager ce qui est désormais qualifié de « cancel culture », y compris en enfermant les internautes dans des chambres d'échos les confortant et souvent les radicalisant dans leurs idées. Ainsi, l'opinion divergente est plus amèrement perçue comme non conforme et la violence est de mise, d'autant plus qu'un sentiment de frustration est constamment alimenté par la structure des plateformes qui encourage une effusion collective contre une victime semblant mériter d'être sujette à de virulentes critiques.

De cette évolution, les créateurs du *Web* s'inquiètent depuis une décennie déjà. Berners-Lee indiquait que le *web* était censé être un lieu social⁵³ : « Berners-Lee, debout devant un tableau noir, dessine un gra-



Ce graphique montre que plus une publication est susceptible de violer les normes communautaires de Facebook, plus elle reçoit d'engagement de la part des utilisateurs, parce que les algorithmes qui maximisent l'engagement récompensent les contenus incendiaires. Dans la note précitée, le fondateur de Facebook a ensuite partagé un autre graphique avec la relation inverse. La compagnie pourrait commencer à pénaliser le contenu limite en lui accordant moins de distribution et d'engagement. Cela se ferait avec plus d'IA : Facebook vise à développer de meilleurs modèles de modération de contenu pour détecter le contenu limite afin de pouvoir le repousser rétroactivement plus bas dans le fil d'actualité. Mais il est important de

noter que les algorithmes qui recommandent le contenu travaillent toujours pour maximiser l'engagement des utilisateurs. Cela signifie donc qu'en réalité, chaque publication toxique échappant aux systèmes de modération continuera à être promue. Le deuxième graphique est donc un peu trop simplifié et ne correspond pas à ce qui se passera réellement avec cette approche. Voy. K. HAOA, *How Facebook got addicted to spreading misinformation*, *MIT Technology Review*, 11 mars 2021, <https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/>.

(49) Abréviation pour *software robots*.
(50) M. LAMO et R. CALO, « Regulating Bot Speech », *UCLA*

Law Review, 16 juillet 2018, <http://dx.doi.org/10.2139/ssrn.3214572> (traduction libre) « We live in a world with artificial speakers with a real impact. So-called “bots” foment political strife, skew online discourse, and manipulate the market place ».

(51) Il a ainsi par exemple été révélé (en 1440 seulement) que la « donation de Constantin », acte par lequel l'empereur romain Constantin I^{er} aurait conféré au pape Silvestre la primauté sur les Églises d'Orient et l'*imperium* (pouvoir impérial) sur l'Occident au IV^e siècle, avait en réalité été rédigée au cours de la deuxième moitié du VIII^e siècle afin de justifier l'expansion des territoires pontificaux, et qu'elle fut de surcroît longtemps encore invoquée pour asseoir les pouvoirs du pape. Voy.

A. MORK (éd.), *Fake for Real - A history of forgery and falsification*, Temporary Exhibition Catalogue - Maison de l'histoire européenne, Luxembourg, Publications Office of the European Union, 2020, pp. 37-40.

(52) En effet, quiconque peut se procurer un faux enregistrement audio pour 10 dollars tous les 50 mots, ou une fausse vidéo de la personne de son choix pour seulement 3 dollars, pour autant qu'elle dispose de 250 photos de cette personne. A. M. WALORSKA, *Deepfakes and disinformation*, Friedrich Naumann Foundation for Freedom, Potsdam, mai 2020, https://fnf-europe.org/wp-content/uploads/2020/06/fnf_deepfakes_broschuere_en_web.pdf, p. 15.

(53) « L'objectif initial était de tra-

phique, comme il a tendance à le faire. Ce tableau classe les groupes sociaux par taille, selon les familles, les groupes de travail, les écoles, les villes, les entreprises, la nation, la planète. Le *Web* pourrait en théorie faire fonctionner les choses sans heurts à tous ces niveaux, ainsi qu'entre eux. C'était en effet l'idée initiale — une étendue organique de collaboration. Mais le *Web* peut se développer dans l'autre sens. Et Berners-Lee s'inquiète que cela "permette aux fous et aux cinglés de trouver dans le monde 20 ou 30 d'autres fous et cinglés qui sont absolument convaincus des mêmes choses. **Permettons-leur de mettre en place des filtres autour d'eux... et développons un nid-de-poule de culture dont ils ne peuvent pas sortir**". *Allons-nous "aboutir à un monde plein de cultures très, très disparates qui ne se parlent pas ?" »⁵⁴.*

Dans un ouvrage récent, V. de Coorebyter s'inquiète également de cette évolution : « On peut s'inquiéter des lynchages qui se développent de manière fulgurante à la moindre parole malheureuse, dissonante ou à contre-courant, éventuellement tirée de son contexte. Le risque est réel de voir triompher une morale *mainstream* qui sanctionne durement des points de vue non conformes, jugés scandaleux, dangereusement réactionnaires, ou simplement trop tièdes quant au soutien apporté à un combat présenté comme indiscutable. Internet favorise ce qu'on appelle les chambres d'écho, la tendance au renforcement des convictions par un mécanisme de répétition, d'approbation mutuelle et de critique sans appel du point de vue opposé. *Il se crée ainsi des entre-soi bloqués sur leur certitude d'avoir raison et sur leur droit à dénoncer la nocivité de l'adversaire, qui est moins un interlocuteur qu'un coupable* »⁵⁵.

Récemment, ce 21 avril, l'ancien président Obama, lors d'une conférence tenue à Stanford⁵⁶, accusait les grandes plateformes de réseaux sociaux d'avoir largement amplifié « les pires instincts de l'humanité », mettant en avant que « [l']une des causes majeures de l'affaiblissement des démocraties tient au profond changement dans nos façons de communiquer et de nous informer ».

2 La *cancel culture* à l'aune de la liberté d'expression et du numérique

A. Le principe de la liberté d'expression : du dogme à la relativité

20. La liberté d'expression et de la presse est un dogme de la Constitution américaine, consacré par le premier amendement de la Constitution en 1791. Elle s'exprime comme suit : « Le Congrès n'adoptera aucune loi relative à l'établissement d'une religion, ou à l'interdiction de son libre exercice ; ou pour limiter la liberté d'expression, de la

presse ou le droit des citoyens de se réunir pacifiquement ou d'adresser au Gouvernement des pétitions pour obtenir réparations des torts subis ». Cette liberté d'expression est revendiquée par les partisans de la *cancel culture* pour justifier des initiatives prises en particulier sur le *web* en la matière. Certes, elle connaît dans la Constitution une limitation. Une punition interviendrait dès lors que le discours inciterait directement à des actes de violence physique ou à des comportements illégaux mais le *discours haineux* (*hate speech*) est protégé par le Premier amendement. Une telle consécration s'explique par la conviction des auteurs de la Constitution que la concurrence des idées librement exprimées par chacun aboutira à la victoire de la vérité. Comme le note le juge Holmes en 1928 dans une décision importante de la Cour suprême des États-Unis (*US c. Schwimmer*)⁵⁷ : « It is the theory of our Constitution that the best of truth is the power of the thought itself accepted by the competition of the market ». (« C'est la théorie de notre Constitution que la meilleure des vérités est la puissance de la pensée elle-même acceptée par la concurrence du marché »⁵⁸).

La consécration de la liberté d'expression par l'article 10 de la Convention européenne des droits de l'homme (ConvEDH) protège de même la liberté d'expression tout en prévoyant des exceptions plus larges : « 1. Toute personne a droit à la liberté d'expression. Ce droit comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontière. Le présent article n'empêche pas les États de soumettre les entreprises de radiodiffusion, de cinéma ou de télévision à un régime d'autorisations. 2. L'exercice de ces libertés comportant des devoirs et des responsabilités peut être soumis à certaines formalités, conditions, restrictions ou sanctions prévues par la loi, qui constituent des mesures nécessaires, dans une société démocratique, à la sécurité nationale, à l'intégrité territoriale ou à la sûreté publique, à la défense de l'ordre et à la prévention du crime, à la protection de la santé ou de la morale, à la protection de la réputation ou des droits d'autrui, pour empêcher la divulgation d'informations confidentielles ou pour garantir l'autorité et l'impartialité du pouvoir judiciaire ». On connaît à cet égard le célèbre arrêt de 1976 de la Cour de Strasbourg : La liberté d'expression « vaut non seulement pour les "informations" ou "idées" accueillies avec faveur ou considérées comme inoffensives ou indifférentes, mais aussi pour celles qui heurtent, choquent ou inquiètent l'État ou une fraction quelconque de la population. Ainsi le veulent le pluralisme, la tolérance et l'esprit d'ouverture sans lesquels il n'est pas de "société démocratique" »⁵⁹.

