REVUE BELGE DE DROIT INTERNATIONAL
2018/2 – Éditions BRUYLANT, Bruxelles

LA LUTTE CONTRE LE TERRORISME PAR LA CENSURE DES « CONTENUS À CARACTÈRE TERRORISTE » : UNE INGÉRENCE JUSTIFIÉE AU DROIT À LA LIBERTÉ D'EXPRESSION ?

PAR

Justine BRAUN*

ET

France LAURENT**

Résumé

La lutte contre l'utilisation des réseaux sociaux par les terroristes est devenue une préoccupation majeure au sein de la politique intérieure des États. Dans cette perspective, la Commission européenne a proposé un règlement visant à lutter contre les contenus terroristes en ligne dans lequel il délègue aux opérateurs privés la tâche de détecter ces contenus au moyen d'outils automatisés et de les supprimer. Cet article démontre que cette proposition, dans sa formulation actuelle, entraîne une ingérence injustifiée au droit à la liberté d'expression à la lumière des critères dégagés par la Cour européenne des droits de l'homme. D'abord, la capacité d'un algorithme de distinguer les propos bénéficiant de la protection du droit à la liberté d'expression de ceux qui incitent à la haine et à la violence est mise en doute. Ensuite, la difficulté de traduire en langage informatique les éléments de définition des « contenus à caractère terroriste » ainsi que le manque de fiabilité de l'intelligence artificielle risquent de conduire à une suppression des contenus au-delà du nécessaire. Enfin, la délégation du contrôle des contenus à ces opérateurs privés entraîne un risque de censure excessive compte tenu de l'évolution du régime de responsabilité applicable à ces derniers.

Abstract

The fight against the use of social networks by terrorists has become a major concern in the domestic policies of states. In this context, the European Commission has proposed a regulation to combat online terrorist content in which it delegates to private operators the task of detecting such content by using automated tools and removing it. This article demonstrates that this proposal, in its

* Avocate au barreau de Liège.
** Chercheuse au Centre de droit international.

current wording, entails an unjustified interference with the right to freedom of expression in the light of the criteria established by the European Court of Human Rights. First, the ability of an algorithm to distinguish between discourses that are protected by the right to freedom of expression and hate speech is questioned. Secondly, the difficulty of translating into computer language the elements of the definition of "terrorist content" and the unreliability of artificial intelligence may lead to the deletion of content beyond what is necessary. Finally, the delegation of content control to private operators also entails a risk of excessive censorship given the evolution of the liability regime applicable to them.

Introduction

Depuis plusieurs années, la lutte contre l'utilisation de l'Internet par les terroristes et la prévention de la radicalisation conduisant à l'extrémisme violent sont devenues des préoccupations majeures dans la politique intérieure des États du monde entier (1). En effet, les groupes terroristes recourent de plus en plus souvent à l'Internet pour diffuser leur propagande, inciter à la violence et recruter des personnes vulnérables et/ou partageant les mêmes pensées idéologiques extrémistes (2).

Dans cette perspective, les États ont à plusieurs reprises exprimé leur souhait de voir les entreprises du secteur informatique accroître leurs efforts afin de procéder à la suppression rapide des « contenus à caractère terroriste » (3) et ont insisté sur l'importance d'une collaboration entre les secteurs public et privé — principalement avec les « Géants de l'Internet » (4) — pour contrer l'exploitation du Web à des fins terroristes.

(1) G7 Taormina Statement on the Fight Against Terrorism and Violent Extremism, 26-27 mai 2017, par. 5, disponible à l'URL : http://www.consilium.europa.eu/media/23562/26-g7-statement-fight-against-terrorism-and-violent-extremism.pdf [consulté le 10 mars 2019]. Voy. égal. « Facebook, Microsoft, Twitter and YouTube Announce Formation of the Global Internet Forum to Counter Terrorism », 26 juin 2017, disponible à l'URL : https://newsroom.fb.com/news/2017/06/global-internet-forum-to-counter-terrorism/ [consulté le 10 mars 2019].

(2) « Hard Questions: Are We Winning the War On Terrorism Online ? », by Monika Bickert, Head of Global Policy Management, and Brian Fishman, Head of Counterterrorism Policy, Facebook, 28 novembre 2017, disponible à l'URL : https://newsroom.fb.com/news/2017/11/hard-questions-are-we-winning-the-war-on-terrorism-online/ [consulté le 10 mars 2019]. Voy. égal. Commission européenne, « Lutte contre le terrorisme sur l'internet : le Forum sur l'internet prône la détection automatique de la propagande terroriste », communiqué de presse, Bruxelles, 6 décembre 2017 ; UNODC, « Utilisation de l'Internet à des fins terroristes », Nations unies, New York, mars 2014, pp. 3-12, par. 1 à 28 ; Fr. Douzet, « Le cyberespace, troisième front de la lutte contre Daesh », *Hérodote*, 2016/1 (n° 160-161), pp. 223-224.

(3) G7 Taormina Statement on the Fight Against Terrorism and Violent Extremism, 26-27 mai 2017, par. 5, *supra* note 1.

(4) « Le G7 et les géants de l'internet s'accordent pour bloquer la propagande 'terroriste' », 20 octobre 2017, disponible à l'URL : https://www.rtbf.be/info/medias/detail_le-g7-et-les-geants-de-l-internet-s-accordent-pour-bloquer-la-propagande-terroriste ?id=9742082 [consulté le 9 mars 2019].

Au niveau régional, l'Union européenne a également exprimé sa volonté de lutter contre la propagation de la radicalisation en ligne (5). La Commission européenne a lancé le « Forum de l'UE sur l'Internet » (6) visant à permettre une coopération plus efficace entre le secteur informatique, les gouvernements et les institutions de l'UE, ainsi que le déploiement de meilleurs outils pour lutter contre la propagande terroriste sur l'Internet (7).

Le 15 juin 2017, Facebook a également fait part de son propre plan d'action pour lutter contre le terrorisme (8), dans lequel il explique qu'il recourt notamment à l'intelligence artificielle et à un algorithme pour identifier, supprimer et empêcher la publication de contenus à caractère terroriste et la création de groupes, de pages ou de profils par des terroristes. Le processus est en grande partie automatisé — à savoir entièrement délégué à l'algorithme — ou semi-automatisé — c'est-à-dire accompagné *in fine* d'une vérification humaine (9). À l'heure actuelle, 99 % des contenus terroristes relatifs à l'État islamique et à Al-Qaïda sont retirés de Facebook avant que la communauté des utilisateurs n'y ait eu accès (10). Plus récemment, Mark Zuckerberg a déclaré avoir une responsabilité vis-à-vis de la sécurité des utilisateurs, ce qui impliquait de décider de ce qui relève notamment de la propagande terroriste et du discours haineux (11).

Lors d'une réunion de haut niveau aux Nations unies en date du 20 septembre 2017, le Royaume-Uni, la France et l'Italie ont exprimé leur souhait d'aller encore plus loin dans la prévention de l'utilisation de l'Internet par les terroristes. Dans leur déclaration commune (12), ces États requièrent

(5) Conclusions du Conseil européen sur la sécurité et la défense, communiqué de presse, 22 juin 2017, par. 1, disponible à l'URL : http://www.consilium.europa.eu/fr/press/press-releases/2017/06/22/euco-security-defence/ [consulté le 11 mars 2019]. Voy. égal. Commission européenne, « Lutte contre le terrorisme sur l'internet : le Forum sur l'internet prône la détection automatique de la propagande terroriste », communiqué de presse, Bruxelles, 6 décembre 2017.

(6) Commission européenne, « Forum de l'UE sur l'internet : réunir les gouvernements, Europol et les entreprises du secteur de l'internet pour lutter contre les contenus à caractère terroriste et les discours à la haine en ligne », Communiqué de presse, 3 décembre 2015, Bruxelles, *supra* note 8.

(7) Commission européenne, Le programme européen en matière de sécurité, COM(2015) 185 final, Strasbourg, 28 avril 2015, p. 16.

(8) « Hard questions : How we coutner terrorism » by Monika Bickert, Director of Global Policy Management, and Brian Fishman, Couternterrorism Policy Manager, 15 juin 2017, disponible à l'URL : https://newsroom.fb.com/news/2017/06/how-we-counter-terrorism/ [consulté le 5 février 2019].

(9) Conseil de l'Europe, « Algorithmes et droits humains — Études sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires, étude du Conseil de l'Europe », DGI(2017)12, 2018. Disponible à l'adresse suivante : https://rm.coe.int/algorithms-and-human-rights-fr/1680795681, p. 21.

(10) « Hard Questions: Are We Winning the War On Terrorism Online ? », *supra* note 2.

(11) M. Zuckerberg, « Four ideas to regulate the Internet », 30 mars 2019, disponible à l'URL : https://newsroom.fb.com/news/2019/03/four-ideas-regulate-internet/ [consulté le 15 avril 2019].

(12) Déclaration du Royaume-Uni, de la France et de l'Italie — Réunion de haut niveau sur la prévention de l'utilisation d'internet par les terroristes — Nations unies, 20 septembre 2017, disponible à l'URL : https://onu.delegfrance.org/Ensemble-lutter-contre-le-terrorisme-sur-internet [consulté le 5 février 2019].

notamment des entreprises technologiques qu'elles suppriment les contenus à caractère terroriste dans un délai d'une à deux heures après leur publication, qu'elles veillent à ce que ces contenus ne soient pas à nouveau téléchargés sur d'autres plateformes, et qu'elles empêchent même leur mise en ligne dès le départ, le tout via l'utilisation d'un algorithme. Très récemment, le gouvernement du Royaume-Uni a dévoilé dans un Livre blanc la proposition que les responsables de médias sociaux soient tenus personnellement responsables des contenus préjudiciables distribués sur leurs plateformes (13). L'Allemagne s'est déjà dotée d'une telle loi, en vigueur depuis le 1er janvier 2018 (14). L'Australie a déjà modifié sa législation suite à l'attentat de Christchurch (15) pour rendre les intermédiaires et leurs dirigeants directement responsables pour les contenus publiés sur leurs plateformes (16).

Dans ce contexte, le problème actuel majeur est que les intermédiaires sont dans la plupart des cas encouragés à supprimer eux-mêmes ce type de contenu, sans base juridique claire et précise (17). Lorsqu'une telle base légale existe, elle s'avère souvent différente d'un État à l'autre. Néanmoins, dans une tentative d'étendre cette politique à l'ensemble des États européens, le 12 septembre 2018, la Commission européenne a proposé un nouveau règlement visant à faire supprimer du web les contenus à caractère terroriste et garantissant que les mêmes obligations s'imposent uniformément dans toute l'Union (18). Le 8 avril 2019, la Commission des libertés civiles, de la justice et des affaires intérieures du Parlement européen a approuvé à 35 voix contre 1 et 8 abstentions la proposition de règlement, en y proposant certaines modifications (19). Le 17 avril 2019, le Parlement européen a proposé plusieurs modifications au texte initial (20), et notamment à

(13) *The Guardian*, «Social media bosses could be liable for harmful content, leaked UK plan reveals», disponible à l'URL: https://www.theguardian.com/technology/2019/apr/04/social-media-bosses-could-be-liable-for-harmful-content-leaked-uk-plan-reveals [consulté le 15 avril 2019].

(14) Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG), 1st September 2017, *Federal Law Gazette I*, pp. 3352 et s.

(15) *Le Monde*, «L'Australie adopte une loi punissant les réseaux sociaux qui ne modèrent pas assez vite les images d'actes terroristes», disponible à l'URL: https://www.lemonde.fr/pixels/article/2019/04/04/l-australie-adopte-une-loi-punissant-les-reseaux-sociaux-qui-ne-moderent-pas-assez-vite-les-images-d-actes-terroristes_5445743_4408996.html [consulté le 15 avril 2019].

(16) Pour le texte de la loi, voy. https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/s1201_aspassed/toc_pdf/1908121.pdf;fileType=application/pdf [consulté le 15 avril 2019].

(17) Conseil de l'Europe, «Algorithmes et droits humains — Études sur les dimensions des droits humains dans les techniques de traitement automatisé des données et éventuelles implications réglementaires», *op. cit.*, p. 21.

(18) Commission européenne, «État de l'Union 2018: la Commission propose de nouvelles règles pour faire supprimer du web les contenus à caractère terroriste», communiqué de presse, 12 septembre 2018, Strasbourg, disponible à l'URL: http://europa.eu/rapid/press-release_IP-18-5561_fr.htm [consulté le 15 avril 2019].

(19) Voy. http://www.europarl.europa.eu/doceo/document/A-8-2019-0193_EN.html?redirect [consulté le 15 avril 2019].

(20) Résolution législative du Parlement européen du 17 avril 2019 sur la proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, COM(2018)0640, C8-0405/2018, 2018/0331(COD).

la définition des contenus à caractère terroriste afin de la rapprocher des définitions contenues dans la directive 2017/541 (21). Si ces modifications pourraient atténuer certaines des critiques émises dans les pages suivantes, elles ne permettent toutefois pas de les résorber au point de rendre vains les développements qui suivront.

Compte tenu de la multiplication, ces dernières années, des actions entreprises par les États pour contraindre les intermédiaires de l'Internet à lutter contre la publication des contenus terroristes en ligne au moyen de l'intelligence artificielle, une analyse en termes de légalité semble s'imposer à ce stade. En effet, tant les différentes législations et déclarations adoptées par les États que la proposition de règlement pour la suppression des contenus à caractère terroriste en ligne, en ce qu'ils prévoient la suppression presque automatique, au moyen d'algorithmes, et la censure *a priori* des messages à contenu terroriste par des acteurs privés, nécessitent à notre sens un examen plus approfondi au regard de la liberté d'expression, droit garanti par l'article 19 du Pacte international relatif aux droits civils et politiques (ci-après le PIDCP) (22), ainsi que par l'article 10 de la Convention européenne des droits de l'Homme (ci-après la CEDH) (23).

