Nations Unies A/HRC/38/35

Assemblée générale

Distr. générale
6 avril 2018
Français
Original : anglais

Conseil des droits de l'homme
Trente-huitième session
18 juin-6 juillet 2018
Point 3 de l'ordre du jour
Promotion et protection de tous les droits de l'homme, civils, politiques, économiques, sociaux et culturels, y compris le droit au développement

Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression

Note du secrétariat

Le Secrétariat a l'honneur de transmettre au Conseil des droits de l'homme le rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, conformément à la résolution 34/18 du Conseil. Dans son rapport, le Rapporteur spécial traite de la réglementation des contenus en ligne générés par les utilisateurs. Il recommande aux États de créer un environnement propice à la liberté d'expression en ligne, et aux entreprises d'appliquer les normes relatives aux droits de l'homme à toutes les étapes de leurs activités. Le droit des droits de l'homme donne aux entreprises les moyens de définir leurs positions de manière à respecter les normes démocratiques et à se soustraire aux demandes autoritaires. Les entreprises et les États devraient au minimum viser une amélioration radicale de la transparence, dès le stade de l'élaboration des règles et jusqu'à leur application, afin de garantir l'autonomie de l'utilisateur, étant donné que les individus exercent de plus en plus leurs droits fondamentaux en ligne.

GE.18-05436 (F) 230418 270418
1805436

Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression

Table des matières

I. Introduction

1. Aux commencements de l'ère numérique, John Perry Barlow déclara qu'Internet allait ouvrir la voie à un monde « où n'importe qui, n'importe où, peut exprimer ses croyances, aussi singulières qu'elles soient, sans peur d'être réduit au silence ou à la conformité »[1]. Bien qu'Internet reste le plus formidable outil jamais conçu pour ce qui est de l'accès mondial à l'information, il est difficile d'y trouver une traduction concrète de ce bel idéal. Le public voit dans les contenus créés par les utilisateurs de la haine, des abus et de la désinformation. Les gouvernements y voient le recrutement de terroristes et des manifestations décevantes de dissidence et d'opposition. Les organisations de la société civile y observent l'externalisation de fonctions publiques telles que la protection de la liberté d'expression, déléguées à des acteurs privés qui n'ont pas d'obligation de rendre des comptes. Bien qu'elles prennent des mesures pour préciser leurs réglementations et leurs relations avec les gouvernements, 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. L'ONU, les organisations régionales et les organes conventionnels ont rappelé que les droits dont les personnes jouissent hors ligne doivent également être protégés en ligne, mais on n'est pas toujours sûr que les entreprises protègent les droits de leurs utilisateurs ou que les États fournissent à celles-ci les incitations juridiques nécessaires à cet effet.

2. Dans le présent rapport, le Rapporteur spécial propose un cadre pour la modération des contenus en ligne générés par les utilisateurs qui donne aux droits de l'homme une place centrale[2]. Il s'emploie à répondre à certaines questions fondamentales : quelles responsabilités les entreprises doivent-elles exercer pour garantir que leurs plateformes ne portent pas atteinte aux droits garantis par le droit international ? Quelles normes devraient-elles appliquer en matière de modération de contenus ? Les États devraient-ils réglementer la modération des contenus commerciaux et, dans l'affirmative, de quelle manière ? La loi exige des États qu'ils fassent preuve de transparence et qu'ils rendent des comptes afin d'atténuer les menaces qui pèsent sur la liberté d'expression : faut-il attendre la même chose des acteurs privés ? Quelle forme revêtent les procédures de protection et les mécanismes de recours à l'ère numérique ?

3. Certaines de ces questions ont été traitées dans les précédents rapports[3]. Le présent rapport s'intéresse essentiellement à la réglementation des contenus générés par les utilisateurs, en se concentrant principalement sur les États et les entreprises de médias sociaux, mais en tenant compte de tous les acteurs du secteur des technologies de l'information et des communications (TIC) concernés. Le Rapporteur spécial expose le cadre juridique des droits de l'homme applicable et présente les approches des entreprises et des États en matière de réglementation des contenus. Il propose des normes et des procédures que les entreprises devraient adopter pour réglementer les contenus dans le respect du droit des droits de l'homme.

4. Le rapport s'appuie avant tout sur l'étude des conditions d'utilisation établies par les entreprises et des rapports présentés à des fins de transparence et sur des sources secondaires. Suite à un appel à commentaires, 21 États et 29 acteurs non étatiques, dont une entreprise, ont soumis des communications. Le Rapporteur spécial s'est rendu dans plusieurs entreprises de la Silicon Valley et s'est entretenu avec d'autres entreprises pour comprendre les approches qu'elles utilisaient en matière de modération de contenus[4]. La

[1] John Perry Barlow, Déclaration d'indépendance du cyberespace, 8 février 1996.

[2] Le terme « modération » désigne la procédure au moyen de laquelle les entreprises Internet déterminent si les contenus générés par les utilisateurs répondent aux normes définies dans leurs conditions d'utilisation et autres règles.

[3] A/HRC/35/22 et A/HRC/32/38.

[4] Le Rapporteur spécial s'est rendu au siège de Facebook, de GitHub, de Google, de Reddit et de Twitter et s'est entretenu avec des représentants de Yahoo/Oath, de Line et de Microsoft. Il s'est également rendu à la Wikimedia Foundation, organisation à but non lucratif. Il espère aller voir des entreprises à Beijing, Moscou, Séoul et Tokyo dans le cadre des activités menées en lien avec le présent rapport.

consultation d'organisations de la société civile à Bangkok et à Genève en 2017 et 2018 et des entretiens en ligne en 2018 avec des experts d'Amérique latine, du Moyen-Orient, d'Afrique du Nord et d'Afrique subsaharienne lui ont été très utiles[5].

II. Cadre juridique

5. Les activités des entreprises du secteur des TIC mettent en jeu le droit au respect de la vie privée, le droit à la liberté de religion, de conviction, d'opinion et d'expression, de réunion et d'association, et le droit de participer à la vie publique, entre autres. Le présent rapport met l'accent sur la liberté d'expression, tout en reconnaissant l'interdépendance des droits, le respect de la vie privée ouvrant notamment la voie à la liberté d'expression[6]. L'article 19 du Pacte international relatif aux droits civils et politiques prévoit des règles établies au niveau mondial et ratifiées par 170 États, qui font écho à la Déclaration universelle des droits de l'homme et garantissent « le droit de ne pas être inquiété pour ses opinions » et celui « de rechercher, de recevoir et de répandre des informations et des idées de toute espèce, sans considération de frontières », par quelque moyen que ce soit[7].

A. Obligations de l'État

6. Le droit des droits de l'homme soumet les États à l'obligation de mettre en place un environnement favorable à la liberté d'expression et de protéger son exercice. En vertu du devoir qui leur incombe de garantir la liberté d'expression, les États sont notamment tenus de promouvoir la diversité et l'indépendance des médias et l'accès à l'information[8]. Des organismes internationaux et régionaux ont par ailleurs instamment invité les États à promouvoir l'accès universel à Internet[9]. Les États doivent également veiller à ce que les entités privées ne compromettent pas l'exercice de la liberté d'opinion et d'expression[10]. Dans les Principes directeurs relatifs aux entreprises et aux droits de l'homme, adoptés par le Conseil des droits de l'homme en 2011, il est souligné que les États ont l'obligation de créer un environnement qui permette aux entreprises de respecter les droits de l'homme (principe 3)[11].