Le message dit de *cancel culture* est donc *a priori* protégé par la liberté d'expression vis-à-vis d'une censure tant publique que privée y compris par les plateformes de communication et d'information. Récemment, le Conseil des ministres de l'Union européenne le rappelle : « With regard to the importance of freedom of speech, states and administrative regulatory authorities as well as private platform providers

vailler ensemble avec d'autres", dit-il. "Le *Web* était censé être un outil créatif, un outil d'expression". Il avait imaginé, par exemple, un travailleur affichant une note sur un site *web* accessible uniquement à des collègues et les faisant réagir en incorporant des hyperliens menant à leurs commentaires ou à d'autres documents pertinents ; ou une famille bi-côtée planifiant de la même manière sa réunion annuelle sur le site familial. Mais le *Web* s'est développé autrement. Robert Cailliau du CERN, le premier collaborateur de Berners-Lee sur le projet, décrit la structure descendante dominante du *Web* : "Il y a un point qui met les données à l'extérieur, et vous n'êtes qu'un consommateur". Il trouve ce modèle un "désastre absolu et total" ». R. WRIGHT, « The man who invented the web », *Time*, 24 juin 2001, <http://content.time.com/time/magazine/article/0,9171,137689,00.html> (traduction libre) « [The web] was meant to be a social place. "The original goal was working together with

others", he says. "The Web was supposed to be a creative tool, an expressive tool". He had imagined, say, a worker posting a memo on a website accessible only to colleagues and having them react by embedding hyperlinks that led to their comments or to other relevant documents ; or a bi-coastal family similarly planning its annual reunion on the family site. But the Web turned out otherwise. Robert Cailliau of CERN, Berners-Lee's earliest collaborator on the project, describes the Web's prevailing top-down structure : "There's one point that puts the data out, and you're just a consumer". He finds this model — whose zenith is the coming wave of so-called push technology — an "absolute, utter disaster" ». Sur la genèse de la création du *Web*, lire l'ouvrage de Q. JARDON, *Alexandria - les pionniers oubliés du Web*, Gallimard, Paris, 2019.

(54) Nous soulignons ; R. WRIGHT, « The man who invented the web », *op. cit.* (note 53), (traduction libre) « Berners-Lee, standing at a

blackboard, draws a graph, as he's prone to do. It arrays social groups by size. Families, workplace groups, schools, towns, companies, the nation, the planet. The Web could in theory make things work smoothly at all of these levels, as well as between them. That, indeed, was the original idea—an organic expanse of collaboration. But the Web can pull the other way. And Berners-Lee worries about whether it will "allow cranks and nut cases to find in the world 20 or 30 other cranks and nut cases who are absolutely convinced of the same things. Allow them to set up filters around themselves ... and develop a pothole of culture out of which they can't climb". Will we "end up with a world which is full of very, very disparate cultures which don't talk to each other ?" ».

(55) Nous soulignons ; V. DE COOREBYTER, « L'internet : démocratie ou démagogie », *op. cit.* (note 39), p. 248.

(56) Sur cette conférence, voir l'édition du *Monde* du 22 avril 2022 : [https://www.lemonde.fr/pixels/article/2022/04/22/barack-obama-appelle-a-reguler-les-reseaux-sociaux-responsables-de-l-affaiblissement-des-democraties_6123194_4408996.html?xtor=EPR-32280629-\[a-la-une\]-20220422-](https://www.lemonde.fr/pixels/article/2022/04/22/barack-obama-appelle-a-reguler-les-reseaux-sociaux-responsables-de-l-affaiblissement-des-democraties_6123194_4408996.html?xtor=EPR-32280629-[a-la-une]-20220422-). L'ex-président met en cause les outils technologiques mis en place par ces plateformes et leur modèle économique fondé sur la publicité à large échelle. Ces algorithmes devraient être soumis à des contrôles de sécurité par une autorité de régulation, au même titre que les voitures, aliments et autres produits de consommation. Selon sa formule, « les outils ne nous contrôlent pas. Nous pouvons les contrôler ».

(57) Judge Holmes, dissenting opinion in *Abrams vs US*, 250 US 616 (1919).

(58) Traduction libre.

(59) C.E.D.H., 7 décembre 1976, arrêt *Handyside c. Royaume-Uni*, req. n° 5493/72, § 49.

should abstain from defining quality content or the reliability of content itself »^{60 61}.

B. Le recours aux juridictions étatiques et comment échapper aux conséquences de la qualification de délit de presse ?

21. De cette liberté d'expression, découle la liberté de la presse, traditionnellement plus protégée encore. Le message électronique public échangé sur les réseaux sociaux, message contribuant à un appel à *cancel culture*, est protégé par la liberté de la presse vis-à-vis d'une censure tant privée que publique. La liberté de la presse, selon une jurisprudence récente tant de la Cour de Strasbourg que de Luxembourg s'applique en effet aux « écrits » imprimés ou numériques (médiat sociaux, *blogs*), accessibles au public, même émanant de simples particuliers, ces derniers étant qualifiés de journalistes même s'ils ne sont ni des professionnels ni soumis à la déontologie journalistique : « Si l'ouverture d'espaces de débat public fait partie du rôle de la presse, l'exercice de cette mission n'est pas réservé aux médias ou aux journalistes professionnels »⁶². On note cette extension qui aurait pour conséquence, conformément à l'article 150 de la Constitution belge⁶³, de soumettre aux seuls jurys d'assises les infractions telles que l'injure, la calomnie ou le harcèlement, dès lors que ces infractions expriment une pensée ou une opinion dans un écrit rendu public y compris sur les réseaux sociaux (alors qualifiés de délits de presse). La conséquence de cette protection par la liberté de la presse des messages publiés par des particuliers sur les réseaux sociaux reviendrait ou aboutirait, vu la lourdeur des procès d'assises, selon nombre d'auteurs, à une quasi-impunité des auteurs de certaines expressions délictueuses.

22. La facilité avec laquelle les réseaux sociaux, par l'anonymat des messages, par l'étendue de la diffusion et par le manque de contextualisation et souvent d'argumentation des propos y tenus, donnent à la désinformation et aux messages de *cancel culture* une ampleur que ne permettaient pas les « écrits » ou les modes de diffusion traditionnels, amène le législateur à s'interroger sur la nécessité d'étendre les hypothèses de correctionnalisation des délits de presse. Dans le cadre des limites fixées par l'article 10 de la CEDH à la liberté d'expression et de la presse, et sur base de l'exemple de la loi Moureaux⁶⁴, le législateur belge entend ainsi multiplier les hypothèses de correctionnalisation d'infractions constitutives de délits de presse afin d'éviter les lourdeurs des procès d'assises. Ainsi, seraient jugés comme simples délits sans passer devant le jury d'assises :

- 1) l'incitation à la discrimination, à la haine, à la violence ou à la ségrégation à l'égard d'autrui, en public, intentionnellement et pour une raison précise ;
- 2) la diffusion d'idées fondées sur la supériorité raciale ou la haine raciale ;
- 3) l'appartenance ou la collaboration à un groupement ou à une association qui, de manière répétée, prône la discrimination ou la ségrégation ;
- 4) le négationnisme ;
- 5) les injures écrites, l'abus de moyens de communication et le harcèlement.