En effet, si le droit à la liberté d'expression n'est pas absolu, les ingérences dans l'exercice de ce droit n'en demeurent pas moins encadrées et limitées. Elles ne peuvent être justifiées que si elles sont prévues par la loi, nécessaires dans une société démocratique, et proportionnées au but légitime poursuivi (24). Ces trois conditions prévalent même lorsque l'ingérence entre dans le cadre de la lutte contre le terrorisme (25).

Il ne fait pas de doute que la lutte contre le terrorisme, la suppression rapide et la censure *a priori* des messages à caractère terroriste poursuivent plusieurs objectifs légitimes. En effet, en ce que ces mesures visent à empêcher notamment la diffusion de la propagande, l'incitation à la violence, la radicalisation et le recrutement de « nouveaux terroristes » via les réseaux sociaux (26),

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

(22) Pacte international des droits civils et politiques, New York, 16 décembre 1966, *R.T.N.U.*, vol. 999, p. 171, article 19.

(23) Convention de sauvegarde des Droits de l'Homme et des Libertés fondamentales, Rome, 4 novembre 1950, *R.T.N.U.*, vol. 1496, p. 232, article 10.

(24) Voy. Pacte international des droits civils et politiques, *op. cit.*, article 19 §3 ; C.D.H., Observation générale No. 34, Article 19 : Liberté d'opinion et liberté d'expression, U.N. Doc. CCPR/C/GC/34, par. 22 ; C.D.H., *Velichkin c. Bélarus*, 20 octobre 2005, requête n° 1022/2001, U.N. Doc. CCPR/C/85/D/1022/2001 (2005).

(25) Conseil de l'Europe, « Lignes directrices du Comité des Ministres du Conseil de l'Europe sur les droits de l'homme et la lutte contre le terrorisme », adoptées par le Comité des ministres lors de sa 804e réunion, Strasbourg, 11 juillet 2002, III. Voy. égal. Cour eur. D.H., arrêt du 2 octobre 2008, affaire *Leroy c. France*, requête n° 36109/03, par. 37.

(26) UNODC, « Utilisation de l'Internet à des fins terroristes », *op. cit.*, pp. 3-12, par. 1 à 28.

elles permettent d'assurer la sauvegarde de la sécurité nationale et de l'ordre public, ainsi que la défense de l'ordre et la prévention de crimes (27).

Néanmoins, les restrictions apportées au droit à la liberté d'expression doivent encore répondre à un « besoin social impérieux » (28) et s'avérer « proportionnées au(x) but(s) légitime(s) poursuivi(s) » (29). Ces restrictions doivent donc permettre la restauration du fonctionnement normal de la société, et ne doivent pas être établies dans le but d'assouvir des volontés politiques contraires aux principes démocratiques (30). Ainsi, s'il n'appartient pas à la Cour de se substituer aux autorités nationales, les motifs invoqués par ces dernières doivent être « suffisants et pertinents » (31). Ces trois critères développés dans l'arrêt *Handyside* ne reçoivent néanmoins pas nécessairement une application simultanée par la Cour, cette dernière préférant l'un ou l'autre selon le cas faisant l'objet du contrôle (32).

S'il est donc vrai que les États bénéficient d'un certain pouvoir d'appréciation pour établir l'existence d'un tel « besoin social impérieux » ainsi que d'une certaine marge d'appréciation pour décider des mesures adéquates, ce pouvoir est loin d'être illimité et s'accompagne d'un contrôle juridictionnel (33). Il appartient normalement aux juridictions nationales et, en cas de recours, à la Cour européenne des droits de l'Homme, de déterminer *in fine* si, compte tenu des circonstances concrètes et particulières de l'affaire qui lui est soumise (34), l'ingérence d'un État partie à la CEDH ne compromet pas l'essence même de la liberté d'expression (35). La Cour devra considérer l'ingérence litigieuse à la lumière de l'ensemble de l'affaire, en particulier la teneur des propos reprochés et le contexte dans lequel ils ont pris place (36).

(27) À titre d'exemple, voy. Cour eur. D.H., affaire *Leroy c. France*, *op. cit.*

(28) Cour eur. D.H., arrêt du 26 novembre 1991, affaire *Observer et Guardian c. Royaume-Uni*, requête n° 13585/88, par. 59.

(29) C.D.H., Observation générale n° 34 précitée, par. 33 ; Cour eur. D.H., arrêt du 7 décembre 1976, affaire *Handyside c. Royaume-Uni*, requête n° 5493/72, p. 24, par. 49.

(30) M. Tümay, « The concept of 'necessary in a democratic society' in restriction of fundamental rights: a reflection from european convention on human rights », *Human Rights Review*, vol. I, n° 2, décembre 2011, p. 2 ; S. Greer, « Les exceptions aux articles 8-11 de la Convention européenne des Droits de l'Homme », in *Dossiers sur les droits de l'Homme No. 15*, Éditions du Conseil de l'Europe, 1997, p. 14.

(31) Cour eur. D.H., arrêt du 7 décembre 1976, affaire *Handyside c. Royaume-Uni*, requête n° 5493/72, p. 24, par. 50.

(32) S. Van Drooghenbroeck, *La proportionnalité dans le droit de la convention européenne des droits de l'homme, prendre l'idée au sérieux*, Bruylant, Bruxelles, 2001, n° 96, pp. 84-85.

(33) Cour eur. D.H., arrêt du 8 juillet 1999, affaire *Ceylan c. Turquie*, requête n° 23556/94, par. 32.

(34) Cour eur. D.H., affaire *Sunday Times c. Royaume-Uni*, *op. cit.*, par. 65 ; Cour eur. D.H., affaire *Lingens c. Autriche*, *op. cit.*, par. 43 ; Cour eur. D.H., arrêt du 26 avril 1991, affaire *Ezelin c. France*, requête n° 11800/85, par. 51 ; Cour eur. D.H., affaire *Ceylan c. Turquie*, arrêt précité, par. 32 et 35.

(35) A. Callamard, « Liberté d'expression et sécurité nationale : équilibrer pour protéger », in *Aperçu et analysé élaborés au profit de la Columbia Freedom of Expression, Modules de formation pour les juges*, janvier 2016, point I.2.1.

(36) Cour eur. D.H., arrêt du 23 septembre 1998, affaire *Lehideux et Isorni c. France*, requête n° 55/1997/839/1045, par. 51.

Pour ce faire, la Cour de Strasbourg effectue une balance des intérêts en conflit dans l'affaire, de manière à les mesurer et à évaluer leur force respective (37). Ce faisant, la Cour est en mesure d'évaluer la compatibilité du but et de la nécessité d'une atteinte au droit à la liberté d'expression avec l'article 10, § 2, de la CEDH (38).

En l'espèce, si ces différentes législations venaient à être soumises au contrôle de la Cour, celle-ci procéderait à une mise en balance entre d'une part, le droit des individus à la liberté d'expression et, d'autre part, le droit légitime d'une société démocratique de se protéger elle-même contre les actes terroristes notamment en cherchant à éradiquer toute radicalisation (39). Notons néanmoins que les critères énoncés aux deux paragraphes précédents ont vocation à s'appliquer à l'examen de cas particuliers par la Cour, et non à l'examen des régimes de limitation dans leur principe. Lorsque la Cour est amenée à évaluer, non pas un cas particulier, mais la conformité d'un régime ou d'une loi en tant que tel, comme ce fût le cas des régimes de surveillance secrète, un changement de perspective s'opère. Dans ce cas, la Cour laisse aux autorités nationales une grande latitude dans le choix des moyens propres à atteindre l'objectif (40). En contrepartie toutefois, la Cour devra être convaincue de l'existence des garanties adéquates et effectives, de nature à éviter les abus éventuels (41). En particulier, « la Cour doit rechercher si les procédures de supervision de la décision et la mise en œuvre des mesures restrictives sont de nature à circonscrire l'ingérence à ce qui est nécessaire dans une société démocratique » (42). Dès lors, les conclusions auxquelles serait susceptible de parvenir la Cour pourraient éventuellement s'avérer différentes selon qu'elle est amenée à se prononcer sur une application particulière du règlement ou sur le régime mis en place par le règlement lui-même, *via* certaines législations nationales.

À cet égard, la Cour a rappelé dans l'affaire *Yildirim c. Turkey* (43) que « de telles restrictions présentent [...] de si grands dangers qu'elles appellent

(37) M. Tümay, « The concept of 'necessary in a democratic society' in restriction of fundamental rights: a reflection from european convention on human rights », *op. cit.*, pp. 4-5 ; Ch. Girard, « La liberté d'expression : état des questions », *Raisons politiques*, 2016/3 (n° 63), p. 21.

(38) S. Greer, « Les exceptions aux articles 8-11 de la Convention européenne des Droits de l'Homme », *op. cit.*, p. 16.

(39) À titre d'exemple, voy. Cour eur. D.H., arrêt du 29 novembre 2011, affaire *Kiliç et Eren c. Turquie*, requête n° 43807/07, par. 25. Voy. égal. M. Geistlinger, « Fight against Terrorism and Limitation of the Freedom of Expression: Some Remarks on Recent Judgments of the European Court of Human Rights », 61 *Collection Papers Fac. L. Nis.* 49 (2012), p. 56.

(40) Cour eur. D.H., décision sur la recevabilité du 29 juin 2006, affaire *Weber et Saravia c. Allemagne*, requête n° 54934/00, par. 106.

(41) Cour eur. D.H., arrêt du 13 septembre 2018, affaire *Big Brother Watch et autres c. Royaume-Uni*, requêtes n^{os} 58170/13, 62322/14 et 24960/15, par. 308.

(42) Cour eur. D.H., arrêt du 4 décembre 2015, affaire *Roman Zakharov c. Russie*, requête n° 14881/03, par. 232.

(43) Cour eur. D.H., arrêt du 18 décembre 2012, affaire *Yildirim c. Turkey*, requête n° 3111/10.

de la part de la Cour l'examen le plus scrupuleux » (44) et que « l'information est un bien périssable et en retarder la publication, même pour une brève période, risque fort de la priver de toute valeur et de tout intérêt » (45). Par conséquent, bloquer l'accès à l'Internet ou supprimer un contenu publié en ligne nécessite « un cadre légal particulièrement strict quant à la délimitation de l'interdiction et efficace quant au contrôle juridictionnel contre les abus éventuels » (46).

Comme l'a soulevé récemment le rapporteur spécial pour la liberté d'opinion et d'expression des Nations unies, « les obligations relatives à la surveillance et à la suppression rapide de contenus [...] ont été renforcées par la mise en place de cadres répressifs susceptibles de compromettre la liberté d'expression » (47). En particulier, c'est d'abord l'utilisation de l'intelligence artificielle — en particulier lorsqu'elle ne s'accompagne pas de vérifications humaines — pour détecter et supprimer certains contenus, qui soulève plusieurs problèmes liés à la justification de l'ingérence. Ensuite, c'est la délégation de cette compétence normalement dévolue à l'État et à ses organes, à des acteurs privés, tels que Facebook, qui apparaît problématique au regard des limitations autorisées tant par la CEDH que par le Pacte. Il serait néanmoins artificiel de tenter de faire entrer chacune de ces problématiques dans le critère de légalité ou dans celui de nécessité, la Cour elle-même s'attardant tantôt sur la légalité, tantôt sur la nécessité, les deux critères étant inextricablement liés. C'est pourquoi nous avons fait le choix d'articuler les développements qui suivront autour des questions soulevées par l'utilisation de l'intelligence artificielle, et qui sont transversales aux critères de légalité et de nécessité.

Il est désormais bien établi que la protection de la liberté d'expression est exclue lorsque le discours concerné est identifié comme une incitation à la haine, en application de l'article 17 de la CEDH (48). Pour les autres types de discours, les limitations doivent s'évaluer au regard des conditions énoncées au § 2 de l'article 10 de la CEDH (ou du § 3 de l'article 19 du PIDCP).

Notre étude analysera les difficultés posées par la mise en œuvre des exclusions ou limitations de discours par les algorithmes, dans le domaine de la prévention du terrorisme, au regard des principes relatifs au respect de la liberté d'expression. Tout d'abord, nous analyserons la capacité des algorithmes de distinguer les propos bénéficiant de la protection du droit à la liberté

(44) *Ibid.*, par. 47.

(45) *Ibid.*

(46) *Ibid.*, par. 64.

(47) Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/HCR/38/35, 6 avril 2018, par. 16.

(48) Cour eur. D.H., arrêt du 27 juin 2017, affaire *Belkacem c. Belgique*, requête n° 34367/14, par. 30 et 31 ; voy. égal. Cour eur. D.H., décision sur la recevabilité du 18 mai 2004, affaire *Seurot c. France*, requête n° 57383/00 ; Cour eur. D.H., décision sur la recevabilité du 20 février 2007, affaire *Pavel Ivanov c. Russie*, requête n° 35222/04.

d'expression de ceux qui incitent à la haine et à la violence, compte tenu du caractère assez flou de la distinction dans la jurisprudence de la CEDH (I). Ensuite, dans la mise en œuvre des limitations à la liberté d'expression, nous mettrons en évidence les interrogations suscitées par la traduction en langage informatique des éléments de la définition européenne du « contenu à caractère terroriste », éléments qui eux-mêmes soulèvent de nombreuses questions (II). Enfin, nous étudierons les difficultés liées au fait de la délégation à des opérateurs privés de fonctions de police normalement dévolues à l'État. Les opérateurs de l'Internet n'ont pas la qualité ni l'impartialité des acteurs étatiques et risquent de censurer au-delà du strict nécessaire, compte tenu de l'évolution du régime de responsabilité qui leur est applicable (III).