7. Les États ne peuvent pas restreindre le droit de ne pas être inquiété pour ses opinions. Aux termes du paragraphe 3 de l'article 19 du Pacte, les restrictions imposées par l'État à la liberté d'expression doivent satisfaire aux conditions bien établies énumérées ci-après :

- *Légalité*. Les restrictions doivent être « fixées par la loi ». Elles doivent notamment être adoptées dans le cadre de procédures juridiques régulières et limiter le pouvoir discrétionnaire du gouvernement d'une manière qui établisse avec « suffisamment de précision » la distinction entre l'expression licite et l'expression illicite d'opinions. Les restrictions adoptées secrètement ne satisfont pas à cet impératif

[5] Le Rapporteur spécial tient à remercier son conseiller juridique, Amos Toh, ainsi que les étudiants de la International Human Rights Clinic de l'University of California, Irvine School of Law.

[6] Voir A/HRC/29/32, par. 16 à 18.

[7] Voir également : Charte africaine des droits de l'homme et des peuples, art. 9 ; Convention américaine relative aux droits de l'homme, art. 13 ; Convention de sauvegarde des Droits de l'Homme et des Libertés fondamentales, art. 10. Voir également la communication du Centro de Estudios en Libertad de Expresión y Acceso a la Información.

[8] Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande, 3 mars 2017, art. 3. Voir également l'observation générale nº 34 (2011) du Comité des droits de l'homme sur la liberté d'opinion et la liberté d'expression, par. 18 et 40 ; A/HRC29/32, par. 61, et A/HRC/32/38, par. 86.

[9] Voir la résolution 32/13 du Conseil des droits de l'homme, par. 12 ; Bureau du Rapporteur spécial pour la liberté d'expression de la Commission interaméricaine des droits de l'homme, *Standards for a Free, Open and Inclusive Internet* (2016), par. 18.

[10] Voir l'observation générale nº 34, par. 7.

[11] A/HRC/17/31.

fondamental[12]. D'une manière générale, la légalité devrait être garantie par un contrôle exercé par des autorités judiciaires indépendantes[13].

- *Nécessité et proportionnalité*. Les États doivent démontrer que la restriction compromet le moins possible l'exercice du droit et protège effectivement l'intérêt public légitime qui est en jeu ou a toutes les chances de le protéger. Les États qui adoptent une législation restrictive et décident de restreindre telle ou telle forme d'expression ne peuvent pas se contenter d'affirmer le caractère nécessaire de la restriction, mais sont tenus de le démontrer[14].

- *Légitimité*. Toute restriction, pour être légale, doit protéger exclusivement les intérêts énumérés au paragraphe 3 de l'article 19 : les droits ou la réputation d'autrui, la sécurité nationale ou l'ordre public, la santé ou la moralité publiques. Les restrictions visant à protéger les droits d'autrui, par exemple, visent « les droits fondamentaux tels qu'ils sont reconnus dans le Pacte et, plus généralement, dans le droit international des droits de l'homme »[15]. Les restrictions visant à protéger le droit au respect de la vie privée, le droit à la vie, le droit à une procédure régulière, le droit d'association et le droit de participer aux affaires publiques, pour n'en citer que quelques-uns, seront légitimes s'il est démontré qu'elles satisfont aux critères de légalité et de nécessité. Le Comité des droits de l'homme attire l'attention sur le fait que les restrictions visant à protéger la « moralité publique » doivent « être fondées sur des principes qui ne procèdent pas d'une tradition unique » et tenir compte des principes de non-discrimination et d'universalité des droits[16].

8. Les restrictions imposées en application du paragraphe 2 de l'article 20 du Pacte, qui exige des États qu'ils interdisent « tout appel à la haine nationale, raciale ou religieuse qui constitue une incitation à la discrimination, à l'hostilité ou à la violence » doivent répondre aux conditions cumulatives de légalité, de nécessité et de légitimité[17].

B. Responsabilité des entreprises

9. Les entreprises Internet sont devenues des plateformes essentielles pour les discussions et les débats, l'accès à l'information, le commerce et le développement humain[18]. Elles collectent et conservent les données personnelles de milliards d'individus, notamment des informations sur leurs habitudes, le lieu où ils se trouvent et leurs activités, et affirment fréquemment jouer un rôle civique. En 2004, Google a proclamé son ambition de réaliser de bonnes choses pour le monde, même au prix d'une renonciation à certains gains à court terme[19]. Le fondateur de Facebook a déclaré désirer développer une infrastructure sociale qui donnerait aux populations les moyens de fonder une communauté mondiale qui fonctionne pour tous[20]. Twitter a promis des politiques qui favorisent − au lieu d'empêcher − une conversation libre à l'échelle de la planète[21]. VKontakte, une entreprise russe de médias sociaux, se proposer d'unir les populations du monde entier, tandis que Tencent, reprenant les termes du Gouvernement chinois, déclare vouloir contribuer à l'édification d'une société harmonieuse et devenir une entreprise citoyenne exemplaire[22].

[12] Ibid., par. 25 ; A/HRC/29/32.
[13] Ibid.
[14] Voir l'observation générale nº 34, par. 27.
[15] Ibid., par. 28.
[16] Ibid., par. 32.
[17] Ibid., par. 50. Voir également A/67/357.
[18] Voir, par exemple, Cour suprême des États-Unis, *Packingham* v. *North Carolina*, avis du 19 juin 2017 ; Cour européenne des droits de l'homme, *Times Newspapers Ltd* c. *Royaume-Uni* (nos 1 et 2) (requêtes nos 3002/03 et 23676/03), arrêt du 10 mars 2009, par. 27.
[19] Déclaration d'enregistrement de titres (formulaire S-1) en vertu de la loi sur les titres de 1933, 18 août 2004.
[20] Mark Zuckerberg, « Building global community », Facebook, 16 février 2017.
[21] Twitter, Formulaire S-1 de déclaration d'enregistrement, 13 octobre 2013, par. 91 et 92.
[22] VKontakte, information communiquée par l'entreprise ; Tencent, « About Tencent ».

10. Rares sont les entreprises qui appliquent les principes des droits de l'homme dans leurs activités ; la plupart de celles qui le font ne sont guidées que par la nécessité de s'adapter aux menaces et aux exigences des gouvernements dans ce domaine [23]. Les Principes directeurs relatifs aux entreprises et aux droits de l'homme établissent toutefois des « norme[s] de conduite générale » que toutes les entreprises sont censées appliquer dans toutes leurs activités, où qu'elles opèrent [24]. Les Principes directeurs ne sont pas contraignants, mais le rôle prépondérant que jouent les entreprises dans la vie publique partout dans le monde plaide fortement en faveur de l'adoption et de la mise en œuvre de ces principes.

11. Les Principes directeurs établissent un cadre en vertu duquel les entreprises devraient, au minimum :

a) Éviter d'avoir des incidences négatives sur les droits de l'homme ou d'y contribuer et s'efforcer de prévenir ou d'atténuer les incidences négatives qui sont directement liées à leurs activités, produits ou services par leurs relations commerciales, même si elles n'ont pas contribué à ces incidences (principe 13) ;

b) Formuler au plus haut niveau des engagements de principes concernant le respect des droits de leurs utilisateurs (principe 16) ;

c) Faire preuve de diligence raisonnable pour identifier les incidences effectives et potentielles de leurs activités sur les droits de l'homme, y remédier et rendre des comptes à ce sujet, notamment en effectuant régulièrement une évaluation des risques et des incidences, en organisant de véritables consultations avec les groupes susceptibles d'être touchés et les autres acteurs concernés, et en prenant des mesures de suivi qui permettent d'atténuer ou de prévenir ces incidences (principes 17 à 19) ;

d) Mettre en place des stratégies de prévention et d'atténuation qui respectent, dans la mesure du possible, les principes des droits de l'homme internationalement reconnus lorsque le droit interne leur impose des obligations contradictoires (principe 23) ;

e) Passer constamment en revue les mesures prises pour respecter les droits, notamment en consultant régulièrement les parties prenantes et en communiquant fréquemment, de manière accessible et efficace, avec les groupes concernés et avec le public (principes 20 et 21) ;

f) Prévoir des mesures de réparation appropriées et notamment établir au niveau opérationnel des mécanismes de réclamation que les utilisateurs puissent saisir sans que cela aggrave leur impression « de n'avoir aucun pouvoir » (principes 22, 29 et 31).