On relève en particulier une proposition du parti socialiste lors de la dernière révision constitutionnelle (proposition de loi du 12 novembre 2019 modifiant l'article 442bis du Code pénal, DOC 55 0738/001) de déclarer comme délit le « harcèlement groupé », défini comme des « faits imposés à une même victime par plusieurs personnes, de manière concertée ou à l'instigation de l'une d'elles, alors même que chacune des personnes n'a pas agi de façon répétée » ou des « faits sont imposés à une même victime, successivement, par plusieurs personnes qui, même en l'absence de concertation, savent ou auraient dû savoir que ces propos ou comportements caractérisent une répétition ». Cette incrimination s'applique particulièrement bien à certaines formes de *cancel culture*, fondées sur la multiplication concertée de messages sur le net. La proposition propose également de déclarer comme délit le *doxing*, à savoir la révélation publique de données à caractère personnel⁶⁵. Toujours à l'occasion des débats sur la révision de la Constitution, les partis Ecolo/Groen et CD&V ont souhaité voir étendue la correctionnalisation des messages racistes et xénophobes instituée par la loi Moureaux aux messages incitant⁶⁶ à la haine, à la violence et à la « discrimination », qui, jusqu'à présent, sont l'objet de procès civils mais peuvent faire l'objet d'une ordonnance de suspension par le président du tribunal de première instance. Il est indéniable que certains messages de *cancel culture* sont visés par ces propositions de loi. Leur consécration législative soit cette extension par création de nouveaux délits se justifierait par les risques nouveaux créés par le phénomène des réseaux sociaux.

C. Plaidoyer pour une approche restrictive de la notion de désinformation et conséquences sur les mesures à envisager pour la *cancel culture*

23. Faut-il suivre ce mouvement de correctionnalisation qui permettrait aux tribunaux répressifs d'être plus facilement saisis ? Notre propos sur ce point est double. Le recours juridictionnel même facilité restera « une denrée rare ». En effet, face à des messages de *cancel culture*, il importe d'agir vite et de ne point devoir attendre l'issue d'un procès qui reste aléatoire dans son contenu mais également dans sa portée souvent limitée au seul territoire national voire européen. Ce sera par conséquent souvent auprès de l'opérateur du réseau social, du moteur de recherche ou autre plateforme de partage de contenus que la personne s'adressera et demandera le blocage, la suppression, ou le retrait de priorité du message en cause. Par ailleurs, et c'est notre second point dans une matière qui touche à une liberté fondamentale, nous soulignons le risque de *Chilling effect*⁶⁷ d'une répression pénale qui demain pourrait être plus facile et étendue. Comme le note J. Englebert, « La haine véhiculée sur les réseaux sociaux est avant tout un révélateur culturel et économique avec d'être un délit. Et la réponse pénale nécessairement individuelle qui lui est donnée est d'une rare médiocrité sociale »⁶⁸. Sans doute, doit-on préférer comme l'affirme le Parlement européen dans une résolution du 20 novembre 2020 « sur le renforcement de la liberté des médias, la protection des journalistes en Europe, les discours de haine, la désinformation et le rôle des plateformes », diverses mesures comme l'éducation aux médias et l'appui y compris financier à une presse pluraliste, indépendante et forte de journalistes respectant les codes de déontologie qui assurent la qualité de l'information diffusée⁶⁹.

(60) Council of the European Union, « Council conclusions on safeguarding a free and pluralistic media system », 27 novembre 2020, n° 37, <https://data.consilium.europa.eu/doc/document/ST-13260-2020-INIT/en/pdf>.

(61) « En ce qui concerne l'importance de la liberté d'expression, les États et les autorités de régulation administrative ainsi que les fournisseurs de plateformes privées devraient s'abstenir de définir un contenu de qualité ou la fiabilité du contenu lui-même » (traduction libre).

(62) C.E.D.H., 14 avril 2009.

(63) À cet égard, voy. l'attendu de l'arrêt de la Cour de cassation du 7 octobre 2020, RG n° P.19.0644.F : L'article 150 de la Constitution « ne fait dépendre la compétence du jury,

ni de la pertinence, ni de l'importance sociale de la pensée ou de l'opinion exprimée, ni du caractère plus ou moins argumenté ou développé de l'écrit incriminé, ni de la notoriété de son auteur ». En l'occurrence, il s'agissait de propos publiés sur un « mur Facebook » dénigrant la conduite d'une échevine locale lors de la passation d'un marché public. Voy. sur ce sujet la note d'observations de J. ENGLEBERT, *R.D.T.I.*, n° 81, Bruxelles, Larcier, 2021, pp. 106-117. La note reprend toute l'évolution de la jurisprudence belge tendant à élargir de plus en plus la notion d'expression journalistique.

(64) La loi du 30 juillet 1981 tendant à réprimer certains actes inspirés par le racisme ou la xénophobie, dite loi Moureaux, est une loi belge correc-

tionnalisant les discours de haine mais aussi contre les discriminations raciales et xénophobes et le harcèlement. On ajoute que des dispositions permettent au président du tribunal de première instance d'ordonner la cessation immédiate de toute autre forme de discrimination non raciale, par exemple basée sur le sexe, l'orientation sexuelle, l'état civil, la naissance, l'âge, l'état de santé actuel ou futur, un handicap ou une caractéristique physique, la conviction religieuse ou philosophique.

(65) Le ministre de la Justice des Pays-Bas a introduit en juillet 2021, un projet de loi contre le *Doxing* (<https://rijksoverheid.nl/actueel/nieuws/21/07/12/strafrechtelijke-aanpak-intimidatie-door-delen-persoonsgegevens>).

(66) À noter que c'est l'incitation (l'appel au passage à l'acte) qui est repris comme critère et non le contenu ou le motif du message.

(67) À traduire par « Effet glaçant ».

(68) J. ENGLEBERT, « La liberté d'expression à l'heure de l'Internet », in Y. POULLET (éd.), *Vie privée, liberté d'expression et démocratie dans la société numérique*, coll. Cahier du CRIDS, n° 47, Bruxelles, Larcier, 2020, p. 139.

(69) Voilà quelques lignes d'actions préconisées par cette résolution parlementaire qui défend une approche positive en faveur de la liberté de la presse : affirmation du rôle des organes de presse comme service public essentiel ; mise en évidence des dangers d'une presse de plus en plus contrôlée par certains pouvoirs privés

24. Indéniablement, certains messages aujourd'hui sont ou demain pourront être incriminés au niveau pénal⁷⁰ et justifieront dès lors une intervention des juridictions pénales voire civiles, dans le cadre de procédures plus ou moins rapides, voire une intervention d'autorités privées en l'occurrence les opérateurs de site *web* de communication et d'information, en particulier les plateformes. Pour autant, là ne s'arrête pas la volonté des États de réguler les messages sur Internet. Les notions de « désinformation » et de « cancel culture » visent des messages qui ne ressortent pas nécessairement aux champs d'application des incriminations pénales voire civiles déjà existantes ou envisagées. Des sanctions civiles ou privées voire des mesures technologiques peuvent sanctionner des messages au-delà de ceux ressortant aux incriminations pénales.