Nous précisons d'emblée que notre analyse en termes juridiques s'inscrit principalement dans le cadre développé au sein du Conseil de l'Europe, et se base sur le régime établit par la Cour européenne des droits de l'homme, bien que ce dernier ne diffère substantiellement pas du régime établi par le PIDCP. Nous limiterons également notre analyse au cas particulier de Facebook, bien que les conclusions tirées de cette étude soient également transposables à d'autres plateformes telles que YouTube ou Twitter.

I. — Les difficultés liées à l'identification par un algorithme des discours incitant à la haine ou la violence, exclus de la liberté d'expression

La Cour européenne des droits de l'homme a souligné que « l'Internet est aujourd'hui devenu l'un des principaux moyens d'exercice par les individus de leur droit à la liberté d'expression et d'information : on y trouve des outils essentiels de participation aux activités et débats relatifs à des questions politiques ou d'intérêt public » (49). En particulier, les réseaux sociaux sont devenus un outil important pour l'exercice de cette liberté d'expression (50). Ainsi, les publications postées sur l'Internet entrent dans le champ d'application de la liberté d'expression (51).

Toutefois, quelles sont les publications protégées ? En règle, si certaines publications bénéficient de la protection du droit à la liberté d'expression — et relèvent effectivement du champ d'application des articles 19 du PIDCP et 10 de la CEDH — d'autres n'en bénéficient pas, et tombent sous le coup des articles 5, § 1, du PIDCP et 17 de la CEDH. Certes, « la liberté d'expression vaut non seulement pour les "informations" ou "idées" accueillies avec faveur

(49) Cour eur. D.H., affaire *Yildirim c. Turquie*, *op. cit.*, par. 54.

(50) B. F. Jackson, « Censorship and Freedom of Expression in the Age of Facebook », *New Mexico Law Review*, vol. 44, 2014, p. 124.

(51) Voy. Cour eur. D.H., Division de la recherche, *Internet : la jurisprudence de la Cour européenne des droits de l'homme (Mis à jour en juin 2015)*, 2011, p. 18 ; ainsi que C.D.H., Observation générale No. 34, *op. cit.*, par. 12.

ou considérées comme inoffensives ou indifférentes, mais aussi pour celles qui heurtent, choquent ou inquiètent l'État ou une fraction quelconque de la population »(52). Néanmoins, « on peut juger nécessaire, dans les sociétés démocratiques, de sanctionner, voire de prévenir, toutes les formes d'expression qui propagent, incitent à, promeuvent ou justifient la haine fondée sur l'intolérance »(53). Ainsi, bien que la liberté d'expression puisse recouvrir des idées heurtant la population, elle s'arrête dès lors qu'elle propage des idées qui incitent à la haine ou à la discrimination, ou encore qui tendent à faire l'apologie du terrorisme. C'est ce qu'expriment les articles 5, § 1, du PIDCP et 17 de la CEDH, lesquels interdisent tout abus de droit et prévoient qu'aucune disposition tant du Pacte que de la Convention ne peut « être interprétée comme impliquant [...] un droit quelconque de se livrer à une activité ou d'accomplir un acte visant à la destruction des droits ou libertés » reconnus dans un des deux instruments(54). Ainsi, lorsque la Cour identifie un discours incitant à la haine, deux voies s'offrent à elle. Dans le premier cas, elle pourra exclure la protection garantie par l'article 10 si les propos sont « dirigés contre les valeurs qui sous-tendent la Convention »(55), comme elle l'a fait à maintes reprises(56). Dans le second cas, si le discours, tout en étant haineux, n'était pas dirigé à l'encontre des valeurs qui sous-tendent la Convention, elle pourra se limiter à vérifier la légalité des limitations prévues au § 2. Si cette première distinction peut déjà prêter à confusion, la jurisprudence de la Cour, ainsi que l'absence de définition du « discours de haine » et de critères clairs pour l'identifier, rendent l'appréciation des contenus particulièrement difficile. Dès lors, nous pouvons légitimement questionner la capacité d'un algorithme, éventuellement assisté d'une vérification humaine, de discerner les propos qui bénéficient de la protection de la liberté d'expression, de ceux

(52) Cour eur. D.H., arrêt du 7 décembre 1976, affaire *Handyside c. Royaume-Uni*, requête n° 5493/72, p. 23, par. 49 ; voy. égal. en ce sens Cour eur. D.H., arrêt du 26 avril 1979, affaire *Sunday Times c. Royaume-Uni*, requête n° 6538/74 ; Cour eur. D.H., arrêt du 8 juillet 1986, affaire *Lingens c. Autriche*, requête n° 9815/82 ; Cour eur. D.H., affaire *Observer et Guardian c. Royaume-Uni*, *op. cit.*

(53) Cour eur. D.H., arrêt du 6 juillet 2006, affaire *Erbakan c.* Turquie, requête n° 59405/00, par. 56.

(54) Pacte international relatif aux droits civils et politiques, *op. cit.*, article 5, par. 1 ; ainsi que Convention de sauvegarde des Droits de l'Homme et des Libertés fondamentales, *op. cit.*, article 17. À ce sujet, voy. égal. Cour eur. D.H., arrêt du 1er juillet 1961, affaire *Lawless c. Irlande (No. 3)*, requête n° 332/57, par. 7 ; ainsi que Fr. Dubuisson, « La lutte contre le terrorisme et la liberté d'expression : le cas de la répression de l'apologie du terrorisme », in S. Jacopin et A. Tardieu (dir.), *La lutte contre le terrorisme*, Paris, Pedone, 2018, p. 275.

(55) Cour eur. D.H., décision sur la recevabilité du 18 mai 2004, affaire *Seurot c. France*, requête n° 57383/00.

(56) Voy. not. Cour eur. D.H., décision du 11 octobre 1979, affaires *Glimmerveen et Hagenbeek c. Pays-Bas*, requêtes nos 8348/78 et 8406/78 ; décision du 6 septembre 1995, affaire *Remer c. Allemagne*, requête n° 25096/94 ; décisions de la Commission des 11 octobre 1979, 6 septembre 1995 et 24 juin 1996 respectivement, affaire *Marais c. France*, requête n° 31159/96 ; décision du 23 septembre 1998, affaire *Lehideux et Isorni c. France*, requête n° 24662/94, par. 47 ;, décision du 24 juin 2003, affaire *Garaudy c. France*, requête n° 65831/01 ;, décision du 16 novembre 2004, affaire *Norwood c. Royaume-Uni*, requête n° 23131/03 ; décision du 13 décembre 2005, affaire *Witzsch c. Allemagne*, requête n° 7485/03.

qui n'en bénéficient pas et relèvent du discours de haine. L'examen du raisonnement tenu par la Cour dans quatre affaires nous amène néanmoins à douter de l'utilité de cette étape, et atteste de son caractère purement symbolique.

Dans l'affaire *Belkacem c. Belgique*, était en cause la condamnation de Monsieur Belkacem, dirigeant et porte-parole de l'organisation « Sharia4Belgium », qui avait incité, dans ses propos, à combattre les personnes non musulmanes et à les dominer. Tout en reconnaissant la possibilité de soustraire certains propos de la protection de l'article 10 de la CEDH via le jeu de l'article 17, la Cour européenne des droits de l'homme a précisé que :

> « L'article 17 ne s'applique qu'à titre exceptionnel et dans des hypothèses extrêmes. [...] Dans les affaires relatives à l'article 10 de la Convention, il ne doit être employé que s'il est tout à fait clair que les propos incriminés visaient à faire dévier cette disposition de sa finalité réelle par un usage du droit à la liberté d'expression à des fins manifestement contraires aux valeurs de la Convention [...]. La question déterminante sur le terrain de l'article 17 est de savoir si les propos du requérant avaient pour but d'attiser la haine ou la violence *et si, en les tenant, il a cherché à invoquer la Convention de manière à se livrer à une activité ou à commettre des actes visant à la destruction des droits et libertés consacrés* » (57) (nous soulignons).

En l'espèce, la Cour avait considéré que le discours concerné ne pouvait relever de la liberté d'expression car une attaque aussi générale et véhémente était en contradiction avec les valeurs de tolérance, de paix sociale et de non-discrimination qui sous-tendent la Convention. Néanmoins, force est de constater que la limite entre les propos qui relèvent de la liberté d'expression et ceux qui n'en relèvent pas reste floue. En effet, dans l'affaire *Gündüz c. Turquie*, le requérant avait revendiqué l'application de la Sharia et avait ouvertement critiqué le gouvernement. La Cour avait alors estimé que « des expressions visant à propager, inciter à ou justifier la haine fondée sur l'intolérance, y compris l'intolérance religieuse, ne bénéficient pas de la protection de l'article 10 de la Convention. Toutefois, [...] le simple fait de défendre la Sharia, sans en appeler à la violence pour l'établir, ne saurait passer pour un "discours de haine" » (58). Il faut ajouter à cela le raisonnement paradoxal tenu par la Cour dans l'affaire *Leroy c. France*. Dans cette dernière, la Cour a considéré qu'un dessin glorifiant les attentats du 11 septembre 2001 ne rentrait pas « dans le champ d'application des publications qui se verraient soustraites par l'article 17 de la Convention à la protection de l'article 10 » (59). D'une part, le message de fond visé par le requérant ne visait pas la négation de droits fondamentaux. D'autre part, « la Cour est d'avis que le dessin litigieux et le commentaire qui l'accompagne ne constituent

(57) Cour eur. D.H., arrêt du 27 juin 2017, affaire *Belkacem c. Belgique*, requête n° 34367/14, par. 30 et 31.

(58) Cour eur. D.H., arrêt du 4 décembre 2003, affaire *Gündüz c. Turquie*, requête n° 35071/97, par. 51.

(59) Cour eur. D.H., affaire *Leroy c. France*, *op. cit.*, par. 27.

pas une justification à ce point non équivoque de l'acte terroriste qui les feraient échapper à la protection garantie par l'article 10 de la liberté de la presse »(60). Néanmoins, et de manière tout à fait étonnante, la Cour a ensuite conclu que la condamnation de l'auteur dudit dessin pour complicité d'apologie du terrorisme ne violait pas l'article 10 de la CEDH, et ce précisément parce que le dessin en cause promouvait effectivement des actes terroristes(61). Dans ce cas, il apparaît peu cohérent d'à la fois conclure que le dessin relève de l'apologie du terrorisme tout en considérant qu'il ne nie pas les valeurs protégées par la convention(62). Par la suite, dans une affaire *Roj TV A/S c. Danemark*(63), la Cour a estimé que la chaîne requérante, qui avait diffusé des vidéos de propagande du PKK, ne pouvait quant à elle bénéficier de la liberté d'expression car elle avait tenté d'utiliser ce droit à des fins contraires à la Convention, notamment en incitant les téléspectateurs à la violence et en soutenant une activité terroriste.

À la lumière de ces différents exemples, force est de constater l'absence de critères clairs permettant d'identifier les propos relevant du « discours de haine », ainsi que la catégorie de ces derniers qui comprendrait les propos allant à l'encontre des valeurs défendues par la Convention. En l'absence de tels critères, l'élaboration d'un algorithme chargé d'analyser ces contenus paraît difficilement concevable. Or, l'examen de légalité des limitations prescrites par l'article 10, § 2, ne trouvera à s'appliquer que dans le premier cas, tandis que les propos ne bénéficiant d'aucune protection pourront être censurés sans conditions tant par les États que par les plateformes elles-mêmes. Néanmoins, puisqu'il est toujours nécessaire, pour se prononcer sur la recevabilité de la requête, de s'intéresser aux propos tenus ainsi qu'à leur qualification, le défi pour les algorithmes chargés de détecter et de supprimer de tels contenus ne se situe pas ici, mais plutôt dans la détection de ce qui constitue effectivement un « contenu à caractère terroriste ».

II. — Les difficultés liées à la définition des « contenus à caractère terroriste » et à leur traduction en langage informatique

Trois problèmes liés à la traduction en langage informatique de la définition des « contenus à caractère terroriste » peuvent être identifiés. Le premier résulte de la difficulté qu'il y a à dégager une définition suffisamment claire

(60) *Ibid.*, par. 27.

(61) *Ibid.*, par. 42-48.

(62) Concernant l'incohérence du raisonnement de la Cour dans l'affaire *Leroy c. France*, voy. not. Fr. Dubuisson, « La lutte contre le terrorisme et la liberté d'expression : le cas de la répression de l'apologie du terrorisme », *op. cit.*, pp. 288-289.

(63) Cour eur. D.H., décision sur la recevabilité, affaire *Roj TV A/S c. Danemark*, requête n° 24683/14, par. 44-49.

du « contenu à caractère terroriste » (A). Le second est lié aux limites actuelles des algorithmes, qui sont peu capables d'interpréter la nature exacte des discours en tenant compte de leur contexte, de l'intention de leur auteur ou de leurs effets potentiels (B). Enfin, le dernier problème tient à la diversité des définitions nationales de l'infraction terroriste qui rend encore plus compliquée toute évaluation globale d'un « contenu terroriste » qu'il s'agirait de supprimer (C). Ces trois éléments peuvent faire douter du caractère « nécessaire et proportionné » du système mis en place, faisant reposer sur les opérateurs de l'Internet et leurs algorithmes la suppression des contenus.