III. Principales préoccupations concernant la réglementation des contenus

12. les États cherchent à façonner l'environnement dans lequel les entreprises modèrent les contenus, tandis que les entreprises fondent l'accès des utilisateurs à leurs plateformes sur l'acceptation des conditions d'utilisation, qui régissent ce qui peut être exprimé et la manière dont cela peut être exprimé.

A. Réglementation par les États

13. Les États exigent régulièrement des entreprises qu'elles imposent des restrictions concernant les contenus manifestement illégaux tels que les représentations de violences sexuelles sur enfants, les menaces directes et crédibles et l'incitation à la violence, considérant que lesdites restrictions doivent aussi respecter les conditions de légalité et de nécessité [25]. Certains États vont beaucoup plus loin et recourent à la censure et à

[23] Communication de l'Institut danois des droits de l'homme. Cf. Déclaration de Yahoo/Oath, 2016.

[24] Principes directeurs, principe 11.

[25] L'Irlande a mis en place des mécanismes de corégulation avec les entreprises pour imposer des restrictions concernant les contenus montrant des violences sexuelles sur enfants : voir la

l'incrimination pour façonner le cadre réglementaire en ligne[26]. Des lois restrictives formulées en termes généraux sur l'« extrémisme », le blasphème, la diffamation, les discours « offensants », les fausses informations (« *fake news* ») et la « propagande » servent souvent de prétexte pour exiger des entreprises qu'elles suppriment des discours légitimes[27]. On constate de plus en plus que les États ciblent spécifiquement les contenus sur les plateformes en ligne[28]. D'autres lois peuvent porter atteinte au droit au respect de la vie privée en ligne d'une manière qui décourage l'exercice de la liberté d'opinion et d'expression[29]. De nombreux États mettent également en place des outils de désinformation et de propagande destinés à limiter l'accessibilité aux médias indépendants et à nuire à leur crédibilité[30].

14. *Protections contre la responsabilité*. Depuis le début de l'ère numérique, de nombreux États ont adopté des règles visant à faire en sorte que les intermédiaires ne soient pas responsables du contenu publié par des tiers sur leurs plateformes. La directive de l'Union européenne sur le commerce électronique, par exemple, établit un régime juridique visant à protéger les intermédiaires contre la responsabilité relative au contenu, sauf lorsque leurs activités vont au-delà du « simple transport », de la forme de stockage dite « caching » ou de l'« hébergement » de l'information fournie par les utilisateurs[31]. L'article 230 de la loi des États-Unis sur la décence dans le domaine des télécommunications (Communications Decency Act) prévoit de manière générale l'immunité pour les fournisseurs de « services informatiques interactifs » qui hébergent ou publient des informations sur d'autres personnes, immunité qui a été réduite depuis[32]. Le régime de responsabilité des intermédiaires du Brésil prévoit que les restrictions concernant un contenu particulier sont soumises à une décision judiciaire[33], tandis que celui de l'Inde prévoit une procédure « de notification et de retrait » soumise à la décision d'un tribunal ou d'un organisme juridictionnel similaire[34]. Les Principes de Manille de 2014 sur la responsabilité des intermédiaires, élaborés par une coalition d'experts de la société civile, définissent les principes essentiels devant guider tout cadre régissant la responsabilité des intermédiaires.

15. *Imposition d'obligations aux entreprises*. Certains États exigent des entreprises qu'elles appliquent aux contenus des restrictions fondées sur des critères juridiques vagues ou complexes, sans examen judiciaire préalable, avec la menace de sanctions sévères. Par exemple, la loi chinoise de 2016 sur la cybersécurité renforce les interdictions vagues relatives à la diffusion de « fausses » informations qui perturbent « l'ordre social ou économique », l'unité nationale ou la sécurité nationale ; elle oblige également les entreprises à surveiller leurs réseaux et à signaler les violations aux autorités[35]. Les plus

communication de l'Irlande. De nombreuses entreprises utilisent un algorithme de reconnaissance d'images pour détecter et éliminer la pornographie mettant en scène des enfants : voir les communications de l'Open Technology Institute, p. 2 et d'ARTICLE 19, p. 8.

[26] Voir A/HRC/32/38, par. 46 et 47. En ce qui concerne les coupures d'accès à Internet, voir A/HRC/35/22, par. 8 à 16 et des exemples de communications du Rapporteur spécial : n[os] UA TGO 1/2017, UA IND 7/2017 et AL GMB 1/2017.

[27] Communications n[os] OL MYS 1/2018 ; UA RUS 7/2017 ; UA ARE 7/2017, AL BHR 8/2016, AL SGP 5/2016 et OL RUS 7/2016. L'Azerbaïdjan interdit l'apologie du terrorisme, de l'extrémisme religieux et du suicide : voir la communication de l'Azerbaïdjan.

[28] Voir les communications n[os] OL PAK 8/2016 et OL LAO 1/2014 ; Association pour le progrès des communications, *Unshackling Expression : A Study on Laws Criminalising Expression Online in Asia*, GISWatch édition spéciale 2017.

[29] A/HRC/29/32.

[30] Voir, par exemple, Gary King, Jennifer Pan et Margaret E. Roberts, *How the Chinese Government fabricates social media posts for strategic distraction, not engaged argument*, *American Political Science Review*, vol. 111, n[o] 3 (2017), p. 484 à 501.

[31] Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000.

[32] Code des États-Unis, titre 47, art. 230. Voir aussi la loi visant à permettre aux États et aux victimes de lutter contre la traite à des fins d'exploitation sexuelle en ligne (H.R. 1865).

[33] *Marco Civil da Internet*, loi fédérale 12.965, art. 18 et 19.

[34] Cour suprême de l'Inde, *Shreya Singhal* v. *Union of India*, décision du 24 mars 2015.

[35] Art. 12 et 47 ; communication sur les droits de l'homme en Chine, 2016, p. 12. Pour des observations sur une version préliminaire de la loi sur la cybersécurité, voir la communication n[o] OL CHN 7/2015. Voir aussi Global Voices, *Netizen Report : Internet censorship bill looms large over Egypt*, 16 mars

grandes plateformes de médias sociaux du pays se seraient vu infliger de lourdes amendes pour avoir manqué à ces obligations[36].

16. Les obligations relatives à la surveillance et à la suppression rapide des contenus créés par les utilisateurs ont également été renforcées au niveau mondial par la mise en place de cadres répressifs susceptibles de compromettre la liberté d'expression, même dans les sociétés démocratiques. La loi allemande relative au respect de la loi sur les réseaux (*NetzDG*) fait obligation aux grandes entreprises de médias sociaux de supprimer les contenus incompatibles avec certaines lois locales et prévoit des peines sévères en cas de non-respect dans des délais très courts[37]. La Commission européenne a été jusqu'à recommander aux États membres de créer des obligations juridiques pour la surveillance active et le filtrage des contenus illégaux[38]. Au Kenya, les lignes directrices sur la diffusion de contenus sur les médias sociaux pendant les élections, adoptées en 2017, obligent les entreprises à supprimer les comptes utilisés pour diffuser des contenus politiques indésirables sur leurs plateformes dans un délai de vingt-quatre heures[39].