L'article 10 de la Conv.EDH exige cependant en toute hypothèse que les mesures tant pénales que civiles voire technologiques prises répondent à des exigences strictes de légalité et de proportionnalité. Ces exigences sont-elles rencontrées par les notions de désinformation et de *cancel culture* ? La proposition de loi espagnole⁷¹ qui entendait punir « l'atteinte à la stabilité institutionnelle du pays », a été rejetée pour atteinte à la liberté d'expression par le Parlement espagnol le 21 mars 2018. Par contre, semblent devoir être admises des mesures qui préviennent pendant la période électorale la dissémination d'informations fausses ou trompeuses visant à fausser le résultat des élections⁷². Sur cette base, il nous apparaît que les mesures combattant la désinformation, au nom d'une démocratie dite militante, nécessiteraient une approche plus prudente. Rappelons la définition du concept de « désinformation », déjà reprise ci-dessus (voy. n° 4), proposée dans l'étude commanditée par la Commission LIBE du Parlement européen et publiée en 2019⁷³, qui retient quatre éléments : premièrement, l'utilisation d'un contenu soit faux, soit prêtant à confusion (*misleading*), soit utilisant des techniques non éthiques de persuasion ; deuxièmement, l'intention de tromper par de tels messages à des fins de profit ou d'intérêt politique ; troisièmement, l'impact sur des sujets d'intérêt général (élections, environnement, santé publique...) troublant le fonctionnement de la société ; et, quatrième, le recours stratégique à la technologie (y compris aux systèmes d'intelligence artificielle) comme mode de dissémination des messages. Au nom de la légalité, de la nécessaire proportionnalité et de la légitimité dans une société démocratique, cette définition ne devrait-elle pas spécifier les sujets de troubles au fonctionnement de la société et réserver aux seuls cas graves, jugés comme tels *in fine* par des commissions indépendantes, qu'elles soient privées (voy. nos réflexions *infra*, n° 32) ou publiques, l'intervention des opérateurs ou de l'État ? À défaut, la notion de désinformation pourrait servir de prétexte à des restrictions illégitimes de la liberté d'expression. Le concept « troubler le fonctionnement de la société », qui nous apparaît trop vague, s'appliquerait, en ce qui concerne le contenu, à des informations en matière de santé publique (« La grippe cette année a provoqué tant de morts et le vaccin retenu pour le combattre est dangereux ! »), à des informations en matière d'environnement (« Trois morts dus à la chaleur à Bruxelles ! »), à un complot des juges contre les libertés individuelles, à un coup d'État en préparation dans tel ou tel parti politique, etc. C'est dans ce même esprit que nous pensons que la *cancel culture* ne devrait faire l'objet de

mesures restrictives que si le message répond aux mêmes conditions que celles de la désinformation et constitue alors une forme particulière de désinformation, ou si le message contrevient à une infraction par ailleurs spécifiée par la loi (diffamation, messages terroristes, discours racistes, pédopornographique, appel à la violence, etc.). Nous estimons également que l'information fautive ou trompeuse, sauf si elle est visée par un cas d'incrimination prévue par la loi, doit recevoir la protection due à la liberté d'expression. Ce n'est que dans la mesure où elle contamine l'ensemble ou du moins une partie importante de l'espace public informationnel ou le domine, qu'elle pourrait apparaître comme contraire à la finalité même de la liberté d'expression, à savoir permettre le débat démocratique. Cette réflexion exclut des messages au contenu certes répréhensible et envoyés avec l'intention de troubler le fonctionnement de la société sur un sujet d'intérêt général (par exemple, affirmer que l'armée prépare un coup d'État militaire) mais dont la portée, vu le média choisi, reste limitée. C'est en ce sens que nous apparaît ne pas devoir faire l'objet de mesures des messages constitutifs de désinformation, véhiculés par des médias dont l'audience est limitée, et que ces mesures devraient être réservées aux très larges plateformes (*infra*, n° 29).

D. Le droit de la protection des données face aux opérations de traitement des données nécessaires à l'élaboration de messages de *cancel culture* ou de désinformation

25. On ajoute que désormais, l'application du Règlement général sur la protection des données (RGPD) et de la transposition de celui-ci par la loi belge du 30 juillet 2018⁷⁴ (en particulier ses articles 24 et s.), fixe des limites aux traitements de données et n'exempte de certains articles du RGPD que les seuls traitements aux fins « d'informer le public, à l'aide de tout média et où le responsable du traitement s'impose des règles de déontologie journalistique »⁷⁵. Même vis-à-vis de ces derniers, il soumet le traitement au respect des principes de proportionnalité et de licéité et confère aux personnes concernées un droit au déréferement, à l'opposition et à l'oubli. Ainsi, une personne physique (non une entreprise), qui, dans le cadre d'une opération de *cancel culture*, se verrait « boycottée », pourrait invoquer de tels droits pour faire prévaloir son intérêt face à la diffusion de messages appelant à son boycott. Ce sera au juge d'opérer la balance entre cet intérêt légitime et celui de l'intérêt général, en l'occurrence le droit de savoir du public⁷⁶.

L'application du RGPD à la matière de la désinformation ou de la *cancel culture* en cas de traitement par l'auteur ou l'initiateur de tels messages portant sur des personnes identifiées ou identifiables⁷⁷ — et l'hypothèse sera fréquente — soulève quelques questions supplémentaires qu'une décision récente de l'autorité belge de protection des données (APD)⁷⁸ met en exergue. En l'occurrence, il s'agissait non d'attaquer des auteurs de désinformation mais bien une a.s.b.l. qui s'était fait fort de lutter contre la désinformation. Néanmoins, les principes mis en exergue par l'APD doivent valoir également pour ceux qui se prévaudraient de la liberté d'expression ou de la presse pour justifier leur traitement de données à caractère personnel dans le cadre d'opérations de désinformation ou de *cancel culture*.

et, tantôt, réduite au silence par la menace externe voire par les gouvernements ; constatation d'une paupérisation d'une presse privée due à la chute des revenus publicitaires (concurrence des plateformes) et de sa soumission croissante aux impératifs d'une opinion qui souhaite sensationnalisme et réaction immédiate ; importance pour la démocratie à la fois d'un journalisme de haute qualité déontologique et donc critique par rapport à ses sources et du pluralisme d'opinions des médias ; nécessité absolue d'une éducation aux médias.

(70) En ce sens, la disposition du Code pénal français (article 322-14) : « Le fait de communiquer ou de divulguer une fausse information dans le but de faire croire qu'une destruction, une dégradation ou une détérioration dangereuse pour les personnes

va être ou a été commise est puni de deux ans d'emprisonnement et de 30.000 EUR d'amende. Est puni des mêmes peines le fait de communiquer ou de divulguer une fausse information faisant croire à un sinistre et de nature à provoquer l'intervention inutile des secours ».

(71) http://www.gppopular.es/wp-content/uploads/2017/12/171219-PNL-Noticias-falsas.pdf?_ga=2.226110172.738521628.1541868032-2076439625.1541868032.

(72) Voy. à ce sujet les réflexions reprises dans le rapport présenté à la Commission LIBE du Parlement européen : *Disinformation and propaganda - impact on the functioning of the rule of law in the EU and its Member States*, rapport déjà cite, pp. 91 et s. et en particulier, le tableau (pp. 47 et s.) des cas d'atteintes

majeures relevées au processus électoraux par la désinformation.

(73) Commission LIBE du Parlement européen, *Disinformation and propaganda - impact on the functioning of the rule of law in the EU and its Member States*, op. cit. (note 20), pp. 14 et s.

(74) Loi belge du 30 juillet 2018 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel.

(75) Article 24, § 1, de la loi du 30 juillet 2018.

(76) Sur ce point, lire T. LÉONARD et Y. POULLET, « L'intérêt général comme arbitre du débat vie privée vs liberté d'expression dans le RGPD », in Y. POULLET (éd.), *Vie privée, liberté d'expression et démocratie dans la société numérique*, coll. Cahiers du CRIDS, n° 47, Bruxelles, Larcier, 2020, pp. 73 et s.

(77) On parle alors de *dox(x)ing* défini comme suit : « Doxing or doxxing is the act of publicly revealing previously private personal information about an individual or organization, usually via the Internet. Methods employed to acquire such information include searching publicly available databases and social media websites (like Facebook), hacking, social engineering. Doxing may be carried out for reasons such as online shaming, extortion, and vigilante aid to law enforcement. It also may be associated with hacktivism » (Wikipedia, v° « Doxing »).

(78) Autorité de protection des données (chambre contentieuse) Décision quant au fond n° 13/2022 du 27 janvier 2022, accessible sur le site de l'APD (<https://www.autoriteprotectiondonnees.be/citoyen/publications/decisions>).