A. — *Le problème lié à l'absence de définition précise des « contenus à caractère terroriste »*

Tel que cela a été précisé, la première difficulté réside dans l'identification d'une définition internationale du « contenu à caractère terroriste », afin de satisfaire au critère de légalité. Ce critère est appréhendé d'une manière similaire par le PIDCP et par la CEDH. En effet, une loi imposant des restrictions à la liberté d'expression doit être libellée avec suffisamment de précision, et être suffisamment accessible et prévisible, pour permettre à un individu d'adapter son comportement en fonction de la règle, et pour permettre à un autre individu de juger de ce comportement (64). Selon le rapporteur spécial sur la promotion et la protection des droits de l'Homme et des libertés fondamentales dans la lutte antiterroriste, « pour que l'interdiction soit prescrite par la loi, il faut que la loi soit suffisamment accessible de sorte que chacun sache dans quelles limites il doit inscrire son comportement ; et qu'elle soit libellée en termes suffisamment précis pour que chacun ait un comportement adapté » (65). En matière de lutte contre le terrorisme, le Comité des droits de l'Homme des Nations unies a également pris soin de préciser l'importance d'adopter des lois compatibles avec le paragraphe 3 de l'article 19 du Pacte, en particulier de définir de manière précise les infractions dans ce domaine, en évitant des formules trop vagues telles que l'« encouragement du terrorisme » et l'« activité extrémiste » (66), ainsi que le fait de « louer », « glorifier » ou « justifier » le terrorisme (67). Ainsi, à propos de l'« encouragement du terrorisme », le Comité a déjà invité le Royaume-Uni à modifier l'article 1er du

(64) C.D.H., Observation générale No. 34, *op. cit.*, par. 25.

(65) Commission des droits de l'homme, Rapport du Rapporteur spécial sur la promotion et la protection des droits de l'homme et des libertés fondamentales dans la lutte antiterroriste, Martin Scheinin, UN Doc. E/CN.4/2006/98, 28 décembre 2005, par. 46.

(66) C.D.H., Observations finales concernant le rapport de la Fédération de Russie, 1er décembre 2003, U.N. Doc. CCPR//CO/79/RUS, par. 20.

(67) C.D.H., Observation générale No. 34 précité, par. 46.

Terrorism Act de 2006 (68) en raison de la définition trop vague et générale qu'il donnait de cette infraction (69).

Si par le passé, les juristes et les auteurs ont pu affirmer qu'il n'existait aucune définition universelle de l'« infraction terroriste » au sens large (70), il est à notre sens à tout le moins possible d'identifier un dénominateur commun aux différentes définitions (71), dont la substance est identique à la définition de l'infraction terroriste reprise dans la directive 2017/541. Si notre intention n'est pas de mettre de côté les débats entourant les contours précis de cette définition, tels que la question de l'exclusion ou l'inclusion des activités conduites par les forces armées en cas de conflit armé ou des mouvements de libération nationale (72), nous nous contentons de les mentionner, préférant centrer notre analyse sur les « contenus à caractère terroriste ».

Les « contenus à caractère terroriste » ne sont définis dans aucun instrument juridique international contraignant. Si cette expression est directement liée à la définition donnée à l'« infraction terroriste » au sens large et aux différentes infractions terroristes plus spécifiques, il n'en demeure pas moins nécessaire d'évaluer la précision de ce que recouvrent les « contenus à caractère terroriste ». Pour les États membres de l'Union européenne, une recommandation de la Commission européenne du 1er mars 2018 définit le contenu à caractère terroriste comme « toute information dont la diffusion constitue une infraction au sens de la directive (UE) 2017/541 ou une infraction terroriste au sens de la législation de l'État membre concerné, ce qui englobe la diffusion d'informations de ce type émanant de groupes ou d'entités terroristes figurant sur les listes établies par l'Union ou les Nations unies,

(68) Terrorism Act 2006 (UK).

(69) C.D.H., Observations finales concernant le rapport du Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, 30 juillet 2008, U.N. Doc. CCPR/C/GBR/CO/6, par. 26.

(70) Commission des droits de l'Homme, Rapport du Rapporteur spécial sur la promotion et la protection des droits de l'homme et des libertés fondamentales dans la lutte antiterroriste, *op. cit.*, par. 25 ; Commission des droits de l'Homme, Le droit à la liberté d'opinion et d'expression, Rapport de M. Ambeyi Ligabo, Rapporteur spécial, présenté en application de la résolution 2002/48 de la Commission, U.N. Doc. E/CN.4/2003/67, 30 décembre 2012, par. 66 ; B. Saul, « Attempts to Define "Terrorism" in International Law », *Netherlands International Law Review (NILR)*, 52, 2005, pp. 57-83 ; G. P. Fletcher, « The Indefinable Concept of Terrorism », *Journal of International Criminal Justice (JICJ)*, 4 (5), 2006, pp. 894-911 ; Th. Weigend, « The Universal Terrorist. The International Community Grappling with a Definition », *JICJ*, 4 (5), 2006, pp. 912-932 ; A. Verdebout, « La définition coutumière du terrorisme d'Antonio Cassese : de la doctrine au Tribunal spécial pour le Liban », *Droit et société*, nº 88, 2014/3, pp. 709-728 ; R. Young, « Defining Terrorism: The Evolution of Terrorism as a Legal Concept in International Law and Its Influence on Definitions in Domestic Legislation », *Boston College International and Comparative Law Review*, vol. 29, nº 1, Winter 2006, pp. 23-102.

(71) Fr. Dubuisson, « La définition du "terrorisme" : débats, enjeux et fonctions dans le discours juridique », *Confluences Méditerranée*, 2017/3 (nº 102), p. 32.

(72) Annexe II du Rapport du Comité spécial créé par la résolution 51/210 de l'Assemblée générale, en date du 17 décembre 1996, Seizième session (8-12 avril 2013), Doc. A/68/37, pp. 16-19.

ou attribuables auxdits groupes ou entités » (73). De manière plus précise et complète, l'article 2(5) de la proposition de règlement relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne du 12 septembre 2018 définit les « contenus à caractère terroriste » comme :

> « une ou plusieurs des informations suivantes, qui :
>
> (a) provoquent à la commission d'infractions terroristes, ou font l'apologie de telles infractions, y compris en les glorifiant, ce qui entraîne un risque que de tels actes soient commis ;
>
> (b) encouragent la participation à des infractions terroristes ;
>
> (c) promeuvent les activités d'un groupe terroriste, notamment en encourageant la participation ou le soutien à un groupe terroriste au sens de l'article 2, paragraphe 3, de la directive (UE) 2017/541 ;
>
> (d) fournissent des instructions sur des méthodes ou techniques en vue de la commission d'infractions terroristes ».

Nous pouvons d'emblée rendre compte de deux problèmes majeurs dans la formulation actuelle de cette définition. Premièrement, si le point (a) est directement inspiré de l'infraction de « provocation publique à commettre une infraction terroriste » définie à l'article 5 de la directive 2017/541, il est plus large que cette dernière. L'article 2(5)(a) de la proposition de règlement ne mentionne aucun élément intentionnel particulier, à l'inverse de l'infraction de « provocation publique à commettre une infraction terroriste » contenue à l'article 5 de la directive. Par ailleurs, les rapporteurs spéciaux aux droits de l'homme ont critiqué le caractère particulièrement large de l'infraction de « provocation à commettre une infraction terroriste » contenue dans la directive, qui englobe les activités de plaidoyer indirectes et la « glorification », tout en fixant un seuil bas en exigeant seulement que le comportement « crée un danger » que des infractions « puissent être commises », par opposition aux actes qui créent un risque réel ou un danger imminent de préjudice (74). Les rapporteurs spéciaux concluent ainsi que « *The définition as it stands could encompass legitimate forms of expression, such as reporting conducted by journalists and human rights organizations on the activities of terrorist groups and on counter-terrorism measures taken by authorities, in violation of the right to freedom of expression* » (75) et recommandent d'adopter la définition modèle de l'incitation au terrorisme énoncée dans les pratiques optimales en matière de lutte antiterroriste qui énonce que :

> « Constitue une infraction le fait de diffuser ou de mettre un message à disposition du public par tout autre moyen, délibérément et illégalement, *avec l'intention*

(73) Commission européenne, Recommandation de la Commission européenne du 1er mars 2018 sur les mesures destinées à lutter de manière efficace contre les contenus illicites en ligne, C(2018) 1177 final, Bruxelles, 1er mars 2018, pp. 11-12, par. 4, h).

(74) Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, *OL* OTH 71/2018, 7 December 2018, p. 3.

(75) *Ibid.*, p. 3.

d'inciter à la commission d'une infraction terroriste, lorsqu'un tel comportement, qu'il préconise expressément ou non la commission d'infractions terroristes, *crée un danger* qu'une ou plusieurs de ces infractions soient commises » (76) (nous soulignons).

Deuxièmement, une observation opérée par l'Agence des droits fondamentaux de l'Union européenne concerne le cadre des définitions de la directive 2017/541 dont s'inspire la proposition de règlement. En effet, les définitions de la « provocation publique au terrorisme », du « recrutement pour le terrorisme » et de « dispenser un entraînement pour le terrorisme » relèvent du domaine pénal. En d'autres termes, ces définitions devront être transposées en droit interne et leurs éléments seront appréciés dans le cadre d'une procédure pénale assortie des garanties du procès équitable. Tel ne sera néanmoins pas le cas des définitions contenues dans la proposition de règlement, qui seront appliquées par différentes entités, y compris des compagnies privées, sans garanties aussi strictes (77).

Dans sa version actuelle, la définition des « contenus à caractère terroriste » inscrite dans la proposition de règlement présente encore des lacunes au regard du critère de légalité. À supposer même que cette définition soit modifiée dans le règlement final afin d'y répondre, ainsi qu'aux critiques soulevées dans les documents mentionnés – et dont nous n'avons pu relater qu'une partie – d'autres problèmes relatifs à la traduction en langage informatique de ces définitions ainsi qu'à la diversité des définitions nationales doivent alors être examinés.

B. — *Le problème lié à la traduction en langage informatique des éléments des définitions*

Tel que cela a déjà été précisé dans l'introduction, les restrictions apportées au droit à la liberté d'expression doivent s'avérer proportionnées au(x) but(s) légitime(s) poursuivi(s) (78). Or, la difficulté qui résulte de la traduction en langage informatique des différents éléments de définition, et le risque de « sur-censure » des contenus qui y est associé, nous amène précisément à douter de la proportionnalité d'une telle ingérence.

Afin de saisir la difficulté de transposer de manière informatique des éléments subjectifs, il semble essentiel à ce stade de revenir sur la notion même d'algorithme. En effet, ces derniers sont usuellement définis comme un ensemble de règles opératoires dont l'application permet de résoudre un

(76) Rapport du Rapporteur spécial sur la promotion et la protection des droits de l'homme et des libertés fondamentales dans la lutte antiterroriste – Dix pratiques optimales en matière de lutte antiterroriste, Martin Scheinin, A/HCR/16/51, Pratique 8, p. 17.

(77) Opinion of the European Union Agency for Fundamental Rights, Proposal for a Regulation on preventing the dissemination of terrorist content online and its fundamental rights implications, 12 February 2019, 2/2019, Vienna, pp. 17-18.

(78) C.D.H., Observation générale No. 34, *op. cit.*, par. 34.

problème énoncé au moyen d'un nombre fini d'opérations (79). Le rapporteur spécial pour la promotion et la protection du droit à la liberté d'opinion et d'expression les définit comme « des suites d'opérations informatiques conçues et encodées par l'homme sous forme d'instructions transformant des données d'entrée en résultats informatifs ou conclusifs » (80). Cela signifie, dans le cas de Facebook, que l'analyse de certains mots, images ou vidéos, associés à certaines personnes et à une multitude de données, lui permet d'identifier des contenus relatifs au terrorisme. Il s'agit toutefois d'une analyse purement informatique, sans lien avec la réalité matérielle, et qui ne tient compte d'aucun facteur extérieur aux données numériques. Nous notons néanmoins que de nouveaux développements permettent désormais de plus en plus d'analyser les sentiments et intentions des utilisateurs de réseaux sociaux grâce aux outils de *Deep learning*, *Sentiment Analysis* et *Intent Analysis* (81). Dans le cas de *l'Intent Analysis*, elle permet notamment d'analyser l'intention de l'utilisateur derrière un message en déterminant si elle concerne une opinion, une actualité, un marketing, une plainte, une suggestion, une appréciation ou une requête (82).

Deux points relatifs à la traduction de ces définitions en langage informatique ont retenu notre attention. Ces problématiques sont, comme nous l'avons précisé, particulièrement liées à la proportionnalité de l'ingérence dans la mesure où elles entraînent le risque de censurer un grand nombre de propos protégés. D'abord, l'appréciation du contexte propre et spécifique à la commission d'une infraction terroriste – ou, dans ce cas précis, d'un « contenu à caractère terroriste » – ne semble actuellement pas à la portée d'un « être informatique » autre qu'un juge, comme en atteste notamment le manque de fiabilité de l'intelligence artificielle (a). Ensuite, il sera démontré qu'un algorithme n'est pas capable d'appliquer les éléments d'intention et de risque qui sont pourtant essentiels à la définition de la « provocation publique à commettre des actes terroristes » et dont est directement inspirée l'exigence de suppression des « contenus à caractère terroriste » (b).

a) *L'appréciation du contexte propre à l'infraction terroriste*

Au sein de l'Union européenne, la directive 2017/541 définit l'infraction terroriste par (i) une série d'actes matériels tels que définis par le droit interne, (ii) qui, par leur nature ou leur contexte, peuvent porter gravement atteinte

(79) Pour une définition plus complète, voy. la Stanford Encyclopedia of Philosophy, disponible à l'URL : https://plato.stanford.edu/entries/turing-machine/#Bib [consulté le 5 août 2019].

(80) Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/73/348, 29 août 2018, par. 5.

(81) Pour des développements plus complets relatifs au *Sentiment Analysis*, voy. B. Liu, *Sentiment Analysis and Opinion Mining*, Graeme Hirst, Morgan & Claypool, 2012.

(82) Voy. https://blog.paralleldots.com/product/contextual-sentiment-analysis-applications/?fbclid=IwAR0GaDnxhIlDcEPFXBoppgF8liNl3VtyY2iqyH5MmkTDj1Z00qdHwdbi1-k [consulté le 5 août 2019].

à un pays ou une organisation internationale (83), et (iii) commis « dans le but de : a) gravement intimider une population ; b) contraindre indûment des pouvoirs publics ou une organisation internationale à accomplir ou à s'abstenir d'accomplir un acte quelconque ; c) gravement déstabiliser ou détruire les structures fondamentales politiques, constitutionnelles, économiques ou sociales d'un pays ou une organisation internationale » (84).