17. Compte tenu de leurs préoccupations légitimes concernant notamment le respect de la vie privée et la sécurité nationale, il est compréhensible que les États souhaitent disposer de réglementations. Toutefois, les règles de ce type comportent des risques pour la liberté d'expression car elles font peser une pression importante sur les entreprises, de sorte que celles-ci peuvent supprimer des contenus licites pour éviter, de manière générale, que leur responsabilité soit engagée. Elles impliquent également la délégation de fonctions de réglementation à des acteurs privés qui ne disposent pas des outils de base pour l'établissement des responsabilités. Les demandes de suppressions rapides et automatiques risquent de donner lieu à de nouvelles formes de restrictions préalables qui menacent déjà la créativité dans le contexte des droits d'auteur[40]. Les questions complexes de fait 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[41].

18. *Suppressions à l'échelle mondiale*. Certains États exigent la suppression, à l'étranger, de liens, de sites Web et de contenus qui seraient contraires à leur législation locale[42]. Les revendications de ce type font craindre que les États ne puissent porter atteinte au droit à la liberté d'expression « sans considération de frontières ». Dans cette logique, il serait possible d'exercer une censure au-delà des frontières, ce dont profiteront les censeurs les plus restrictifs. Ceux qui souhaitent des suppressions devraient être tenus de présenter

2018 ; République d'Afrique du Sud, projet de loi portant modification de la loi sur les films et les publications (B 61-2003).

36 PEN America, Forbidden Feeds: Government Controls on Social Media in China (2018), p. 21.

37 Loi visant à améliorer l'application de la loi sur les réseaux sociaux (*NetzDG*), juillet 2017. Voir la communication nº OL DEU 1/2017.

38 Commission européenne, recommandation relative à des mesures visant à lutter efficacement contre le contenu illicite en ligne (dernière mise à jour : 5 mars 2018).

39 Voir la communication nº OL KEN 10/2017 ; Javier Pallero, *Honduras: new bill threatens to curb online speech*, Access Now, 12 février 2018.

40 Voir Commission européenne, Proposition de directive du Parlement européen et du Conseil sur le droit d'auteur dans le marché unique numérique, COM (2016) 593 final, art. 13 ; Daphne Keller, *Problems with filters in the European Commission's platforms proposal*, Center for Internet and Society, Stanford Law School, 5 octobre 2017 ; communication de la Fundación Karisma, 2016, p. 4 à 6.

41 Dans le droit de l'Union européenne, les moteurs de recherche doivent déterminer la validité des réclamations présentées au titre du « droit à l'oubli numérique ». Cour de justice de l'Union européenne, *Google Spain* c. *Agencia Española de Protección de Datos et Mario Costeja González* (affaire C-131/12), arrêt de la Cour (grande chambre) du 13 mai 2014 ; communication d'ARTICLE 19, p. 2 et 3 et d'Access Now, p. 6 et 7 ; Google, *Updating our'right to be forgotten' Transparency Report* ; Theo. Bertram et autres, *Three Years of the Right to be Forgotten* (Google, 2018).

42 Voir, par exemple, PE24N America, *Forbidden Feeds*, p. 36 et 37 ; Cour suprême du Canada, *Google Inc.* c. *Equestek Solutions Inc.* jugement du 28 juin 2017 ; Cour de justice de l'Union européenne, *Google Inc.* c. *Commission nationale de l'informatique et des libertés (CNIL)* (affaire C-507/17) ; communication de Global Network Initiative, p. 6.

leurs demandes dans chacun des pays concernés, dans le cadre d'une procédure juridique et judiciaire régulière.

19. *Demandes des États non fondées sur la législation nationale*. Les entreprises font une distinction entre les demandes de suppression de contenus présumés illégaux soumises par les voies judiciaires régulières et les demandes de suppression fondées sur les conditions d'utilisation des entreprises[43]. Les demandes de suppression soumises par les voies judiciaires ne s'appliquent généralement que dans l'État qui présente la demande, tandis que celles fondées sur les conditions d'utilisation des entreprises s'appliquent généralement à l'échelle mondiale. Les pouvoirs publics cherchent de plus en plus souvent à obtenir la suppression de contenus sans passer par les procédures judiciaires voire sans présenter de demande de suppression fondée sur les conditions d'utilisation des entreprises[44]. Plusieurs États ont créé des unités administratives spécialisées chargées de signaler des contenus aux entreprises pour qu'elles les suppriment. L'unité de l'Union européenne chargée du signalement des contenus sur Internet, par exemple, signale les contenus à caractère terroriste et extrémiste violent en ligne et coopère avec les fournisseurs de services en ligne aux fins de la suppression de ces contenus[45]. L'Australie dispose de mécanismes de signalement similaires[46]. En Asie du Sud-Est, des partis politiques alliés des gouvernements tenteraient d'utiliser les demandes fondées sur les conditions d'utilisation des entreprises pour faire taire les critiques dont ils font l'objet[47].

20. Les États font également pression sur les entreprises pour qu'elles accélèrent la suppression de contenus à l'aide de mesures non contraignantes qui ne sont pour la plupart pas tout à fait transparentes. Au Pakistan, l'accès à YouTube a été bloqué pendant trois ans, ce qui a contraint Google à élaborer une version locale de la plateforme qui permette de répondre aux demandes de suppression de contenus « offensants » formulées par le Gouvernement[48]. Facebook et Israël auraient décidé de coordonner leurs efforts et de faire collaborer leurs personnels pour surveiller et supprimer les « incitations » en ligne. Les détails de cet accord n'ont pas été divulgués, mais le Ministre israélien de la justice a affirmé qu'entre juin et septembre 2016, Facebook avait accédé à la quasi-totalité des demandes de suppression des « incitations » présentées par le Gouvernement[49]. Les dispositions prises aux fins de la coordination avec les États des mesures concernant les contenus font craindre encore davantage que les entreprises exercent des fonctions publiques sans le contrôle des tribunaux ou d'autres mécanismes d'établissement des responsabilités[50].

21. Le code de conduite visant à combattre les discours de haine illégaux en ligne établi en 2016 par l'Union européenne a trait à un accord sur la suppression de contenus conclu entre l'Union européenne et quatre grandes entreprises, par lequel celles-ci s'engagent à collaborer avec des « signaleurs de confiance » et à promouvoir des « contre-discours indépendants »[51]. Si la promotion de contre-discours est une réponse intéressante face à un contenu à caractère « extrémiste » ou « terroriste », la pression exercée en faveur de telles

[43] Comparer la section du rapport de transparence de Twitter sur les demandes de retrait (janvier-juin 2017) et celle sur les rapports de conditions d'utilisation par des autorités gouvernementales (janvier-juin 2017). Voir aussi Facebook, Questions/Réponses sur les demandes gouvernementales.

[44] Communications d'ARTICLE 19, p. 2, et de Global Network Initiative, p. 5.

[45] Union européenne, Unité de signalement des contenus sur Internet, Year One Report, sect. 4.11 ; communications de l'European Digital Rights (EDRi), p. 1, et d'Access Now, p. 2 et 3.

[46] Communication de l'Australie.

[47] Southeast Asian Press Alliance, p. 1.

[48] Communication de la Digital Rights Foundation.

[49] Communication de 7amleh − The Arab Center for the Advancement of Social Media.

[50] Association pour le progrès des communications, p. 14, et 7amleh.