26. En l'espèce, une ASBL, à l'occasion de faits mettant en cause la garde du corps du président de la République française, s'était penchée sur une analyse des tweets publiés à cette occasion pour chercher à découvrir si oui ou non les réactions multiples émises à ce propos constituaient une manœuvre de désinformation à la solde de la Russie. À la suite de la publication de l'étude et au vu des contestations nombreuses soulevées par cette dernière, l'auteur de l'étude avait alors procédé, sans doute maladroitement, à la publication des tweets et des éléments biographiques rassemblés à propos de leurs émetteurs. De nombreuses personnes avaient alors porté plainte pour violation de leur vie privée. Le responsable des traitements affirmait que les données récoltées à l'origine de son travail étaient toutes des données « publiques » (tweets et moteurs de recherche) et que dès lors le RGPD ne s'appliquait pas. La chambre contentieuse de l'APD rappelle à ce propos que « [l]e caractère public des données personnelles disponibles sur les réseaux sociaux et confirmé par les CGV de Twitter ne signifie pas que celles-ci perdent la protection conférée par le RGPD, y compris celle liée au caractère sensible de telles données, soit par nature soit dans le cadre de la finalité du traitement ». À ce premier argument, le responsable en ajoutait un deuxième, se prévalant devant l'APD de sa liberté d'expression et de recherche pour justifier tant la collecte des informations qui avaient servi à l'analyse que leur publication. À cet égard, l'APD rappelle le principe de l'extension de liberté de la presse aux auteurs de *blogs* et de messages publics sur la toile : « les défenderesses ne doivent pas démontrer d'appartenance à une profession de journaliste ou la soumission à des règles déontologiques particulières pour pouvoir invoquer l'exception de traitement journalistique dans le cadre de leur liberté d'expression. Exiger la démonstration d'une appartenance professionnelle au journalisme (ex. souscription d'un code de déontologie) ne permettrait pas de donner effet utile au droit à la liberté d'expression prévu à l'article 11 de la Charte dans le contexte où la législation belge applicable au moment des faits ne prévoyait pas encore la possibilité explicite d'invoquer une exception couvrant la liberté d'expression en dehors d'un contexte de journalisme, contrairement à l'article 85 du RGPD ». L'APD ajoute même que cette finalité journalistique l'emporte même si elle est combinée avec d'autres finalités, par exemple commerciales et qu'« une démarche journalistique peut être fondée sur l'intérêt légitime du responsable de traitement dans le cadre de son droit fondamental à la liberté d'expression, ou encore, son intérêt légitime à exercer un droit de réponse numérique ». Ceci dit, l'autorité souligne la nécessité de respecter les principes de base du RGPD, tant celui de minimisation que celui de sécurité : « la finalité poursuivie pouvait être réalisée d'une manière moins attentatoire aux intérêts, droits et libertés des personnes concernées, en publiant — et traitant — des données dûment pseudonymisées (notamment en ce qui concerne le fichier « Rumeurs & Items » diffusé par l'ASBL et en prévoyant des restrictions d'accès) ». Enfin, « si l'APD a considéré que l'ONG était dispensée de son obligation d'information individuelle des personnes au sujet des données personnelles traitées pour la réalisation de l'étude (voy. exception journalistique), au motif que cela aurait pu compromettre cette étude et sa publication ultérieure, elle a en revanche considéré que la publication de données sensibles utilisées pour l'étude, non dûment pseudonymisées, était dépourvue de base légale en raison de l'atteinte disproportionnée portée aux droits des auteurs de tweets concernés. Leur consentement était également requis pour la publication de telles données sensibles non pseudonymisées »⁷⁹.

L'argument de la liberté de recherche que les auteurs de messages de *cancel culture* invoqueront lorsque le contenu des messages émis a nécessité de patientes recherches sur le *web*, reçoit à la lumière de la décision ci-dessus résumée la réponse suivante. Tout d'abord, même si les données collectées à la base de la recherche ont été rendues publiques, leur utilisation à des fins de recherche n'est pas libre dans la mesure où ces données restent des données à caractère personnel et

doivent être traitées pour une finalité légitime, soit compatible avec la finalité pour laquelle elles ont initialement été collectées en l'occurrence par *Twitter* et rendues publiques, soit pour une des finalités légitimes prévues par l'article 6 du RGPD voire s'il s'agit de données sensibles à l'article 6 combiné avec l'article 9. Sur la compatibilité du traitement avec la finalité initiale, en complément des principes spécifiques aux traitements de données personnelles à des fins scientifiques prévus à l'article 89 du RGPD, la chambre contentieuse se réfère à l'article 6.4 du RGPD pour estimer que la finalité de recherche dans le chef de l'ASBL est incompatible avec la finalité de collecte initiale : « il y a lieu de tenir compte des attentes des personnes concernées (article 6.4.a et b) dans le cas d'espèce : cela implique que lorsque des personnes ont diffusé leurs données auprès d'un premier responsable de traitement (ex. *Twitter*) et ne s'attendent pas raisonnablement à faire l'objet d'un traitement ultérieur particulier par un autre responsable de traitement (par exemple, un profilage politique, voir l'article 4.4 du RGPD), une telle réutilisation des données pour de nouvelles finalités n'est possible qu'avec leur consentement sauf à disposer d'une autre base légale ». Le responsable du traitement doit donc baser son traitement sur une base de licéité distincte de la base de licéité initiale. Sans doute, l'APD reconnaît que constitue un intérêt légitime la recherche de la « vérité » et la contribution au débat public mais qu'en toute hypothèse il importe alors non seulement de vérifier si les intérêts des personnes concernées mises en cause ne prévalent pas et en tout cas de veiller au respect de « garanties appropriées », dont l'absence est patente en l'occurrence⁸⁰.

E. Le droit face aux technologies de modération des contenus et aux acteurs de la diffusion massive (les plateformes)

1. Les technologies de modération des contenus

27. Quoi qu'il en soit, il importe de montrer comment la technologie peut intervenir dans la modération des contenus, les possibles initiatives des plateformes en la matière et le regard du droit vis-à-vis de ce qui peut apparaître comme une censure privée. Les solutions technologiques, proposées en particulier par les plateformes mais également à disposition des autorités publiques qui peuvent y contraindre les plateformes à la suite d'une décision au fond ou en référé, suivent le principe bien connu : « Technology is the problem, it might be also the solution » (« Si la technologie est le problème, elle peut être aussi la solution »). La modération en ligne peut aboutir à filtrer, supprimer, bloquer, déprioriser voire supprimer un compte (tel le blocage du compte de l'ancien président Trump par *Twitter* et *Facebook*). Ainsi, pas moins de 2,5 millions de messages, photos et autres contenus ont été bloqués par Facebook pendant la période du 1^{er} janvier au 31 mars 2018 pour contrariété aux règles de Facebook en matière de messages haineux⁸¹.

La détection d'articles « erronés » ou illégaux peut emprunter diverses techniques. Soit suivant le modèle *end-to-end*, le système mis en place utilise des données « labellisées » dangereuses ou porteuses de désinformation (ainsi, on épinglera les messages utilisant les mots « pas de pitié pour... »). Soit, et sans doute aura-t-on ici recours à des techniques d'intelligence artificielle, on détectera des faits erronés repris dans des messages tantôt par la comparaison du contenu de l'article avec des preuves externes, tantôt en vérifiant si l'allégation est appuyée par d'autres sources déterminées ou en utilisant des informations non structurées provenant du *Web*. Des systèmes plus sophistiqués permettent de détecter un style mensonger ou, par l'analyse de métadonnées plutôt que du contenu même de l'article (par exemple les attributs de l'éditeur...), ce qui permet de repérer les *bots* sociaux. On souligne à cet égard que 99.3 % des faux comptes détectés par *Facebook* durant le troisième trimestre de 2020 furent repérés par les

(79) <https://www.autoriteprotection-donnees.be/citoyen/sanction-pour-traitement-massif-de-donnees-twitter-liees-a-laffaire-benalla-a-des-fins-de-profilage-politique>.

(80) Dans l'affaire précitée, l'APD estime qu'il aurait fallu tenir un registre des traitements non seulement parce que les données traitées sont sensibles par nature mais aussi et surtout de par la finalité politique poursuivie.