Si les problèmes résultant de la diversité des dispositions internes seront examinés plus loin, la condition relative au contexte est d'une importance particulière en matière de terrorisme. Dès lors, son appréciation apparaît essentielle lorsqu'on s'intéresse à la traduction en langage informatique de cette définition. En particulier, un arrêt rendu par la Cour constitutionnelle belge au sujet de cette condition nous apparaît particulièrement pertinent pour attester de l'importance de cette condition (85). Cet arrêt concernait la transposition par la Belgique de la définition de l'infraction terroriste reprise dans la décision-cadre de 2002 relative à la lutte contre le terrorisme (86) et la critique portait sur le critère de légalité. En effet, la loi belge définissait l'infraction terroriste comme l'infraction qui « de par sa nature ou son contexte, peut porter gravement atteinte à un pays ou à une organisation internationale » (87), ainsi que par un élément intentionnel directement repris de la décision-cadre. La ministre de la Justice avait par ailleurs précisé, lors des travaux préparatoires de cette loi, qu'« il appartiendra aux cours et tribunaux d'apprécier au cas par cas si, par le contexte dans lequel l'infraction est commise, celle-ci porte gravement atteinte à un pays ou à une organisation internationale » (88). Dans cet arrêt, la Cour a considéré que le principe de légalité n'empêchait pas que la loi attribue un pouvoir d'appréciation au juge (89). Après avoir reconnu que l'élément intentionnel pourrait donner lieu à des difficultés d'interprétation, elle conclut que la loi est constitutionnelle car le juge pourra apprécier ces éléments constitutifs de l'infraction en tenant compte des éléments propres à chaque situation particulière (90). Ainsi, la Cour conclut à la constitutionnalité de la loi en raison de l'appréciation judiciaire qui pourra être faite de chaque situation.

Une telle appréciation n'est pas transposable de manière algorithmique et Facebook, qui n'a pas la qualité ni les ressources d'un acteur juridictionnel, n'a pas de légitimité pour apprécier lui-même ce qui relève du terrorisme.

(83) Directive (UE) 2017/541 précitée, article 3, § 1.

(84) *Ibid.*, article 3, § 2.

(85) C.C., 13 juillet 2005, n° 125/2005.

(86) Cette définition est en substance la même que celle reprise dans la directive 2017/541. Voy. décision-cadre du Conseil du 13 juin 2002 relative à la lutte contre le terrorisme (2002/475/JAI), *J.O.C.E.*, L 164, 22 juin 2002.

(87) C. pén., article 137, § 1er, inséré par la loi du 19 décembre 2003 relative aux infractions terroristes, *M.B.*, 29 décembre 2003, article 3.

(88) *Doc. parl.*, Chambre, 2003-2004, DOC 51-0258/004, p. 14.

(89) C.C., 13 juillet 2005, n° 125/2005, B.6.2.

(90) C.C., 13 juillet 2005, n° 125/2005, B.7.2.

L'aggravation du risque de censure dû à l'utilisation d'outils automatisés a été soulevée par le rapporteur spécial pour la promotion et la protection du droit à la liberté d'opinion et d'expression, qui précise que « la modération de contenu pilotée par IA présente plusieurs inconvénients, notamment la difficulté d'évaluer le contexte *et de prendre en compte la grande variabilité des indices langagiers, des significations et des particularités linguistiques et culturelles* » (91) (nous soulignons). De plus, « à la différence des humains, les algorithmes sont actuellement incapables d'évaluer le contexte culturel, de détecter l'ironie d'un discours ou de procéder à l'analyse critique requise pour reconnaître avec précision, par exemple, un contenu "extrémiste" ou un discours haineux, en conséquence de quoi *ils risquent davantage de procéder par défaut au blocage ou à la restriction de certains contenus*, et ainsi de porter atteinte au droit qu'a chaque utilisateur d'être entendu et d'accéder aux moyens d'information sans restrictions ni censure » (nous soulignons) (92).

En effet, tel que nous pouvons le constater, ce système n'est pas totalement fiable. Actuellement, Facebook fait l'objet de critiques virulentes pour avoir censuré plusieurs photos d'œuvres d'art qui ont été assimilées à des photos pornographiques, telles que notamment des photos du tableau « La liberté guidant le peuple » réalisé par le peintre Eugène Delacroix (93), ou encore de statuette du Paléolithique supérieur nommée « la Vénus de Willendorf » (94). À ce sujet, l'entreprise américaine vient d'être déclarée responsable pour avoir désactivé le compte d'un utilisateur qui avait publié une photo du tableau du peintre Gustave Courbet intitulé « L'Origine du monde » représentant un sexe féminin, et ce « sans préavis ni justificatif » (95).

(91) Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/73/348, 29 août 2018, par. 15.

(92) Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/73/348, 29 août 2018, par. 29.

(93) « Facebook censure le sein nu de Marianne », 17 mars 2018, disponible à l'URL : https://www.lalsace.fr/actualite/2018/03/17/facebook-censure-le-sein-nu-de-marianne [consulté le 15 mars 2019] ; « Facebook censure les seins nus de "La liberté guidant le peuple", 18 mars 2018, disponible à l'URL : http://www.lesoir.be/146182/article/2018-03-18/facebook-censure-les-seins-nus-de-la-liberte-guidant-le-peuple [consulté le 15 mars 2019] ; « Censure de *La Liberté guidant le peuple* : Facebook reconnaît avoir fait "une erreur" », 19 mars 2018, disponible à l'URL : http://www.lefigaro.fr/theatre/2018/03/19/03003-20180319ARTFIG00073-censure-de-la-liberte-guidant-le-peuple-facebook-reconnait-avoir-fait-une-erreur.php [consulté le 15 mars 2019].

(94) J. Lausson, « Facebook a censuré provisoirement la Vénus de Willendorf », 2 mars 2018, disponibles à l'URL : https://www.numerama.com/politique/333256-facebook-a-censure-provisoirement-la-venus-de-willendorf.html [consulté le 15 mars 2019] ; « "Vénus de Willendorf" : Facebook présente ses excuses pour avoir censuré la photo de la statuette », 2 mars 2018, disponible à l'URL : https://www.20 minutes.fr/culture/2230519-20180302-venus-willendorf-facebook-presente-excuses-avoir-censure-photo-statuette [consulté le 15 mars 2019].

(95) « "L'Origine du monde" censurée : pourquoi Facebook risque une condamnation », 30 janvier 2018, disponible à l'URL : https://culturebox.francetvinfo.fr/arts/peinture/l-origine-du-monde-censure-pourquoi-facebook-risque-une-condamnation-268655 [consulté le 15 mars 2019] ; « "L'Origine du monde" de Gustave Courbet censurée sur Facebook », 22 février 2011, disponible à l'URL : https://culturebox.francetvinfo.fr/lorigine-du-monde-de-gustave-courbet-censuree-sur-facebook-50467 [consulté le 15 mars 2019].

L'utilisateur avait alors assigné Facebook en justice afin d'obtenir la réactivation de son compte, au nom de « la liberté d'expression sur les réseaux sociaux » (96). Bien que le règlement de Facebook interdise les publications contenant de la nudité et la pornographie sur son site, ce même règlement autorise la publication de « photos de peintures, sculptures et autres œuvres d'art illustrant des personnages nus » (97). À la date du 15 mars 2018, Facebook a été reconnu responsable et fautif par le tribunal de grande instance de Paris pour la fermeture du compte de l'utilisateur. L'entreprise américaine n'a fait l'objet d'aucune condamnation, uniquement parce que le tribunal a considéré que l'utilisateur en question n'avait pas subi de préjudice dans le cas d'espèce (98). Nous pouvons encore mentionner le système DeepText d'Instagram, pour lequel le mot « mexicain » a été assimilé à une insulte, car les textes ayant alimenté les bases de données associaient ce terme à l'adjectif « illégal », qui était lui-même codé comme négatif dans l'algorithme (99).

In casu, dans le contexte de la suppression des publications et des contenus à caractère terroriste, comment être totalement sûr que Facebook ne va pas aller au-delà de la suppression des publications prônant le terrorisme ou la radicalisation, par exemple en supprimant des publications exprimant l'opinion de l'opposition au gouvernement d'un État, ou encore des publications qui ne cherchent pas à glorifier mais bien à dénoncer les actes terroristes ? Ces exemples montrent toute la difficulté pour un algorithme, qui ne dispose que d'informations numériques, d'apprécier le contexte propre à chaque contenu publié.

Au regard de ces conclusions, et à moins qu'il soit démontré qu'une évaluation du contexte basée exclusivement sur des données récoltées sur le Web soit effectivement possible, une suppression automatique des contenus apparaît problématique au regard du critère de proportionnalité de l'ingérence. Certes, la quantité de données téléchargées quotidiennement rend pratiquement impossible le suivi et l'inspection des contenus sans avoir recours aux

(96) *Ibid.*

(97) Voy. les standards de la communauté de Facebook, disponibles à l'URL : https://www.facebook.com/communitystandards#nudity [consulté le 15 mars 2019] ; voy. égal. O. CHICHEPORTICHE, « L'Origine du Monde censuré par Facebook : responsable mais pas condamné », 16 mars 2018, disponible à l'URL : http://www.zdnet.fr/actualites/l-origine-du-monde-censure-par-facebook-responsable-mais-pas-condamne-39865606.htm [consulté le 15 mars 2019].

(98) J. VERGELY, « Facebook et la censure de "L'Origine du monde" : une faute, mais sans préjudice, dit le tribunal », 15 mars 2018, disponible à l'URL : http://www.telerama.fr/medias/facebook-vs-lorigine-du-monde-la-justice-considere-quil-y-a-eu-faute,-mais-ne-condamne-pas, n5528912.php [consulté le 15 mars 2019].

(99) N. THOMPSON, « Instagram's Kevin Systrom wants to clean up the &#%@! Internet », *Wired*, 14 août 2017. Disponible à l'adresse suivante : www.wired.com/2017/08/instagram-kevin-systrom-wants-to-clean-up-the-internet [consulté le 15 mars 2019] ; voy. égal. Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/73/348, 29 août 2018, par. 15.

outils utilisant l'intelligence artificielle (100). D'ailleurs, Facebook affirme lui-même « *at Facebook's scale neither human reviewers nor powerful technology will prevent all mistakes* » (101). Toutefois, si Facebook affirme que plusieurs millions de contenus liés au terrorisme, dont ceux relatifs à Al-Qaeda, ont été supprimés (102), il n'existe pas de chiffres statistiques relatifs aux contenus supprimés qui entraient — ou non — dans la sphère de protection de la liberté d'expression ainsi que dans les limitations autorisées. Il semble donc particulièrement difficile de mener une étude en termes d'échelle, et de conclure que les contenus abusivement supprimés l'ont été de manière proportionnée à l'objectif poursuivi, compte tenu du pourcentage de contenus véritablement terroristes qui auront été supprimés grâce aux outils automatisés.

b) *La difficile transposition des éléments spécifiques : l'intention de l'auteur ainsi que le risque qu'une infraction terroriste soit commise en raison du contenu publié*

L'appréciation du contexte propre à l'infraction terroriste par des outils automatisés ne constitue pas le seul problème au regard de la justification de l'ingérence. En effet, les infractions terroristes spécifiques requièrent fréquemment une intention particulière de l'auteur du message et le risque qu'une infraction terroriste soit commise. En effet, puisque la publication de messages à contenu terroriste sur les réseaux sociaux et sur Facebook en particulier pose principalement problème en termes de « provocation publique à commettre des actes terroristes », les éléments constitutifs de cette infraction et leur traduction en langage informatique seront successivement examinés.

La « provocation publique à commettre des actes terroristes » est définie par la Convention du Conseil de l'Europe pour la prévention du terrorisme et par la directive européenne 2017/541, comme consistant en « la diffusion ou toute autre forme de mise à disposition du public d'un message, avec *l'intention d'inciter* à la commission d'une infraction terroriste, lorsqu'un tel comportement, qu'il préconise directement ou non la commission d'infractions terroristes, *crée un danger* qu'une ou plusieurs de ces infractions puissent être commises » (103). Si une telle définition pose peu de problèmes en cas de provocation directe, les frontières restent floues avec ce qui relève de la liberté d'expression lorsqu'il s'agit de provocation indirecte. Cette difficulté semble néanmoins partiellement résolue par l'exigence supplémentaire d'un

(100) T. Van Benthem, « Social media actors in the fight against terrorism: technology and its impact on human rights », *Cambridge International Law Journal*, 2018, vol. 7, n° 2, p. 289.

(101) « Hard Questions: What Are We Doing to Stay Ahead of Terrorists ? », *op. cit.*

(102) *Ibid.*

(103) Voy. Convention du Conseil de l'Europe pour la prévention du terrorisme, Varsovie, 16 mai 2005, *e.v.* le 1er juin 2007, *S.T.C.E.*, No. 196, Article 5, § 1, ainsi que directive (UE) 2017/541 précitée, article 5.