[51] Le terme « signaleurs de confiance » fait référence au statut accordé à certaines organisations, qui leur permet de signaler des contenus illégaux au moyen d'un système ou d'un canal de signalement spécial, qui n'est pas accessible aux utilisateurs traditionnels. Commission européenne, *Code of Conduct on countering illegal hate speech online: First results on implementation* (décembre 2016).

solutions présente le risque que les plateformes soient transformées en vecteurs d'une propagande qui sortirait largement du cadre des préoccupations légitimes[52].

B. Modération des contenus par les entreprises

Respect de la législation nationale par les entreprises

22. Chaque entreprise s'engage en principe à se conformer à la législation du territoire dans lequel elle exerce ses activités. Facebook indique que, si après un examen juridique minutieux, il détermine qu'un contenu est illégal au regard de la législation locale, il le rend indisponible dans le pays ou le territoire concerné[53]. Tencent, le propriétaire de l'application mobile de dialogue en ligne et de médias sociaux WeChat, va beaucoup plus loin et exige que toute personne utilisant la plateforme en Chine et tout citoyen chinois utilisant la plateforme « n'importe où dans le monde » respecte les restrictions applicables aux contenus qui sont prévues par la législation ou la politique chinoises[54]. En outre, plusieurs entreprises collaborent entre elles et avec des organismes de réglementation pour supprimer les représentations de violences sexuelles sur enfants[55].

23. L'engagement pris de se conformer à la législation d'un État peut être difficile à respecter lorsque cette législation est vague, sujette à diverses interprétations ou incompatible avec le droit des droits de l'homme. À titre d'exemple, les lois contre l'« extrémisme » qui ne définissent pas le terme clef donnent aux organismes publics le pouvoir discrétionnaire de faire pression sur les entreprises pour qu'elles suppriment des contenus pour des motifs contestables[56]. De même, des pressions sont souvent exercées sur les entreprises pour qu'elles se conforment à des lois nationales qui érigent en infraction les contenus dits blasphématoires, critiques à l'égard de l'État, diffamatoires à l'égard des fonctionnaires ou faux, par exemple. Comme expliqué plus loin, les Principes directeurs prévoient des outils visant à réduire le plus possible les incidences négatives de ces lois sur les utilisateurs. D'autres directives sur la façon d'utiliser ces outils ont été élaborées par l'Initiative mondiale des réseaux, une initiative multipartite qui aide les entreprises du secteur des TIC à gérer les problèmes qu'elles rencontrent dans le domaine des droits de la personne[57]. L'un de ces outils est la transparence : de nombreuses entreprises rendent compte chaque année du nombre de demandes formulées par des gouvernements qu'elles reçoivent et qu'elles exécutent, en les classant par État[58]. Toutefois, les entreprises ne divulguent pas toujours suffisamment d'informations sur la façon dont elles répondent aux demandes des gouvernements, et ne rendent pas régulièrement compte des demandes des gouvernements fondées sur les conditions d'utilisation[59].

[52] Les mêmes entreprises ont créé le Forum mondial contre le terrorisme sur Internet dans le but de mettre au point des outils technologiques sectoriels en vue d'éliminer de leurs plateformes les contenus à caractère terroriste. Google, *Update on the Global Internet Forum to Counter Terrorism*, 4 décembre 2017.

[53] Facebook, Questions/Réponses sur les demandes gouvernementales. Voir aussi Google, Demandes légales de suppression ; Twitter, Règles et politiques ; Reddit, politique relative au contenu.

[54] Tencent, Terms of Service : Introduction ; Tencent, Agreement on Software License and Service of Tencent Weixin.

[55] Organisation des Nations Unies pour l'éducation, la science et la culture, *Fostering Freedom Online : The Role of Internet Intermediaries* (Paris, 2014), p. 56 et 57.

[56] Voir Maria Kravchenko, *Inappropriate enforcement of anti-extremist legislation in Russia in 2016*, SOVA Center for Information and Analysis, 21 avril 2017 ; Danielle Citron, *Extremist speech, compelled conformity, and censorship creep*, Notre Dame Law Review, vol. 93, nº 3 (2018), p. 1035 à 1071.

[57] Initiative mondiale des réseaux, Principes de liberté d'expression et de respect de la vie privée, sect. 2. Les entreprises de médias sociaux qui participent à l'Initiative sont notamment Facebook, Google, Microsoft/LinkedIn et Yahoo/Oath.

[58] Voir plus loin le paragraphe 39. De plus, Automattic, Google, Microsoft/Bing et Twitter font partie des entreprises qui publient régulièrement, mais pas nécessairement de façon exhaustive, des demandes formulées par des gouvernements relatives à des suppressions et à la propriété intellectuelle dans la base de données Lumen.

[59] Ranking Digital Rights, 2017 Corporate Accountability Index, p. 28.

Normes de modération des entreprises

24. Les entreprises du Web exigent de leurs utilisateurs qu'ils respectent les conditions d'utilisation et les « normes communautaires » qui régissent l'expression sur leurs plateformes[60]. Ces conditions d'utilisation, que les utilisateurs sont tenus d'accepter pour pouvoir utiliser la plateforme, déterminent les juridictions compétentes pour la résolution des différends et réservent aux entreprises un pouvoir discrétionnaire concernant leurs actions relatives aux contenus et aux comptes[61]. Les politiques relatives aux contenus des plateformes constituent l'un des volets de ces conditions d'utilisation ; elles définissent les limitations concernant ce que les utilisateurs peuvent exprimer et la façon dont ils peuvent le faire. La plupart des entreprises ne fondent pas explicitement leurs normes relatives aux contenus sur un corpus de lois particulier qui pourraient régir l'expression, comme le droit national ou le droit international des droits de l'homme. Cependant, le géant chinois des moteurs de recherche, Baidu, interdit les contenus qui sont « contraires aux principes de base consacrés par la Constitution » de la République populaire de Chine[62].

25. L'élaboration de politiques de modération de contenus nécessite généralement l'intervention de conseillers juridiques, de gestionnaires de politiques publiques et de produits, ainsi que de cadres supérieurs. Les entreprises peuvent constituer des équipes de « confiance et de sécurité » chargées de lutter contre les spams, la fraude et les abus, et des équipes chargées de lutter contre les contenus terroristes[63]. Certaines ont mis au point des mécanismes permettant de solliciter la contribution de groupes externes sur des aspects spécialisés des politiques relatives aux contenus[64]. L'augmentation exponentielle des contenus créés par les utilisateurs a donné lieu à l'élaboration de règles détaillées, qui sont en constante évolution. Ces règles varient selon une série de facteurs, allant de la taille de l'entreprise, de ses revenus et de son modèle économique, à l'image et à la réputation de la plateforme, à son niveau de tolérance face au risque et à la façon dont elle souhaite que ses utilisateurs participent[65].

Domaines de préoccupation concernant les normes relatives aux contenus

26. *Des règles vagues*. L'interdiction faite par les entreprises de proférer des menaces ou de promouvoir le terrorisme[66], d'apporter un appui aux dirigeants d'organisations dangereuses ou d'en faire l'éloge[67] et de publier des contenus qui encouragent les actes terroristes ou incitent à la violence[68] est, tout comme la législation antiterroriste, formulée de manière excessivement vague[69]. Les politiques des entreprises en matière de haine, de harcèlement et de mauvais traitements n'indiquent pas clairement ce qui constitue une infraction. L'interdiction par Twitter des « comportements qui incitent à craindre un groupe protégé » et la distinction faite par Facebook entre les « attaques directes » contre des

[60] Jamila Venturini et autres, *Terms of Service and Human Rights: An Analysis of Online Platform Contracts* (Rio de Janeiro, Revan, 2016).