Elle reproche de surcroît aux auteurs de l'étude de ne pas avoir procédé à une analyse d'impact relative à la protection des données (*Privacy Impact Assessment*) vu les risques élevés (nombre de personnes visées et caractère politique des données) d'atteinte à la protection des données. Elle rappelle que l'obligation d'information des personnes concernées est certes limitée en cas de traitements

opérés à des fins de recherche doit cependant être respectée dans toute la mesure du possible et par des moyens adéquats. Enfin, elle invoque la nécessaire transparence des méthodes utilisées pour la recherche et l'exploitation des résultats : « la convocation de telles sources, si elle a lieu dans le cadre d'une démarche scientifique, requiert, selon la chambre contentieuse, plus de trans-

parence, notamment sur la méthodologie de collecte des données et leur degré d'anonymisation/de pseudonymisation ».

(81) Facebook, *Community Standards Enforcement Report*, disponible en ligne sur <https://transparency.facebook.com/community-standards-enforcement#hate-speech>.

systèmes d'intelligence artificielle mis en place par l'entreprise avant même que les utilisateurs ne les signalent⁸².

28. Toutefois, on pointera les défauts intrinsèques de telles techniques de détection des messages considérés comme de la désinformation en général, et qui pourraient s'appliquer aux messages de *cancel culture* répondant aux conditions de la désinformation (voy. *supra*). Les programmes d'analyse textuelle ne repèrent pas nécessairement les messages illicites, diffamatoires ou haineux (faux négatifs) ou, plus grave, repèrent des messages faussement qualifiés comme tels (faux positifs). Ces « erreurs » peuvent être attribuées à des biais volontaires ou non (par exemple, dans le contexte de la propagande russe dans le cadre de la guerre ukrainienne, toute information contenant le mot « guerre » est systématiquement bloquée) ou viennent de l'indéniable difficulté d'analyse de textes aux significations multiples suivant leur contexte, complexes et éventuellement contradictoires.

De telles mesures peuvent, si elles ne font pas l'objet de balises, remettre en cause la liberté d'expression puisqu'eu égard à l'importance de cette liberté, « les États et les autorités de régulation administratives ainsi que les fournisseurs de plateformes privées devraient s'abstenir de définir un contenu de qualité ou la fiabilité du contenu lui-même »⁸³. La mise en place de mesures de détection et de blocage par les plateformes privées soulève en particulier la crainte d'une censure privée qui pourrait suivre des règles différentes de celles mises en place par nos États démocratiques. Sans doute, les plateformes feront valoir que leur censure trouve dans la « loi contractuelle » sa légitimité : les utilisateurs d'*Instagram*, de *Facebook* (ou *Meta*), de *Twitter*, etc. n'ont-ils pas signé la charte d'utilisation des services offerts par ces plateformes ? À cette objection, nous opposons deux arguments : le premier est que la liberté d'expression est opposable tant vis-à-vis des autorités publiques que des autorités privées et ne peut souffrir donc que les limitations en accord avec le droit. Le second considère que les opérateurs de très larges plateformes, par leur position sur le marché, peuvent être considérés comme des « gatekeepers » de l'espace public de discussions et d'informations que constitue la toile. Les services ainsi offerts par ces plateformes devraient être considérés comme des services universels sur le modèle des services de communication. Comme l'accès à l'infrastructure de communication désormais libéralisé a été réglementé afin d'offrir, à chacun, un service de qualité défini par la loi, de même ces plateformes devraient selon notre opinion être réglementées en ce qui concerne tant les conditions d'accès que la qualité des services offerts (respect de la vie privée, absence de manipulations et respect de la liberté d'expression).

2. La régulation des plateformes

29. L'Union européenne évolue vers un système réglementaire de contrôle et de régulation des opérateurs de plateformes, en particulier des très larges plateformes, c'est-à-dire celles qui offrent un service à plus de 10 % de la population européenne. Sans entrer dans le détail, citons le *Digital Markets Act*, qui interdit certaines pratiques de *ranking* et de manipulation de l'offre sur les plateformes, et le *Digital Services Act*, adopté tout récemment ce 5 juillet 2022. Créer un espace informationnel de confiance, objectif du deuxième règlement, s'obtient par diverses mesures valables cumulativement suivant la catégorie d'intermédiaires dont l'entreprise en question relève. Ainsi, tous les fournisseurs de services intermédiaires doivent respecter les règles déduites des droits fondamentaux et coopérer avec les autorités (rapports d'activité et mise en œuvre des décisions des autorités judiciaires ou administratives).

Les plateformes en ligne sont tenues par des obligations supplémentaires, y compris détecter et s'opposer à la présence en ligne de produits, de services ou de contenus illégaux ou de désinformation, aidées en cela par l'action de « signaleurs de confiance » (*trusted flaggers*), accessibles aux internautes et avec lesquels les plateformes sont priées de coopérer. On note que si le principe est louable, aucune définition n'est donnée de la notion de désinformation et on peut craindre dès lors que les plateformes soient saisies de demandes bien

au-delà des limites strictes de la notion pourtant prônée par les travaux du *High Level Group of Experts on disinformation* dont le contenu a précédé et inspiré la proposition de règlement. En outre, les places de marché en ligne (*online marketplaces*) sont tenues de tracer les utilisateurs professionnels présents sur leurs sites afin d'identifier les vendeurs de biens ou services illégaux ; doit être prévue la possibilité pour les utilisateurs de contester les décisions de modération de contenus prises par les plateformes ; sont imposées des mesures de transparence, en particulier en ce qui concerne les systèmes de recommandation ainsi que toute option permettant aux bénéficiaires du service de modifier ou d'influencer les principaux paramètres de recommandation. On ajoute à toutes ces obligations mais, cette fois, pour les seules très grandes plateformes en ligne, l'obligation d'effectuer des évaluations des risques systémiques provoqués par ou liés au fonctionnement et à l'utilisation de leurs services et de prendre des mesures raisonnables et efficaces visant à atténuer ces risques. Elles doivent également se soumettre à des audits externes et indépendants. Les fournisseurs de très grandes plateformes en ligne qui utilisent des systèmes de recommandation doivent, outre leurs obligations de transparence, fournir au moins une option pour chacun de leurs systèmes de recommandation qui n'est pas basée sur le profilage. Dans les cas où de telles plateformes affichent de la publicité en ligne sur leur interface en ligne, elles font également l'objet d'obligations de transparence spécifiques. Elles accorderont l'accès à des données aux chercheurs pour comprendre comment évolue le risque en ligne ; elles nomment un ou plusieurs agent(s) de conformité pour assurer le respect des obligations énoncées dans le règlement ; et se voient imposer des obligations spécifiques et supplémentaires en matière de rapports sur la transparence. On note que des telles règles sont applicables aux plateformes installées dans des pays tiers à partir du moment où leurs services ciblent et atteignent des citoyens européens (pour les obligations liées aux grandes plateformes, comme dit plus haut, 10 % de la population européenne).

30. La régulation des contenus par les plateformes, en particulier par les très larges plateformes, est sans doute le point le plus délicat du débat. Celles-ci font valoir leur liberté contractuelle qui les autorise à imposer à leurs clients des limites à leur expression et de justifier les mesures de blocage des messages voire des comptes. Au-delà du cas Trump, sur lequel nous reviendrons, on sait combien la lutte contre la *coronavirus disease* (Covid) a justifié le blocage de nombre de comptes par *Twitter*, *LinkedIn*, *Facebook* y compris d'hommes et de femmes de sciences ou de chercheurs et chercheuses. À ces arguments, ne faut-il pas opposer que la liberté d'expression s'impose même vis-à-vis des pouvoirs privés et en tout cas exige le respect du principe de proportionnalité, faute de quoi on peut parler de censure privée⁸⁴, bien plus insidieuse et incontrôlable que celle publique ? C'est dans cette perspective que la régulation publique européenne impose des balises comme nous l'avons dit. En ce sens, la résolution du Parlement européen du 25 novembre 2020 « sur le renforcement de la liberté des médias : la protection des journalistes en Europe, les discours de haine, la désinformation et le rôle des plateformes », exhorte au contrôle des outils technologiques : « souligne, à cet égard, que la suppression de contenus en ligne, en l'absence de décision judiciaire déterminant leur caractère illégal, a une forte incidence sur la liberté d'expression et la liberté d'information ; demande des analyses d'impact régulières des mesures volontaires prises par les fournisseurs de services et les plateformes en matière de lutte contre la désinformation ; insiste sur les obligations des États membres de respecter, de protéger et de garantir les droits fondamentaux ; souligne que l'utilisation d'outils automatisés pour la modération du contenu peut mettre en péril la liberté d'expression et d'information, et que la politique et la stratégie numériques de l'Union doivent prévoir des voies de recours et des garanties appropriées, dans le plein respect des dispositions pertinentes de la Charte des droits fondamentaux de l'Union européenne et de la CEDH ».