élément intentionnel particulier (104). Prenant en considération les réflexions du Commissaire aux droits de l'Homme du Conseil de l'Europe (105), le rapport explicatif de la Convention du Conseil de l'Europe pour la prévention du terrorisme insiste sur le fait que l'on ne peut parler de « provocation publique à commettre une infraction terroriste » que si deux conditions sont réunies (106). D'une part, l'auteur doit avoir expressément l'intention d'inciter à la commission d'une infraction terroriste, et la provocation doit être commise illégalement et intentionnellement. D'autre part, « l'acte considéré doit créer un risque de commission d'une infraction terroriste » (107). Si l'infraction de « provocation au terrorisme » est désormais mieux encadrée, sa traduction en langage informatique pose question. En effet, concernant cette deuxième condition, le rapport explicatif précise que « pour évaluer si un tel risque est engendré, il faut prendre en considération la nature de l'auteur et du destinataire du message, ainsi que le contexte dans lequel l'infraction est commise [...]. L'aspect significatif et la nature crédible du risque devraient être pris en considération lorsque cette disposition est appliquée [...] ». Si ces conditions délimitent plus spécifiquement l'infraction de provocation au terrorisme, le véritable problème tient à la traduction numérique de ces deux éléments. En effet, l'intention expresse requise dans le chef de l'auteur et le risque de commission d'un acte terroriste, qui constituent des critères subjectifs, peuvent certes être décelés par un juge, mais peuvent difficilement l'être par un algorithme, lequel fonde son analyse sur des informations purement numériques et peut difficilement apprécier le contexte (108), tel que cela a déjà été souligné. À nouveau, la Cour constitutionnelle belge a récemment insisté sur l'importance de cette seconde condition, déclarant inconstitutionnelle une loi du 6 août 2016 qui supprimait de la définition de l'infraction de provocation au terrorisme les mots suivants : « lorsqu'un tel comportement, qu'il préconise directement ou non la commission d'infractions terroristes, crée le risque qu'une ou plusieurs de ces infractions puissent être commises » (109). La Cour y rappelle également son arrêt n° 9/2015 du 28 janvier 2015 (110), où elle a jugé que « la référence au risque que soient commises une ou plusieurs des

(104) Voy. Fr. Dubuisson, « La lutte contre le terrorisme et la liberté d'expression : le cas de la répression de l'apologie du terrorisme », *op. cit.*, pp. 277-278.

(105) Voy. Avis du Commissaire aux droits de l'Homme, Alvaro Gil-Robles, sur le projet de convention pour la prévention du terrorisme, Comm. D.H. (2005)1, 2 février 2005, disponible à l'URL : https://wcd.coe.int/ViewDoc.jsp?id=979807&Site=COE [consulté le 10 mars 2019].

(106) Rapport explicatif de la Convention du Conseil de l'Europe pour la prévention du terrorisme, Varsovie, 16 mai 2005, par. 99, disponible à l'URL : http://www.conventions.coe.int/Treaty/FR/Reports/Html/196.htm [consulté le 10 mars 2019].

(107) *Ibid.*, par. 99-100.

(108) M. Hertig Randall, « Freedom of Expression in the Internet », *Swiss. Rev. Int'l & Eur.*, vol. 26, 2016, p. 244 ; A. Ollo, « EDRi writes to EU Commissioner Gabriel about tackling illegal content online », 20 octobre 2017, https://edri.org/edri-writes-to-eu-commissioner-gabriel-about-tackling-illegal-content-online/ [consulté le 5 mars 2019].

(109) C.C. (Belgique), 15 mars 2018, n° 31/2018, B.7.3. à B.8.

(110) C.C. (Belgique), 28 janvier 2015, n° 9/2015.

infractions était définie de manière suffisamment claire pour être compatible avec le principe de légalité et qu'il appartient au juge d'exercer son pouvoir d'appréciation et d'examiner si ce risque est fondé sur des "indices sérieux" en tenant compte de l'identité de la personne qui diffuse le message ou le met à la disposition du public, de son destinataire, de sa nature et du contexte dans lequel il est formulé »(111). Comme en atteste ce dernier exemple, le pouvoir d'appréciation du risque *in concreto* par le juge apparaît comme une condition essentielle pour satisfaire au critère de légalité. Dès lors, la traduction en langage informatique d'une telle définition — en particulier du risque — et son application par des acteurs privés, restent fortement problématiques.

À ce sujet, Facebook reconnaît lui-même qu'elle fait à présent face à un défi, dans le sens où un algorithme qui parvient à détecter une image d'iconographie terroriste ne distingue pas nécessairement le terroriste qui partage une image, d'une personne qui partage la même image dans le but d'informer la communauté des internautes(112). Ainsi, il est possible que le soutien écrit d'une personne à une cause particulière puisse être détecté et perçu par les algorithmes de Facebook comme relevant de la « provocation publique à commettre une infraction au terrorisme »(113). Autrement dit, les messages visant à informer et à dénoncer le terrorisme risquent d'être assimilés aux véritables messages de propagande terroriste, et sont les cibles potentielles d'une censure *a priori*.

Ainsi, à moins qu'il puisse être démontré qu'une analyse de risques soit réalisable au moyen d'outils automatisés, il semble encore prématuré à ce stade de déléguer un tel pouvoir d'appréciation à ces nouveaux outils, sans contrôle juridictionnel en parallèle.

C. — *Le problème lié à la prise en compte de la diversité des définitions de l'infraction terroriste en droit interne*

Tel que cela a été mis en exergue, la définition de l'infraction terroriste contenue dans la directive (UE) 2017/541 renvoie à une première définition commune aux États membres de l'UE, mais également à la législation de ces États membres, qui peut donc s'avérer différente dans chacun d'entre eux. Cette différence est particulièrement notable en ce qui concerne l'infraction de « glorification » du terrorisme adoptée par plusieurs États.

(111) C.C., (Belgique), 15 mars 2018, n° 31/2018, B.7.5.

(112) « Hard Questions: Are We Winning the War On Terrorism Online? », by Monika Bickert, Head of Global Policy Management, and Brian Fishman, Head of Counterterrorism Policy, Facebook, 28 novembre 2017, *supra* note 2.

(113) S. M. Boyne, « Free Speech, Terrorism, and European Security: Defining and Defending the Political Community », *Pace L. Rev.*, vol. 30, 2010, p. 451.

Sur le plan national, depuis 2001, les États ont pris de plus en plus de mesures antiterroristes, en usant de définitions de plus en plus vagues (114). De telles lois, avec de telles définitions, risquent de qualifier énormément de personnes, telles que les opposants politiques et les défendeurs des droits humains, de terroristes (115). Puisque la présente contribution a notamment pour but d'analyser la conformité de la déclaration commune du Royaume-Uni, de la France et de l'Italie au droit à la liberté d'expression, nous avons examiné plus attentivement les définitions consacrées par le droit interne de ces trois pays.

Tout d'abord, concernant le Royaume-Uni, la définition de l'« infraction terroriste » adoptée dans le *Terrorism Act 2000* (116) et applicable au *Terrorism Act de 2006* a été considérée comme trop large par le *Joint Committee on Human Rights*, et comportait un risque considérable d'incompatibilité avec le droit à la liberté d'expression énoncé à l'article 10 de la CEDH (117). En effet, cette loi criminalise, outre les actes généralement considérés comme « terroristes », d'autres formes de comportements qui ne peuvent être considérés comme du « terrorisme » (118). D'une part, la définition peut être lue comme incluant les rassemblements et manifestations légitimes. D'autre part, la définition inclut de nombreux actes ou types de comportements qui, tout en étant illégaux, n'atteignent pas le niveau auquel les mesures extraordinaires intrusives prévues par la législation antiterroriste peuvent être utilisées à juste titre (119). Le Comité des droits de l'homme a d'ailleurs demandé à plusieurs reprises aux États d'adopter une définition plus précise de cette infraction (120). Ensuite, si la définition donnée par la France des faits constitutifs d'acte de terrorisme peut déjà faire l'objet de critiques en raison de l'utilisation de termes vagues et imprécis (121), nous reviendrons plus spécifiquement sur l'infraction de « l'apologie du terrorisme », qui apparaît la plus problématique au regard du principe de légalité lorsqu'on s'intéresse aux contenus à caractère terroriste. Enfin, bien que moins d'informations

(114) P. Hoffman, « Human Rights and Terrorism », *Hum. Rts. Q.*, vol. 26, 2004, p. 938.

(115) *Ibid.*

(116) Terrorism Act 2000 (UK), article 1.

(117) House of Commons, Joint Committee on Human Rights, The Council of Europe Convention on the Prevention of Terrorism, 2007, First Report of Session 2006-07, HL 26/HC 247, p. 12, par. 26.

(118) Article 19, *The impact of UK anti-terror laws on Freedom of Expression*, Londres, avril 2006, p. 4, disponible à l'URL : https://www.article19.org/data/files/pdfs/analysis/terrorism-submission-to-icj-panel.pdf [consulté le 5 février 2019].

(119) *Ibid.*, p. 5.

(120) Voy. Observations finales concernant le rapport du Canada, 20 avril 2006, U.N. Doc. CCPR/C/CAN/CO/5, par. 12 ; Observations finales concernant le rapport de la Norvège, 21 avril 2006, U.N. Doc. CCPR/C/NOR/CO/5, par. 9 ; Observations finales concernant le rapport de l'Islande, 25 avril 2005, U.N. Doc. CCPR/CO/83/ISL, par. 10.

(121) C. pén. (France), article 421-1 modifié la loi n° 2014-1353 du 13 novembre 2014 renforçant les dispositions relatives à la lutte contre le terrorisme, *op. cit.*, article 1 ; Observations finales concernant le rapport de la France, 17 août 2015, U.N. Doc. CCPR/C/FRA/CO/5, par. 10.

nous soient accessibles concernant l'Italie, outre le fait que cet État n'ait pas adopté une loi spécifique autorisant le retrait automatique des contenus à caractère terroriste sur l'Internet, sa loi relative au terrorisme (122) introduit le crime de « comportement inspiré par le terrorisme » (123). Ce dernier est défini de manière très ouverte, puisque le terrorisme renvoie lui-même à « tout autre comportement défini comme terroriste ou commis aux fins du terrorisme par des conventions ou d'autres dispositions de droit international contraignantes pour l'Italie ». Encore une fois, une définition aussi large apparaît problématique du point de vue du respect des droits de l'homme (124).

Concernant plus particulièrement l'infraction de « glorification » ou d'« apologie » du terrorisme (selon l'expression utilisée), elle n'existe actuellement dans aucun instrument juridique international. Elle est reprise dans le *Terrorism Act de 2006* du Royaume-Uni et dans le Code pénal français (125) et apparaît également problématique face au critère de légalité. En effet, si la résolution 1624 du Conseil de Sécurité a appelé les États à réprimer la glorification du terrorisme qui inciterait à commettre d'autres actes, elle n'a pas condamné la glorification en tant que telle, et a réaffirmé le droit à la liberté d'expression et les conditions prévues à l'article 19 du PIDCP (126). La France a pourtant fait le choix, en 2014, d'incriminer expressément « le fait de provoquer directement à des actes de terrorisme *ou* de faire publiquement l'apologie de ces actes » (127), sans définir précisément la notion d'« apologie ». Or, l'utilisation de la conjonction de coordination « ou » indique que l'apologie du terrorisme ne doit pas être comprise comme un élément de l'infraction de provocation au terrorisme, mais bien comme une infraction distincte. Il faut ajouter à cela que l'application qui a été faite de la notion d'apologie du terrorisme dans plusieurs affaires atteste d'une conception très large de cette infraction (128). Nous venons pourtant de souligner les deux conditions nécessaires à la « provocation au terrorisme » — l'intention de l'auteur et le risque — qui ont été considérées comme essentielles par le Conseil de l'Europe afin de ne pas porter atteinte à la liberté d'expression, conditions qui

(122) Loi nº 55 du 31 juillet 2005 contenant des « mesures urgentes pour lutter contre le terrorisme international ».

(123) C. pén. (Italie), article 207*sexies*.

(124) Voy., à ce sujet, A. Saccucci, « Italian 2005 Anti-Terrorism Legislation in the Light of International Human Rights Obligations », *Italian Y.B. Int'l L.*, vol. 15, 2005, pp. 181-183.

(125) C. pén. (France), article 421-2-5, modifié par la loi nº 2014-1353 du 13 novembre 2014 renforçant les dispositions relatives à la lutte contre le terrorisme, *J.O.R.F.*, nº 0263 du 14 novembre 2014, p. 19162, article 5.

(126) Résolution 1624 du Conseil de Sécurité, S/RES/1624 (2005), 14 septembre 2005, par. 4.

(127) C. pén. (France), article 421-2-5, modifié par la loi nº 2014-1353 du 13 novembre 2014 renforçant les dispositions relatives à la lutte contre le terrorisme, *J.O.R.F.*, nº 0263 du 14 novembre 2014, p. 19162, article 5.

(128) Cour eur. D.H., affaire *Leroy c. France*, *op. cit.* ; Nîmes, 20 septembre 2013, RG nº 13/00687 ; Corr. Paris, *M. le Procureur de la République, et autres et autres/D.M.*, jugement du 18 mars 2015 ; voy. égal. Fr. Dubuisson, « La lutte contre le terrorisme et la liberté d'expression : le cas de l'apologie du terrorisme », *op. cit.*, pp. 285-295.

font défaut pour l'apologie du terrorisme. Plusieurs rapports ont d'ailleurs souligné l'importance de respecter le « dénominateur commun » établi par la Convention de 2005 (129). L'Espagne, de son côté, a pu considérer qu'une telle interdiction était inconstitutionnelle (130). Par ailleurs, dans l'affaire *Karatas v. Turkey*, laquelle concernait la publication d'un poème qui pouvait être considéré comme glorifiant à certains égards des actes de violence, la Cour européenne des droits de l'homme a considéré qu'une interdiction générale de la glorification de la violence ne saurait être justifiée (131). « La notion d'apologie permet dès lors une interprétation beaucoup plus ouverte de ce qui constitue un discours incriminé, le critère étant celui du caractère "favorable" de celui-ci envers des actes terroristes » (132). De plus, cette interdiction très vague peut avoir un effet sur les débats relatifs à des questions d'intérêt public (133). Dès lors, « *glorification raises key concerns, not only in terms of the goals associated with the right to freedom of expression, but also with regard to the clarity and the définition of the law itself* » (134). Appliquée aux messages postés sur Facebook, une telle définition apparaît encore plus problématique au regard de la liberté d'expression, puisque tous les contenus touchant de près ou de loin au terrorisme pourraient se voir censurés. En effet, une publication qui, d'une manière ou d'une autre, émettrait des commentaires susceptibles d'être perçus comme positifs sur un acte terroriste, pourrait être associée à la glorification du terrorisme, voire à l'incitation au terrorisme (135).