[61] Accord d'utilisation de Baidu (Le retrait et la suppression de tout contenu de cette plateforme se fait à la seule discrétion de Baidu, pour quelque raison que ce soit.) ; Conditions d'utilisation de Tencent (Nous nous réservons le droit de bloquer ou de supprimer votre contenu pour quelque raison que ce soit, y compris si nous le jugeons approprié ou si les lois et règlements applicables l'exigent) ; Conditions d'utilisation de Twitter (Nous pouvons suspendre ou résilier votre compte, ou cesser de vous fournir tout ou partie des Services, à tout moment, pour quelque raison que ce soit ou sans raison.).

[62] Conditions d'utilisation de Baidu, sect. 3.1.

[63] Monika Bickert, *Hard questions: how we counter terrorism*, 15 juin 2017.

[64] Voir, par exemple, le Conseil Confiance et Sécurité de Twitter et le programme YouTube Trusted Flagger.

[65] Sarah Roberts, *Content Moderation* (Université de Californie à Los Angeles, 2017). Voir également la communication d'ARTICLE 19, p. 2.

[66] Règles et politiques de Twitter (Groupes extrémistes violents).

[67] Standards de la communauté Facebook (organisations dangereuses).

[68] Politiques de YouTube (politiques relatives aux contenus visuels choquants ou violents).

[69] Voir A/HRC/31/65, par. 39.

caractéristiques protégées et les contenus simplement « désagréables ou offensants » sont des critères subjectifs et mouvants aux fins de la modération des contenus[70].

27. *Haine, harcèlement et abus*. Le manque de précision des politiques en matière de discours haineux et de harcèlement a entraîné des plaintes liées à une application incohérente des politiques qui pénalise les minorités tout en renforçant la position des groupes dominants ou puissants. Les utilisateurs et la société civile signalent des violences et des abus à l'égard des femmes, y compris des menaces physiques, des commentaires misogynes, l'affichage d'images intimes sans le consentement des intéressées ou d'images truquées et la publication de données et d'informations personnelles dans l'intention de nuire (« doxing »)[71] ; des menaces contre les personnes privées de leur droit de vote[72], les races et les castes minoritaires[73] et les groupes ethniques victimes de persécutions violentes[74] ; et des abus à l'égard des réfugiés, des migrants et des demandeurs d'asile[75]. Dans le même temps, des plateformes auraient supprimé des contenus militants émanant de la communauté des lesbiennes, gays, bisexuels, transgenres et queers[76], des plaidoyers contre des gouvernements répressifs[77], des informations sur les nettoyages ethniques[78] et des critiques de phénomènes racistes et de structures de pouvoir[79].

28. L'ampleur et la complexité de la lutte contre l'expression de la haine soulèvent des défis à long terme et peuvent amener les entreprises à restreindre cette expression même lorsqu'elle n'est pas clairement liée à des conséquences néfastes (comme l'appel à la haine lié à l'incitation au sens de l'article 20 du Pacte international relatif aux droits civils et politiques). Les entreprises devraient toutefois énoncer les fondements de ces restrictions et démontrer la nécessité et la proportionnalité de toute mesure prise à l'égard de contenus (comme les suppressions ou les suspensions de compte). Une transparence véritable et constante en matière d'application des politiques relatives aux discours haineux, supposant l'établissement de rapports sur les cas précis, peut également fournir des éclaircissements que même les explications les plus détaillées ne pourraient pas apporter[80].

29. *Contexte*. Les entreprises soulignent l'importance du contexte dans leur évaluation de l'applicabilité des restrictions générales[81]. Néanmoins, l'attention portée au contexte n'a pas empêché la suppression de représentations de la nudité ayant une valeur historique, culturelle ou éducative[82] ; de récits historiques et documentaires de conflits[83] ; de preuves de crimes de guerre[84] ; de contre-discours s'opposant aux propos de groupes animés par la

[70] Standards de la communauté Facebook (Discours haineux) ; Règles et politiques de Twitter (Politique en matière de conduite haineuse).

[71] Amnesty International, Toxic Twitter : A Toxic Place for Women; communication de l'Association pour le progrès des communications, p. 2.

[72] Communications de 7amleh et de l'Association pour le progrès des communications, p. 15.

[73] Ijeoma Oluo, « Facebook's complicity in the silencing of black women », Medium, 2 août 2017 ; communications du Center for Communications Governance, p. 5 et de l'Association pour le progrès des communications, p. 11 et 12.

[74] Déclaration de la Rapporteuse spéciale sur la situation des droits de l'homme au Myanmar, Yanghee Lee, lors de la trente-septième session du Conseil des droits de l'homme, 12 mars 2018.

[75] Communication de l'Association pour le progrès des communications, p. 12.

[76] Communication d'Electronic Frontier Foundation, p. 5.

[77] Ibid. ; communications de l'Association pour le progrès des communications et de 7amleh.

[78] Betsy Woodruff, « Facebook silences Rohingya reports of ethnic cleansing », The Daily Beast, 18 septembre 2017; Communication d'ARTICLE 19, p. 9.

[79] Julia Angwin et Hannes Grasseger, « Facebook's secret censorship rules protect white men from hate speech but not black children", ProPublica, 28 juin 2017.

[80] Voir plus loin par. 52 et 62.

[81] Twitter, « Notre approche en matière d'élaboration de politiques et notre philosophie relative à l'application de ces dernières » ; Politiques YouTube (The importance of context) ; Richard Allan, « Hard questions: who should decide what is hate speech in an online global community ? » Facebook Newsroom, 27 juin 2017.

[82] Communications d'OBSERVACOM, p. 11, et d'ARTICLE 19, p. 6.

[83] Communication de WITNESS, p. 6 et 7.

[84] Ibid.

haine[85] ; ou d'actions destinées à contester des propos racistes, homophobes ou xénophobes ou à en demander la suppression[86]. L'examen adéquat du contexte peut être compromis par le manque de temps et de ressources alloués aux modérateurs humains, par une dépendance excessive à l'automatisation ou par une compréhension insuffisante des nuances linguistiques et culturelles[87]. Les entreprises ont exhorté les utilisateurs à compléter les contenus controversés par des informations contextuelles, mais l'applicabilité et l'efficacité de ces instructions sont incertaines[88].

30. *Exigences relatives au nom réel*. Pour faire face aux violences en ligne, certaines entreprises ont des exigences en matière d'« authenticité de l'identité »[89] ; d'autres abordent les questions d'identité avec plus de souplesse[90]. L'efficacité des exigences relatives au nom réel en tant que garde-fou contre les violences en ligne est discutable[91]. En effet, cette exigence relative aux noms réels a fait se démasquer des blogueurs et des militants qui utilisaient des pseudonymes pour se protéger, ce qui les a exposés à de graves dangers physiques[92]. Cela a également entraîné le blocage de comptes d'utilisateurs et de militants de la communauté des lesbiennes, gays, bisexuels, transsexuels, transgenres et queers, d'artistes travestis et d'utilisateurs dont le nom n'était pas anglais ou pas usuel[93]. Étant donné que l'anonymat en ligne est souvent nécessaire à la sécurité physique d'utilisateurs vulnérables, les principes des droits de l'homme s'appliquent par défaut à la protection de l'anonymat et ne peuvent faire l'objet que des restrictions qui viseraient à protéger leur identité[94]. Des règles relatives à l'usurpation d'identité rédigées en termes précis, qui limitent la possibilité pour les utilisateurs de représenter une autre personne d'une manière prêtant à confusion ou trompeuse, peuvent être un moyen plus proportionné de protéger l'identité, les droits et la réputation des autres utilisateurs[95].