31. Dans cet esprit, la Commission européenne, après avoir laissé une place importante à l'autorégulation des plateformes, a souhaité imposer via des lignes directrices contraignantes⁸⁵, dans lesquelles doivent

(82) Facebook, *Community Standards Enforcement Report*, disponible en ligne sur <https://transparency.facebook.com/community-standards-enforcement#fake-accounts>.

(83) Council of the European Union, « *Council conclusions on safeguarding a free and pluralistic media system* », *op. cit.* (note 60), n° 37.

(84) Sur ce point, parmi d'autres, lire

S. TOSZA, « Internet service providers as law enforcers and adjudicators. A public role of private actors », *CL&SR*, 2021, n° 43, 105614 et R. BADOUARD, « Les enjeux de la mo-

dération des contenus sur le web », *Revue européenne des médias*, n° 59, 2021.

(85) *Guidance for strengthening the Code of Practice - un encadrement*

s'inscrire les chartes et autres codes de conduite des opérateurs privés de ces plateformes, un certain nombre de règles. Ainsi, on relève :

- la nécessité de développer des services évitant toute manipulation des personnes vulnérables et de prendre des mesures contre les comportements manipulateurs qui pourraient être opérés via les plateformes. Les lignes directrices ajoutent l'obligation des plateformes de publier la politique suivie à cet égard ;
- l'obligation de donner les moyens d'agir aux utilisateurs s'ils sont victimes de messages illicites ou violents, mais également de mettre à la disposition des utilisateurs des mesures technologiques permettant de vérifier l'authenticité, l'exactitude et l'origine de l'information ;
- la contribution au renforcement de l'éducation aux médias ;
- la responsabilisation quant aux systèmes de recommandation ou de sélection mis en place par l'engagement à rendre les systèmes de recommandation transparents en ce qui concerne les critères utilisés pour hiérarchiser ou déprioriser les informations, avec la possibilité pour les utilisateurs de personnaliser les algorithmes de classement ;
- la garantie de visibilité des sites d'informations fiables présentant un intérêt pour le public ;
- l'obligation d'avertir les utilisateurs en cas d'interaction avec de la désinformation et de la mésinformation ;
- la garantie de l'accès à des indicateurs de fiabilité et de la visibilité des labels de vérification des faits ;
- le renforcement des moyens d'agir de la communauté des chercheurs et des vérificateurs de faits par l'accès aux données et l'obligation de coopérer ;
- enfin, la garantie d'utiliser voire d'intégrer la référence aux services des vérificateurs de faits dans les services offerts par ces opérateurs.

3. La question des « comités de surveillance » mis en place par les plateformes

32. Les plateformes, du moins les « very large », ont pris soin dans le cadre de leur autorégulation de mettre sur pied des organes plus ou moins indépendants chargés d'intervenir au cas où un internaute s'insurgerait du blocage d'un message ou pire de son compte. L'exemple le plus connu est celui de Facebook devenu Meta, qui, en mai 2020, annonça la création d'un *Oversight Board* (Conseil de surveillance) dont l'objectif est de « Protéger la liberté d'expression en prenant des décisions indépendantes et fondées sur des principes concernant des éléments de contenu importants, et en émettant des avis consultatifs sur les politiques en matière de contenu sur la plateforme »⁸⁶. Le rapport d'octobre 2021 de cet organe dit « indépendant » composé de personnes prestigieuses est impressionnant : « le Conseil de surveillance a affirmé avoir reçu en l'espace de neuf mois, pas moins de 524.000 demandes d'examen de cas, avec une augmentation significative des appels au fil des trimestres »⁸⁷. Cette augmentation se matérialise surtout dans le rapport publié en décembre 2021 qui fait état de presque 330.000 recours supplémentaires⁸⁸. L'affaire « Trump », jugée le 5 mai 2021⁸⁹, mais également d'autres affaires témoignent du souci de cet organe de ne pas se référer uniquement à la Charte Facebook dans l'examen des affaires. Ainsi, le Comité s'est référé explicitement à l'article 19 du Pacte international des droits civils et politiques (PIDCP) pour exiger conformément au principe de proportionnalité la nécessité de limiter dans le temps (6 mois) l'interdiction faite à l'ex-président d'accéder à son compte Facebook. D'autres principes semblables par ailleurs à ceux décrits *supra* en ce qui concerne l'article 10 de la Convention européenne des droits de l'homme découlent de ce pacte international : le principe de légalité exige que le texte de limitation de la liberté d'expression soit clair, intelligible et par ailleurs connu des personnes à qui il s'applique. Ainsi, le Conseil reproche à Facebook de ne pas définir le terme « désinformation », pourtant utilisé largement pour bloquer ou retirer des messages. La question de la

légitimité du but poursuivi renvoie à des motifs comme la protection des enfants, les exigences de la santé publique, la sécurité publique, la non-discrimination, les droits et libertés des tiers. Pour revenir au principe de proportionnalité, Charvin conclut : « Le Conseil a admis "qu'il peut y avoir des cas où des mots qui sont dégradants dans un contexte peuvent être plus anodins, ou même donner plus de pouvoir, dans un autre ». Il a par exemple mis en avant les responsabilités de Meta dans un monde « qui connaît vraisemblablement une recrudescence du soutien et de l'acceptation de l'idéologie néonazie » où « l'antisémitisme » est en hausse. La lutte contre la désinformation dans un contexte de pandémie mondiale et les défis que rencontre l'entreprise à cet égard ont également été soulignés de manière pragmatique.

33. Ces réflexions pourraient accréditer l'idée que l'autorégulation, mettant en place des « juridictions privées » compétentes et surtout indépendantes — encore faudrait-il que ces qualités soient exigées et vérifiées par l'autorité publique — et se soumettant aux principes universels de la liberté d'expression, rejoigne les exigences des cours et tribunaux, mis en place par l'autorité publique. Ces principes de légalité, de légitimité et de proportionnalité ont également certaines incidences sur l'opposabilité des chartes ou codes de conduite qui constituent le fondement contractuel de la possibilité d'actions des plateformes. Il est certain que la rédaction actuelle des chartes pêche par un contenu trop vague, peu transparent et sans doute aux sanctions disproportionnées. Cette conclusion optimiste ne peut cependant être suivie totalement dans la mesure où les organes de contrôle mis en place par l'autorégulation entendent parfois au nom de l'identité affirmée par les plateformes aller plus loin que les États : ainsi, en matière de racisme, le Conseil de surveillance de Facebook a entendu aller plus loin dans la répression de messages que ne l'auraient fait des juridictions étatiques. Cette divergence d'appréciation nous apparaît certes valable dans le cas où il s'agit d'un réseau réellement privé fondé sur une identité forte à laquelle adhèrent les membres mais peut-on accepter qu'un réseau qui se veut universel puisse bénéficier de cette possibilité de s'écarter des standards internationaux en matière de liberté d'expression et de leur contrôle par les autorités juridictionnelles étatiques ? Ainsi, des plateformes comme Facebook, YouTube, WhatsApp, Instagram comptent respectivement, la première, plus de 3 milliards (un internaute sur deux), la deuxième, 2,3 milliards, la troisième, 2 milliards, la quatrième, 1, 3 milliards de clients et toutes relèvent sauf YouTube (Google) du même conglomérat, à savoir META. C'est donc à raison que nombre de défenseurs de nos libertés et les associations civiles s'inquiètent de la puissance de telles sociétés sur la régulation de nos sociétés. Charvin nous rappelle à cet égard que, dans son livre visionnaire « *Online courts and the future of Justice* », Richard Susskind affirmait que « si les services du secteur privé devaient être systématiquement préférés aux tribunaux publics, nous risquerions d'être régis moins par la Loi du pays que par des normes sociales imprévisibles et de compromis, dont aucune ne cherche directement et publiquement à faire respecter la justice ».