(129) Rapport du Rapporteur spécial sur la promotion et la protection des droits de l'homme et des libertés fondamentales dans la lutte antiterroriste, Martin Scheinin, A/HRC/4/26/Add.3 14 décembre 2006, par. 26-27 ; Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, Frank La Rue, 16 mai 2011, A/HRC.17/27, par. 33-34.

(130) H. Fenwick, G. Phillipson, *Media Freedom Under the Human Rights Act*, Oxford, Oxford University Press, 2006, p. 533.

(131) Cour eur. D.H., arrêt du 8 juillet 1999, affaire *Karatas v. Turkey*, requête n° 23168/94, par. 50-52.

(132) Fr. Dubuisson, « La lutte contre le terrorisme et la liberté d'expression : le cas de l'apologie du terrorisme », *op. cit.*, p. 283.

(133) Article 19, *The impact of UK anti-terror laws on Freedom of Expression*, Londres, avril 2006, disponible à l'URL : https://www.article19.org/data/files/pdfs/analysis/terrorism-submission-to-icj-panel.pdf [consulté le 5 février 2019], p. 1.

(134) D. Murray, « Freedom of Expression, Counter-Terrorism and the Internet in Light of the UK Terrorist Act 2006 and the Jurisprudence of the European Court of Human Rights », *Neth. Q. Hum. Rts*, vol. 27, 2009, p. 344.

(135) S. M. Boyne, « Free Speech, Terrorism, and European Security: Defining and Defending the Political Community », *Pace L. Rev.*, vol. 30, 2010, p. 460.

III. — Les conséquences de la délégation de la censure des messages à contenu terroriste aux acteurs privés

Parallèlement aux problèmes relatifs à la définition des « contenus à caractère terroriste » et à leur traduction en langage informatique, on constate, notamment dans la proposition de règlement, une délégation de la fonction d'identifier et de supprimer les contenus à caractère terroriste aux opérateurs privés. Si cette délégation se justifie par le nombre exponentiel de contenus publiés chaque jour, l'absence de décision préalable à la suppression émanant d'une autorité étatique compétente nécessite une attention particulière. Dans la mesure où cette fonction est désormais majoritairement déléguée aux opérateurs privés, l'absence de contrôle juridictionnel préalable et des garanties qui y sont associées apparaît problématique (A). En particulier, ces opérateurs privés sont mus par des préoccupations principalement économiques, et le régime de responsabilité qui leur est applicable entraîne un risque de « sur-censure » des contenus publiés (B).

A. — *L'absence de contrôle juridictionnel préalable à la suppression des contenus*

La proposition de règlement, dans son article 6, exige des opérateurs qu'ils prennent des mesures proactives pour « (a) empêcher la remise en ligne de contenus qui ont été supprimés ou dont l'accès a été bloqué parce qu'ils sont considérés comme revêtant un caractère terroriste ; (b) [...] détecter, [...] identifier et [...] supprimer sans délai les contenus à caractère terroriste, ou [...] bloquer l'accès à ceux-ci » (136). Ce faisant, les États européens demandent à ces entreprises technologiques, en tant qu'opérateurs privés, de procéder à une censure *a priori* des contenus à caractère terroriste. C'est ce que fait désormais Facebook, puisque l'entreprise américaine soutient qu'actuellement, 99 % des contenus à caractère terroriste relatifs à l'État islamique et à Al-Qaïda sont retirés du site avant que la communauté des utilisateurs n'y ait eu accès (137). Or, aucun contrôle extérieur n'est exercé sur ce type de censure. Dans sa version actuelle, la proposition de règlement établit uniquement que les fournisseurs « prévoient des garanties efficaces et adéquates pour assurer l'exactitude et le bien-fondé des décisions prises au sujet de ces contenus » et requiert qu'il soit procédé à « des vérifications humaines *lorsque cela se justifie*

(136) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, C(2018)640 final, Bruxelles, 12 septembre 2018, article 6.

(137) « Hard Questions: Are We Winning the War On Terrorism Online? », by Monika Bickert, Head of Global Policy Management, and Brian Fishman, Head of Counterterrorism Policy, Facebook, 28 novembre 2017, *supra* note 2.

(nous soulignons)»(138). À ce sujet, dans sa Recommandation sur le guide des droits de l'homme pour les utilisateurs de l'Internet, le Comité des Ministres du Conseil de l'Europe a précisé que «des mesures générales de blocage ou de filtrage ne devraient être prises *par les pouvoirs publics* [...] que sur la base d'une décision au sujet de l'illégalité de ce contenu prise par une autorité nationale compétente». De plus, le Comité des ministres précise également que «les pouvoirs publics devraient veiller à ce que tous les filtres soient évalués avant et pendant leur mise en œuvre, afin de veiller à ce que tous les effets du filtrage soient en adéquation avec l'objectif de la restriction et donc justifiés dans une société démocratique, afin d'éviter tout blocage injustifié des contenus»(139).

Si la proposition de règlement relatif aux contenus à caractère terroriste précise les entités habilitées à supprimer ces contenus, ou à en ordonner la suppression(140), il n'exige *a priori* à aucun stade l'intervention d'une autorité judiciaire indépendante et impartiale. En effet, la suppression aura lieu à l'initiative soit des autorités compétentes d'un État membre, soit des prestataires de services eux-mêmes.

Dans le premier cas, bien que la censure des messages soit *in fine* exécutée par le prestataire de services, l'injonction de suppression pourrait, suivant la proposition de règlement, être décidée par les autorités compétentes de l'État membre concerné(141). Il convient néanmoins de noter que la proposition n'émet aucun critère relatif à la nature de ces autorités(142), ni au degré d'institutionnalisation de ces dernières(143). L'exposé des motifs de la proposition de règlement mentionne seulement que l'injonction pourra être émise «en tant que décision administrative ou judiciaire par une autorité compétente d'un État membre»(144). Le rapporteur spécial pour la liberté d'expression a pourtant insisté sur l'importance, pour les États, de ne «limiter la publication de contenus qu'en vertu d'une ordonnance délivrée par un organe judiciaire indépendant et impartial, dans le respect des garanties d'une procédure régulière et des normes de légalité, de nécessité et de

(138) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, *op. cit.*, article 9.

(139) Recommandation CM/Rec(2014)6 du Comité des Ministres aux États membres sur un guide des droits de l'homme pour les utilisateurs d'internet, 16 avril 2014, par. 49, disponible à l'URL: https://rm.coe.int/16804b8447 [consulté le 16 mars 2019].

(140) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, *op. cit.*, articles 4, 5 et 6.

(141) *Ibid.*, article 4.

(142) *Ibid.*, article 17.

(143) Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, *op. cit.*, p. 6.

(144) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, exposé des motifs, p. 4.

légitimité » (145). Si la Cour européenne n'exige pas nécessairement l'intervention d'une autorité judiciaire, elle estime néanmoins souhaitable, dans le cas des mesures de surveillance secrètes, que le contrôle soit confié à un juge, car le contrôle juridictionnel offre les meilleures garanties d'indépendance, d'impartialité et de procédure régulière (146). Dans l'éventualité où le contrôle n'est pas supervisé par une autorité judiciaire, la Cour a encore pu développer d'autres critères relatifs à l'indépendance des autorités chargées de mettre en œuvre le contrôle (147). Dans son état actuel, le règlement exige seulement de ces autorités qu'elles communiquent les motifs de leur décision (article 4, § 3, b), et qu'elles « disposent de la capacité nécessaire et de ressources suffisantes pour atteindre les objectifs et remplir les obligations qui leur incombent en vertu du présent règlement » (article 12). Compte tenu du peu d'obligations relatives à la formation et au processus de décision de ces autorités, on ne peut affirmer qu'elles offriront dans chaque État des garanties suffisantes contre les abus.

Dans le second cas, la suppression peut être directement décidée par les prestataires de services (148). À la différence des mesures prises par l'État, qui devraient en principe être assorties des garanties effectives et adéquates contre les abus, les mesures prises par des opérateurs privés tels que Facebook manquent de transparence ainsi que de référence à des critères connus du public (149). Ainsi, par exemple, si nous avons pu identifier des éléments communs de définition de l'infraction terroriste, rien n'indique que les critères établis par Facebook soient exclusivement basés sur cette définition. En effet, Facebook définit un acte terroriste comme « tout acte de violence prémédité contre des individus ou une propriété, perpétré par une personne n'appartenant pas à un gouvernement en vue d'intimider une population civile, un gouvernement ou un organisme international, avec pour objectif d'atteindre un but politique, religieux ou idéologique » (150). Le caractère excessivement vague de cette définition a d'ailleurs été critiqué par la rapporteuse spéciale aux droits de l'homme dans la lutte antiterroriste, car elle crée un risque de "sur-censure et d'arbitraire (151). En effet, cette définition ne fait aucune

(145) Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/HRC/38/35, 6 avril 2018, par. 66.

(146) Cour eur. D.H., arrêt du 4 décembre 2015, affaire *Roman Zakharov c. Russie*, requête nº 14881/03, par. 233 ; Cour eur. D.H., *Big Brother Watch et autres c. Royaume-Uni*, précité, par. 309.

(147) *Ibid.*, par. 322-386.

(148) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, C(2018)640 final, Bruxelles, 12 septembre 2018, article 6.

(149) J. M. Balkin, « Old-School/New-School Speech Regulation », *Harvard L. R.*, vol. 127, 2014, p. 2341.

(150) Facebook, Standards de la Communauté, disponible à l'URL : https://fr-fr.facebook.com/communitystandards/dangerous_individuals_organizations [consulté le 16 avril 2019].

(151) « UN human rights expert says Facebook's 'terrorism' définition is too broad », 3 septembre 2018, disponible à l'URL : https://www.un.org/victimsofterrorism/fr/le-rapporteur-sp%C3%A9cial-sur-la-promotion-et-la-protection-des-droits-de-l%E2%80%99homme-et-des-libert%C3%A9s

référence au contexte dans lequel l'acte de violence doit prendre place. Par ailleurs, aucune information n'est donnée sur ce que constitue « tout acte de violence ». La suppression par Facebook de certains contenus entrant dans cette définition irait dès lors bien au-delà des limitations autorisées. Le rapporteur spécial pour la liberté d'expression et d'opinion a sur ce point souligné que « les questions complexes de faits et de droit devraient en général être tranchées par des institutions publiques et non par des acteurs privés dont les procédures peuvent être incompatibles avec les règles d'une procédure régulière et dont les motifs sont essentiellement d'ordre économique » (152).

Si la proposition de règlement relative aux contenus à caractère terroriste impose plusieurs obligations aux fournisseurs de services en matière de transparence (153), il n'en reste pas moins que la décision de suppression est dans certains cas entièrement déléguée à un opérateur privé (154). Le Conseil de l'Europe a pu souligner que la délégation à des opérateurs privés, par les États, des mesures à prendre, permettait à ces opérateurs de mettre en œuvre des solutions que la loi n'autoriserait pas les États à ordonner eux-mêmes (155). De cette manière, cette délégation aux opérateurs privés permettrait aux États d'outrepasser leurs obligations en matière de liberté d'expression.

Enfin, si des mécanismes de plainte sont également prévus par la proposition de règlement, la même remarque générale s'impose une fois encore. La proposition de règlement exige seulement des fournisseurs de service qu'ils « établissent des mécanismes accessibles et efficaces » (156) et qu'ils « examinent dans les meilleurs délais toute réclamation qu'ils reçoivent et rétablissent sans tarder les contenus en cause dès lors qu'il était injustifié de les supprimer ou d'en bloquer l'accès » (157). Tel que cela a été souligné par plusieurs rapporteurs aux droits de l'homme, ces mécanismes n'offrent toutefois

[consulté le 29 décembre 2018]; voy. également Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/HCR/38/35, 6 avril 2018, par. 26.

(152) Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, Daivd Kaye, A/HCR/38/35, 6 avril 2018, par. 17.

(153) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, C(2018)640 final, Bruxelles, 12 septembre 2018, article 8.

(154) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, C(2018)640 final, Bruxelles, 12 septembre 2018, article 6.

(155) Conseil de l'Europe, « Algorithmes et droits humains… », *op. cit.*, p. 22.

(156) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, *op. cit.*, article 10, § 1.

(157) *Ibid.*, article 10, § 2.

par les mêmes garanties qu'une procédure devant les autorités judiciaires, et ne permettent pas de réparation pour le préjudice subi (158).

B. — *Le risque de « sur-censure » comme conséquence du régime de responsabilité accru des intermédiaires de l'Internet pour la publication des messages à contenu terroriste*

Le flou qui entoure le régime de responsabilité des intermédiaires de l'Internet, ou, dans certains cas, la responsabilité accrue de ces derniers, entraîne également certaines conséquences par rapport à la proportionnalité de l'ingérence. Il n'existe actuellement aucun régime uniforme international de responsabilité des intermédiaires (fournisseurs d'accès à l'Internet, hébergeurs, plateformes de médias sociaux et moteurs de recherche) pour les contenus illicites qui seraient publiés par des tiers sans leur intervention (159). Si certains États tels que la Chine et le Japon ont opté pour un modèle de responsabilité stricte (160) qui impose aux intermédiaires de surveiller les contenus afin de vérifier leur conformité avec la loi, d'autres tels que les États-Unis offrent une immunité totale aux intermédiaires (161). L'UE accorde par principe, dans la directive relative au commerce électronique, une large immunité aux intermédiaires qui fournissent simplement un accès technique à l'Internet ou aux hébergeurs de contenus, à condition qu'ils se conforment à certaines exigences (162). En effet, les intermédiaires ne sont soumis à aucune obligation de surveillance des contenus stockés ou transmis et disposent d'une immunité de principe, mais peuvent perdre cette dernière s'ils n'agissent pas promptement pour retirer ou bloquer l'accès à ces contenus illégaux dès qu'ils prennent connaissance de ces derniers (système de *notice and take down*) (163).