31. *Désinformation*. La désinformation et la propagande compromettent l'accès à l'information et la confiance du public à l'égard des médias et des institutions étatiques. Les entreprises subissent de plus en plus de pressions visant à ce qu'elles luttent contre la désinformation diffusée au moyen de liens vers de faux articles ou de faux sites Web de tiers, de faux comptes et de publicités trompeuses, et par la manipulation du classement des recherches[96]. Toutefois, étant donné que les formes d'action brutales, telles que le blocage de sites Web ou les mesures de suppression, risquent de porter gravement atteinte à la liberté d'expression, les entreprises devraient formuler avec soin toute politique portant sur la désinformation[97]. Elles ont adopté une variété de mesures, y compris des arrangements avec des tiers chargés de vérifier les informations, l'application renforcée des politiques publicitaires, la surveillance accrue des comptes suspects, la modification des algorithmes d'édition des contenus et de classement des recherches, et la formation des utilisateurs au repérage des fausses informations[98]. Certaines mesures, en particulier celles qui renforcent

[85] Communication d'Electronic Frontier Foundation, p. 5.

[86] Communication de l'Association pour le progrès des communications, p. 14.

[87] Voir Allan, « Hard questions ».

[88] Politiques YouTube (The importance of context) ; Standards de la communauté Facebook (Discours haineux).

[89] Standards de la communauté Facebook (Utilisation de votre véritable identité). À noter que Facebook autorise à présent des exceptions à sa politique sur les noms réels au cas par cas, mais ces exceptions sont jugées insuffisantes par certains : communication d'Access Now, p. 12. Baidu exige même l'utilisation d'informations d'identification personnelle : voir l'accord d'utilisation de Baidu.

[90] Centre d'assistance Twitter, « Aide sur la création d'un nom d'utilisateur » ; Instagram, « Démarrer sur Instagram ».

[91] J. Nathan Matias, « The real name fallacy », Coral Project, 3 janvier 2017.

[92] Communication d'Access Now, p. 11.

[93] Dia Kayyali, « Facebook's name policy strikes again, this time at Native Americans », Electronic Frontier Foundation, 13 février 2015.

[94] Voir A/HRC/29/32, par. 9.

[95] Règles et Politiques de Twitter (Politique en matière d'usurpation d'identité).

[96] Ibid. ; Allen Babajanian et Christine Wendel, « #FakeNews: innocuous or intolerable ? », Wilton Park report 1542, avril 2017.

[97] Déclaration conjointe 2017.

[98] Communications de l'Association pour le progrès des communications, p. 4 à 6, et d'ARTICLE 19, p. 4.

les restrictions sur le contenu des informations, peuvent menacer les sources d'information indépendantes et non traditionnelles ou les contenus satiriques[99]. Les autorités étatiques ont pris des positions qui peuvent refléter des attentes démesurées quant à la capacité de la technologie à résoudre à elle seule de tels problèmes[100].

Procédures et outils de modération utilisés par les entreprises

32. *Signalement, suppression et filtrage avant publication automatisés*. L'ampleur des contenus générés par les utilisateurs a conduit les plus grandes entreprises à développer des outils de modération automatisés. L'automatisation est utilisée principalement pour le signalement des contenus à des fins d'analyse par l'homme et parfois pour leur suppression. Les outils automatisés d'analyse de la musique et des vidéos qui servent à repérer les violations du droit d'auteur au stade du téléchargement soulèvent des préoccupations quant au caractère excessif de blocages, et les appels à l'extension du filtrage du téléchargement à des domaines liés au terrorisme et à d'autres domaines font craindre l'instauration de régimes complets et disproportionnés de censure préalable à la publication[101].

33. L'automatisation peut présenter un intérêt pour les entreprises qui analysent d'énormes volumes de contenu généré par les utilisateurs, avec des outils allant des filtres par mots-clefs et de la détection des spams aux algorithmes de *hash-matching* et au traitement du langage naturel[102]. Le *hash-matching* est très utilisé pour repérer les images d'agressions sexuelles sur enfants, mais son application aux contenus à caractère « extrémiste » − qui nécessite généralement une analyse du contexte − est difficile à réaliser sans règles claires concernant l'« extrémisme » ou sans une évaluation par l'homme[103]. Il en va de même pour le traitement du langage naturel[104].

34. *Signalement (*flagging*) par l'utilisateur et signalement « digne de confiance »*. Les fonctionnalités de signalement à la disposition des utilisateurs donnent aux individus la possibilité de soumettre des plaintes concernant des contenus inappropriés à des modérateurs de contenus. Ces fonctionnalités ne permettent généralement pas de discussions nuancées sur les limites souhaitables (par exemple sur les raisons pour lesquelles un contenu, s'il peut être offensant, ne devrait tout compte fait pas être supprimé)[105]. Elles sont aussi « exploitées » pour accroître la pression sur des plateformes afin qu'elles suppriment des contenus qui soutiennent des minorités sexuelles ou les musulmans[106]. De nombreuses entreprises ont établi des listes spécialisées de « signaleurs » (*flaggers*) « dignes de confiance », généralement des experts ou des utilisateurs influents ou parfois, paraît-t-il, des signaleurs publics[107]. Il existe peu, voire pas, d'informations publiques expliquant le choix des signaleurs spécialisés, l'interprétation que ceux-ci font des normes juridiques ou communautaires ou leur influence sur les décisions de l'entreprise.

35. *Évaluation par l'homme*. L'automatisation est souvent complétée par une analyse effectuée par un être humain ; en effet, les plus grandes entreprises de médias sociaux

[99] Association pour le progrès des communications, p. 5.

[100] Voir la communication nº OL ITA 1/2018. Cf. Commission européenne, *A Multi-Dimensional Approach to Disinformation: Final Report of the Independent High-level Group on Fake News and Disinformation* (Luxembourg, 2018).

[101] Le Royaume-Uni de Grande-Bretagne et d'Irlande du Nord aurait mis au point un outil permettant de détecter et de supprimer automatiquement les contenus à caractère terroriste au stade du téléchargement. Ministère britannique de l'intérieur, « New technology revealed to help fight terrorist content online », 13 février 2018.

[102] Center for Democracy and Technology, *Mixed Messages ? The Limits of Automated Media Content Analysis* (novembre 2017), p. 9.

[103] Communication de l'Open Technology Institute, p. 2.

[104] Center for Democracy and Technology, *Mixed Messages ?*, p. 4.

[105] Sur les fonctionnalités de signalement à la disposition des utilisateurs, voir à titre général Kate Crawford et Tarleton Gillespie, « What is a flag for? Social media reporting tools and the vocabulary of complaint », *New Media and Society*, vol. 18, nº 3 (mars 2016), p. 410 à 428.

[106] Ibid., p. 421.

[107] Aide YouTube, Programme YouTube Trusted Flagger ; Aide YouTube, « Get involved with YouTube contributors ».

mettent en place de grandes équipes de modérateurs de contenus chargées d'analyser les contenus signalés[108]. Ces contenus peuvent être acheminés vers les modérateurs, qui sont généralement autorisés à prendre une décision − souvent en quelques minutes − sur le caractère approprié ou non du contenu, puis à le supprimer ou à l'autoriser. Dans les cas où il est difficile de déterminer si un contenu particulier est ou non approprié, les modérateurs peuvent l'envoyer pour examen aux équipes de contenu du siège social de l'entreprise. Des responsables de l'entreprise − généralement des équipes chargées de la politique générale ou des équipes « confiance et sécurité », qui comprennent des conseillers juridiques − prennent alors les décisions relatives à la suppression de contenus. La divulgation par les entreprises d'informations sur les discussions qui aboutissent à la suppression de contenus, qu'il s'agisse de récapitulatifs ou d'informations sur des cas précis, est limitée[109].