Conclusions

34. Nos propos appellent les conclusions suivantes :

— La *cancel culture* est un phénomène souvent dénoncé tant pour des raisons légitimes (par exemple, lorsque ce qui est dénoncé constitue la diffamation d'autrui, sanctionnée alors également pénalement, ou une désinformation), que pour des motifs illégitimes (en particulier, lorsque la personne dénonciatrice utilise ce concept pour dénigrer l'opinion d'autrui et limiter sa liberté d'expression). Il importe d'emblée d'avoir un jugement critique sur les messages qualifiés comme tels, dans la mesure où le combat contre la *cancel culture* peut être tan-

dur des « online intermediaries » : Orientations de la Commission européenne, 26 mai 2021. Ces orientations visent à renforcer le *Code of Practice on Disinformation* de 2018 qui avait été signé par les grandes plateformes. Cette autorégulation par le secteur faisait l'objet d'une obligation de rapporter à la Commission les activités développées dans ce cadre.

(86) Préambule de la Charte du

Conseil de surveillance (ci-après Charte), traduction par V. NIDOR, « Le Conseil de surveillance de Facebook, service après-vente de la liberté d'expression ? », *Recueil Dalloz*, 2020, p. 1474.

(87) Conseil de surveillance, « Oversight Board transparency reports - Q4 2020, Q1 & Q2 2021 », 2021, p. 8.

(88) Sur ces chiffres et d'autres, lire

l'article de Baptiste Charvin, « L'an I du Conseil de surveillance de Meta - Puissance et faiblesses d'une instance conçue comme l'avenir de la régulation des contenus sur les plateformes ». L'auteur y décrit soigneusement le fonctionnement de ce Conseil. À noter les réflexions du sénateur Kennedy privé de l'accès à son compte et par là même de tout recours contre Facebook dans la me-

sure où seuls les clients actuels de Facebook pourraient saisir le Comité de surveillance.

(89) E. FARNoux, « La liberté d'expression et l'ordre public dans *lea lex electronica* : sur la décision *Trump vs. Facebook* du Conseil de surveillance de Facebook du 5 mai 2021 », *Gaz. Pal.*, 27 juillet 2021, pp. 12 et s.

tôt une volonté de baliser à juste titre la liberté d'expression, tantôt une limitation non légitime de la liberté d'expression.

— Le rôle de la technologie dans le développement de la *cancel culture* est évident. La rapidité de la diffusion et la possibilité de s'adresser au monde entier sont certes des caractéristiques qui expliquent ce rôle mais, au-delà, le *business model* sur lequel fonctionne l'économie des plateformes et qui se base sur l'économie de l'attention et les profits générés par la publicité, justifie pour ces dernières l'utilisation de systèmes d'intelligence artificielle permettant le ciblage des messages et constitue un facteur clé du phénomène de *cancel culture*.

— La technologie, grâce à des systèmes à nouveau d'intelligence artificielle (reconnaissance automatique de mots, d'opinions, d'auteurs...), peut à l'inverse constituer un moyen de lutter contre les messages dits de *cancel culture*, en permettant de les détecter, voire de les bloquer de manière automatique. Il importe de souligner les risques liés à l'utilisation de systèmes d'IA pour lutter contre le phénomène et parfois leur inadéquation, ainsi que la tentation pour les États de déléguer la compétence d'application de la loi dans la recherche des messages suspects. En particulier, la mise en place de ces technologies par des plateformes, véritables *gatekeepers* jouant un rôle prédominant dans notre société de l'information et de la communication, soulève des inquiétudes.

— Face à ces risques, l'Europe développe une politique réglementaire visant à préserver à la fois la liberté d'information et de presse mais également l'intérêt général, en reconnaissant l'action des *fact checkers*, en renforçant les lignes directrices applicables aux opérateurs de service de communication et d'information et surtout en contrôlant les outils technologiques de sélection des messages utilisés par les *very large platforms*. On regrette que la notion de désinformation n'ait pas été définie dans les textes réglementaires et que les limites strictes des mesures susceptibles d'être prises n'aient pas été de ce fait déterminées.

— Si nous pouvons reconnaître l'apport des organes de contrôle mis en place par ces plateformes, nous estimons que ces derniers doivent travailler sous le contrôle ultime des juges et dans le respect des principes de légalité, légitimité et proportionnalité affirmés par les textes internationaux consacrant la liberté d'expression et d'opinion. Ces principes exigent que les chartes d'autorégulation, sur lesquelles s'appuient les opérateurs pour « censurer » les messages, soient transparentes, fassent l'objet d'un réel accord de leurs clients et aient un contenu clair et précis. Par ailleurs, nous mettons en doute que ces limitations puissent valoir lorsque ces plateformes sont devenues un intermédiaire quasi obligé d'accès à l'information et à la communication. Cette situation justifie, selon notre opinion, une réglementation par l'autorité publique de ce qui est à considérer dès lors comme un service universel. Comme l'écrivent Berns et Regeluth⁹⁰, « [c]e qui est donc visé, c'est une compréhension extensive de la liberté d'expres-

sion, comme droit d'accès au débat d'intérêt général, laissant entendre donc que ce débat et donc l'espace public réclament d'être activement protégé, voire conquis : la liberté d'expression n'est donc pas envisagée seulement comme une liberté négative de nature individuelle mais comme une contribution nécessaire à la vie démocratique considérée comme habitée de rapports de lutte et de domination ». Cette conception de la liberté d'expression justifie une politique proactive en faveur de l'accès à tous des lieux de communication et d'information et ce, face aux phénomènes de restriction de l'espace public que constitue l'appropriation par les plateformes privées d'une partie importante de l'espace public.

— Nous soulignons dès lors l'importance du respect dans toute la mesure du possible de la liberté d'expression et de la presse. Selon notre opinion, la *cancel culture* ne devrait être l'objet de mesures de répression que si elle constitue de la désinformation au sens strict ou si son message est constitutif d'une infraction pénalement incriminée comme la diffamation, le message raciste ou de haine, la violation de la vie privée, etc.

— À l'attitude négative qui consiste à rechercher et punir les messages certes dérangeants et parfois heurtants, nous préférons une attitude plus positive qui ouvre à chacun la capacité d'un regard critique sur le flot d'informations généré par la toile. Nous plaidons en particulier pour la revalorisation d'une presse pluraliste, indépendante, et au personnel respectueux d'une véritable déontologie journalistique, nous prônons une éducation critique aux médias, nous encourageons la recherche à mieux comprendre les mécanismes technologiques, sociologiques, économiques et psychologiques qui expliquent le phénomène dangereux de la désinformation et de la polarisation des internautes.

— Sans doute, nos conclusions sont-elles empreintes d'idéalisme. Face à la puissance des opérateurs de plateformes, y compris vis-à-vis de nos États de moins en moins souverains et peut-être de plus en plus complices, face à la demande toujours plus croissante des citoyens de sensationnalisme et d'immédiateté, un changement du modèle économique de la toile est-il possible ? Au modèle de l'autorégulation, nos États même unis sous la bannière européenne pourront-ils imposer une réglementation effective au service des libertés y compris de la lutte contre la manipulation et l'ébranlement de nos démocraties, lorsqu'eux-mêmes — et la crise de la Covid l'a bien montré — sont tentés de suivre une voie autre ?

Noémi BONTRIDDER

Chercheuse au CRIDS/NADI

Yves POULLET

Professeur à l'UNamur et l'UCLille, co-président du NADI⁹¹

(90) T. BERNs et T. REGELUTH, *Éthique de la communication et de l'information*, éd. de l'Université libre de

Bruxelles, 2021, p. 191.

(91) Si vous souhaitez prolonger la discussion, n'hésitez pas à noe-

mi.bontridder@unamur.be et yves.poullet@unamur.be.