(158) Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, the Special Rapporteur on the right to privacy and the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism, *op. cit.*, p. 8.

(159) Joint Declaration on Freedom of Expression and the Internet, 1st June 2011, The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, 2.a.

(160) Article 19, « Intermédiaires Internet : Dilemme de la responsabilité », 2013, p. 8, disponible à l'URL : https://www.article19.org/data/files/WEB_French.pdf [consulté le 2 août 2019].

(161) Voy. Communication Decency Act, 47 U.S.C. section 230.

(162) Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (« directive sur le commerce électronique »), *J.O.U.E.*, L 178/1, 17 juillet 2000, articles 12 à 15.

(163) *Ibid.*, article 14 ; voy. égal. Fr. Dubuisson, « Les restrictions à l'accès au contenu d'Internet et le droit à la liberté d'expression », *Internet et le droit international*, Colloque SFDI de Rouen, Pedone, 2014, pp. 108-109.

Toutefois, puisqu'une connaissance des contenus illicites est requise dans le chef des intermédiaires pour engager leur responsabilité, ce régime n'impose aucune obligation de surveillance des contenus *a priori*. Les intermédiaires restent libres, à leur discrétion, de surveiller les contenus dont ils assurent la transmission.

Depuis 2015, toutefois, une responsabilité accrue semble peser sur certains de ces intermédiaires, suite à l'intervention de la Cour européenne des droits de l'homme. Dans l'affaire *Delfi AS c. Estonie* (164), la société Delfi, qui gérait un portail d'actualités sur lequel les internautes pouvaient laisser des commentaires, fut tenue responsable en Estonie de la publication de commentaires injurieux et illégaux laissés par des internautes anonymes. La Cour de Strasbourg, lors de l'examen de la proportionnalité de cette ingérence dans le droit à la liberté d'expression, s'est attardée sur le rôle joué par la société Delfi. Premièrement, elle a considéré que la société avait un intérêt particulier, économique, à la publication des commentaires. Dès lors, la société jouait un rôle suffisamment actif (165). Deuxièmement, puisqu'il est souvent difficile de retrouver les auteurs des commentaires anonymes, il n'était pas disproportionné de faire peser une telle responsabilité sur la société. Troisièmement, les mécanismes de surveillance mis en place par la société, à savoir un système de notice ainsi qu'un contrôle automatique des expressions employées, n'étaient pas suffisants selon la Cour (166). Au vu du raisonnement tenu par la Cour dans cette affaire, la notion d'intermédiaire passif semble se restreindre. Le rôle joué par ces derniers se veut également de plus en plus actif dans la détection des contenus illégaux, au risque de voir leur responsabilité engagée. Comme elle le rappellera dans une affaire ultérieure, les portails d'information sur l'Internet ont des devoirs et des responsabilités en vertu de l'article 10, § 2, pour les commentaires générés par les utilisateurs lorsque ces commentaires sont clairement des expressions illégales, qui équivalent à un discours de haine ou une incitation à la violence (167). La Cour a toutefois pris le soin de distinguer la société en cause dans cette affaire d'autres types d'intermédiaires :

> « La présente affaire ne concerne pas d'autres types de forums sur Internet susceptibles de publier des commentaires provenant d'internautes, par exemple les forums de discussion ou les sites de diffusion électronique, où les internautes peuvent exposer librement leurs idées sur n'importe quel sujet sans que la discussion ne soit canalisée par des interventions du responsable du forum, ou encore les plateformes de médias sociaux où le fournisseur de la plateforme ne produit aucun contenu et où le fournisseur de contenu peut être un particulier administrant un site ou un blog dans le cadre de ses loisirs » (168).

(164) Cour eur. D.H., arrêt du 16 juin 2015, affaire *Delfi AS c. Royaume-Uni*, requête n° 64569/09.
(165) *Ibid.*, par. 146.
(166) *Ibid.*, par. 159.
(167) Cour eur. D.H., arrêt du 2 février 2016, affaire *Magyar Tartalomszolgáltatók Egyesülete et Index.hu Zrt c. Hongrie*, requête n° 22947/13, par. 63.
(168) *Ibid.*, par. 116.

Si l'affaire *Delfi* semble exclure les plateformes de médias sociaux, parmi lesquelles se retrouve Facebook, la Commission européenne n'a pas tardé à inviter de telles plateformes à une collaboration accrue. Dans une recommandation de mars 2018 (169), la Commission avait expressément invité les prestataires de services d'hébergement à adopter une approche préventive et proactive (170). La directive relative à la lutte contre le terrorisme établit dans ce domaine une véritable obligation de lutter contre les contenus en ligne de provocation publique au terrorisme (171). Enfin, la proposition de règlement relatif aux contenus terroristes en ligne prévoit également que les fournisseurs de services d'hébergement prennent des mesures proactives pour protéger leurs services contre la diffusion de contenus à caractère terroriste (172).

Cette recommandation ne disposant d'aucune force juridique contraignante, et la proposition de règlement relatif aux contenus terroristes n'étant encore qu'une proposition susceptible de modifications, il apparaît à première vue que les médias sociaux tels que Facebook continuent à bénéficier pleinement du régime d'immunité. Toutefois, en pratique, la pression exercée sur ces acteurs privés amène ces derniers à censurer tout contenu qui pourrait s'avérer illégal. À nouveau, on peut douter de l'impartialité de ces acteurs dans la censure de certains contenus (173). Ces acteurs, comme le montrent plusieurs exemples tirés de l'actualité (174), ont effectivement tendance à privilégier leurs intérêts à la liberté d'expression, et à censurer plutôt trop que trop peu.

Notons toutefois qu'au sein de l'UE, l'Allemagne a récemment adopté un régime strict de responsabilité pour les contenus publiés par l'intermédiaire des plateformes, et dont l'application peut aboutir à des amendes allant jusqu'à 5 millions d'euros lorsque les plateformes échouent à supprimer les contenus illégaux (175). Le rapporteur spécial pour la liberté d'expression et

(169) Commission européenne, Recommandation de la Commission européenne du 1er mars 2018 sur les mesures destinées à lutter de manière efficace contre les contenus illicites en ligne, C(2018)1177 final, Bruxelles, 1er mars 2018.

(170) *Ibid.*, par. 18 et 36-37.

(171) Directive (UE) 2017/541 précitée, article 21.

(172) Commission européenne, Proposition de règlement du Parlement européen et du Conseil relatif à la prévention de la diffusion de contenus à caractère terroriste en ligne, *op. cit.*, article 6.

(173) Fr. DUBUISSON, « Les restrictions à l'accès au contenu d'Internet et le droit à la liberté d'expression », *op. cit.*, p. 109, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank LaRue, 16 May 2011, A/HRC/17/27, par. 42-43.

(174) Voy. *Huffigton Post*, « Le musée du Jeu de Paume censuré par Facebook pour une photo de nu », 7 mars 2013, disponible à l'URL : https://www.huffingtonpost.fr/2013/03/07/jeu-de-paume-facebook-censure-photo-nu_n_2826838.html [consulté le 2 mars 2019]; *Courrier International*, « Monde arabe : quand Facebook censure les femmes "dévoilées" », 13 novembre 2012, disponible à l'URL : https://www.courrierinternational.com/article/2012/11/09/monde-arabe-quand-facebook-censure-les-femmes-devoilees [consulté le 2 mars 2019].

(175) Network Enforcement Act (Netzdurchsetzunggesetz, NetzDG), 1st September 2017, *Federal Law Gazette I*, pp. 3352 et s.

d'opinion a fait part à l'Allemagne de ses nombreuses craintes relatives à un tel régime accru de responsabilité, et en particulier :

> « *censorship measures should not be delegated to private entities* (A/HRC/17/31). *States should not require the private sector to take steps that unnecessarily or disproportionately interfere with freedom of expression, whether through laws, policies or extralegal means* (A/HRC/32/38). *(…) The provisions imposing high fines for non-compliance with the obligations set out in the bill raise concerns, as these obligations as mentioned above may represent undue interference with the right to freedom of expression and privacy. The high fines raise proportionality concerns, and may prompt social networks to remove content that may be lawful. (…) Further, I am concerned with the lack of judicial oversight with respect to the responsibility placed upon private social networks to remove and delete content. Any legislation restricting the right to freedom of expression and the right to privacy* must be applied by a body which is independent of any political, commercial, or unwarranted influences in a manner that is neither arbitrary nor discriminatory (A/HRC/17/27) » (176) (nous soulignons).

La France semble vouloir se diriger dans la même direction, puisqu'une proposition de loi visant à lutter contre la haine sur l'Internet a été adoptée par l'Assemblée nationale le 9 juillet 2019 (177). Tout comme son homologue allemande, elle entend tenir responsables les plateformes de réseaux sociaux pour les contenus illicites qui ne seraient pas retirés dans un délai de 24 heures (178).

Compte tenu de ces différents éléments, il convient de conclure que, en l'état actuel du droit international des droits de l'homme, une censure *a priori* et une suppression rapide *a posteriori* des contenus à caractère terroriste sur les réseaux sociaux par des acteurs privés n'apparaît pas assortie des garanties suffisantes. De ce fait, le régime de suppression des contenus par les acteurs privés offert dans la proposition de règlement peut difficilement être considéré comme proportionné à l'objectif poursuivit, et donc nécessaire dans une société démocratique.

Conclusion

La présente contribution met en évidence le fait que si la censure *a priori* et la suppression rapide *a posteriori* des contenus à caractère terroriste, telles que prévues dans la déclaration commune de plusieurs États, dans la proposition de règlement ou encore dans la loi allemande, et auxquelles procède l'entreprise américaine Facebook (179), poursuivent plusieurs buts légitimes,

(176) Rapporteur spécial pour la promotion et la protection du droit à la liberté d'opinion et d'expression, Communication n° OL DEU 1/2017, 1er juin 2017.

(177) Proposition de loi du 20 mars 2019 visant à lutter contre la haine sur internet, n° 1785, adoptée par l'Assemblée nationale le 9 juillet 2019, T.A. n° 310.

(178) *Ibid.*, article 1.

(179) « Hard Questions: Are We Winning the War On Terrorism Online? », *op. cit.*

ces restrictions posent des problèmes considérables en ce qui concerne les critères de légalité et de nécessité. Dès lors, celles-ci ne sauraient être considérées comme des ingérences justifiées au droit à la liberté d'expression au sens des articles 19, § 3, du PIDCP et 10, § 2, de la CEDH.

À côté de ce constat, il peut sembler paradoxal que ces mêmes plateformes ne soient soumises à aucune obligation internationale en ce qui concerne la protection de la liberté d'expression et les droits de l'homme. Comme le souligne le rapporteur spécial sur la protection et la promotion du droit à la liberté d'opinion et d'expression, « les entreprises restent des régulateurs énigmatiques, qui créent une sorte de 'droit des plateformes' qui manque de clarté et de cohérence et dans lequel les mécanismes de responsabilisation et les voies de recours sont flous » (180). Si les principes directeurs relatifs aux entreprises et aux droits de l'homme établissent une responsabilité à charge des entreprises d'assurer le respect des droits de l'homme (181) — en particulier l'obligation de faire preuve de diligence raisonnable — aucun régime contraignant ne leur est directement applicable. Les standards de la communauté Facebook, en particulier, ne doivent pas respecter les exigences strictes relatives au critère de légalité des limitations.

Ce dernier constat nous amène à interroger la pertinence du modèle actuel des droits de l'homme et en particulier du régime des limitations dans un contexte où apparaissent ces nouveaux acteurs et outils permettant tant la réalisation de ces droits que leur violation. L'écart entre le régime applicable aux États et celui applicable à ces opérateurs privés pose nécessairement question et mériterait qu'on s'y attarde.

En effet, à l'heure actuelle, il est impossible de nier le fait que l'Internet constitue un outil particulièrement utile pour les organisations terroristes. De ce fait, les États et les entreprises du secteur informatique ont raison de vouloir agir afin d'empêcher la mauvaise utilisation notamment des réseaux sociaux par les terroristes. À ce sujet, l'Office des Nations unies contre la drogue et le crime a lui-même reconnu le fait que « si la responsabilité de la lutte contre l'utilisation de l'Internet à des fins terroristes incombe *in fine* aux États membres, la coopération des principaux acteurs du secteur privé est essentielle en termes d'efficacité » (182). Néanmoins, la lutte contre le terrorisme ne confère pas aux États et aux géants du Web une « carte blanche » en la matière (183), et ne doit pas être mise en œuvre à n'importe quel prix.

(180) Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, A/HCR/38/35, 6 avril 2018, par. 1.

(181) OHCHR, Principes directeurs relatifs aux entreprises et aux droits de l'homme, 2011, principes n^{os} 11 à 24.

(182) UNODC, « Utilisation de l'Internet à des fins terroristes », *op. cit.*, p. 154, par. 478.

(183) H. Keller, M. Sigron, « State Security voy. Freedom of Expression: Legitimate Fights against Terrorism or Suppression of Political Opposition? », *Human Rights Law Review*, 2010, p. 167.

En conclusion, cette contribution a été réalisée avec la volonté de mettre en évidence le fait que si ces initiatives apparaissent légitimes du point de vue de l'État, la censure *a priori* et la suppression rapide *a posteriori* des contenus à caractère terroriste, par le biais d'outils informatisés recourant le plus souvent à l'intelligence artificielle, semblent être mises en œuvre de manière prématurée. Il aurait convenu de mettre en place un système plus précis et d'adopter les normes légales appropriées avant de mettre ces mesures en œuvre, de manière à ne pas limiter injustement et de manière disproportionnée le droit fondamental à la liberté d'expression des individus.