36. *Mesures prises à l'égard d'un compte ou d'un contenu*. L'existence d'un contenu inapproprié peut amener l'entreprise à prendre une série de mesures. L'entreprise peut limiter la suppression du contenu en question à la juridiction dont le contenu relève, viser un ensemble de juridictions ou viser une plateforme entière ou un ensemble de plateformes. Elle peut appliquer des limites d'âge, émettre des avertissements ou décider d'une démonétisation[110]. Les infractions peuvent entraîner la suspension temporaire du compte, tandis que les récidives peuvent entraîner la désactivation du compte. Il est très rare, mis à part en ce qui concerne l'application de la législation sur les droits d'auteur, que les entreprises prévoient des procédures de « contre-notification » permettant aux utilisateurs qui publient du contenu de contester une suppression.

37. *Notification*. Il est fréquent d'entendre dénoncer le fait que les utilisateurs qui publient des contenus qui font l'objet d'un signalement ou les personnes qui se plaignent d'abus ne sont pas nécessairement informés de la suppression du contenu ou du fait que des mesures ont été prises[111]. Même lorsque les entreprises publient des notifications, celles-ci n'indiquent généralement que la mesure prise et un motif général. Une entreprise au moins a tenté de contextualiser davantage ses notifications, mais on ne sait pas vraiment si le fait de donner des détails supplémentaires dans des notifications générales permet d'apporter des explications suffisantes dans tous les cas[112]. La transparence et les notifications vont de pair : une transparence solide au niveau opérationnel, qui permet aux utilisateurs de mieux comprendre la politique de la plateforme en matière de suppression de contenu réduit la pression sur les notifications dans les cas individuels, tandis qu'une transparence globale plus faible augmente la probabilité que les utilisateurs ne soient pas en mesure de comprendre les suppressions de contenu individuelles en l'absence de notifications adaptées aux cas particuliers.

38. *Recours et réparations*. Les plateformes permettent de former des recours contre toute une série de mesures allant de la suppression de profils ou de pages à la suppression de messages, de photos ou de vidéos[113]. Toutefois, même en cas de recours, les réparations offertes aux utilisateurs s'avèrent limitées ou inopportunes, au point d'être inexistantes et, en tout état de cause, opaques pour la plupart des utilisateurs et même pour des experts de la société civile. Il se peut, par exemple, que le rétablissement du contenu soit une mesure insuffisante si la suppression a entraîné un préjudice particulier − tel qu'une atteinte à la réputation ou un préjudice physique, moral ou financier − pour la personne qui l'avait publié. De même, les suspensions de compte ou les suppressions de contenus au cours de

[108] Voir Sarah Roberts, « Commercial content moderation: digital laborers' dirty work », *Media Studies Publications*, paper 12 (2016).

[109] Voir Wikipedia: BOLD, revert, discuss cycle. Les modérateurs de Reddit sont encouragés à aider les nouveaux utilisateurs et les utilisateurs désorientés en leur apportant des explications sur les règles applicables, en leur donnant des « trucs » ou en leur fournissant des liens (voir Reddit Moddiquette).

[110] Politiques de YouTube (Nudité ou contenu à caractère sexuel) ; Aide YouTube, « Creator influence on YouTube ».

[111] Communications d'ARTICLE 19, p. 7, et de l'Association pour le progrès des communications, p. 16.

[112] Voir https://twitter.com/TwitterSafety/status/971882517698510848/.

[113] Electronic Frontier Foundation et Visualizing Impact, « How to appeal », onlinecensorship.org. Facebook et Instagram n'autorisent les recours que contre les suspensions de comptes. Cf. communication de GitHub, p. 6.

manifestations ou de débats publics peuvent avoir un effet considérable sur les droits politiques, mais il n'y a pas de réparations prévues par les entreprises dans de tels cas.

Transparence

39. Les entreprises rédigent des rapports de transparence dans lesquels elles publient des données agrégées sur les demandes de suppression de contenus et de communication de données d'utilisateurs émanant de l'État. Ces rapports montrent le type de pressions auxquelles les entreprises sont soumises. Les rapports de transparence recensent, pays par pays, le nombre de demandes de suppression juridiques[114], le nombre de demandes à la suite desquelles certaines mesures ont été prises ou des restrictions ont été appliquées à des contenus[115] et, de plus en plus, des descriptions et des exemples de fondements légaux[116].

40. Cependant, comme le conclut la principale étude sur la transparence de l'Internet, les entreprises divulguent le moins d'informations possible sur la manière dont les règles et mécanismes *privés* d'autorégulation et de corégulation sont conçus et mis en œuvre[117]. En particulier, la divulgation des mesures prises à la suite de demandes privées de suppression de contenu au titre des conditions d'utilisation est « incroyablement limitée »[118]. Les normes relatives au contenu sont rédigées en termes généraux, ce qui laisse aux plateformes une grande marge de manœuvre, sur laquelle les entreprises ne donnent pas suffisamment de précisions. Le contrôle dont elles font l'objet de la part des médias et du public a conduit les entreprises à compléter les politiques générales par des articles explicatifs publiés sur des blogs[119] et par des exemples hypothétiques limités[120], mais ceux-ci n'apportent pas suffisamment d'éclaircissements sur la manière précise dont les règles internes sont élaborées et appliquées[121]. Alors que les conditions d'utilisation sont généralement disponibles dans les langues locales, les rapports de transparence et les blogs d'entreprise et contenus connexes ne le sont pas, ce qui rend les choses encore moins claires pour les utilisateurs non anglophones. En conséquence, les utilisateurs, les autorités publiques et la société civile expriment souvent leur insatisfaction face à l'imprévisibilité des mesures prises au titre des conditions d'utilisation[122]. Leur engagement insuffisant, conjugué aux critiques croissantes du public, contraint les entreprises à évaluer, réviser et défendre constamment leurs règles.

IV. Principes des droits de l'homme applicables à la modération des contenus par les entreprises

41. Le fondateur de Facebook a récemment affirmé qu'il espérait un mode de fonctionnement dans le cadre duquel l'entreprise pourrait prendre en considération plus fidèlement les valeurs de la communauté d'un endroit à l'autre[123]. Un tel mode de fonctionnement, de même que les normes pertinentes, est prévu par le droit des droits de l'homme. Les normes privées, qui varient en fonction du modèle économique de

[114] Rapport de transparence de Twitter : Demandes de retrait (janvier à juin 2017) ; Rapport de transparence de Google : Demandes gouvernementales de suppression de contenu ; 2016 Reddit Inc., Transparency Report. Facebook ne précise pas le nombre total de demandes reçues par pays.

[115] Voir, par exemple, Rapport de transparence de Facebook (France) (janvier à juin 2017) ; Rapport de transparence de Google : Demandes gouvernementales de suppression de contenu (Inde) ; Rapport de transparence de Twitter (Turquie).

[116] Ibid.

[117] Communication de Ranking Digital Rights, p. 4. Les italiques figurent dans l'original.

[118] Ibid., p. 10.

[119] Voir Elliot Schrage, « Introducing hard questions », Facebook Newsroom, 15 juin 2017 ; Twitter Safety, « Enforcing new rules to reduce hateful conduct and abusive behavior », 18 décembre 2017.

[120] Voir, par exemple, Règles YouTube (Règles concernant les contenus visuels choquants ou violents).

[121] Angwin et Grasseger, « Facebook's secret censorship rules ».

[122] Communications de : Ranking Digital Rights, p. 10 ; OBSERVACOM, p. 10 ; Association pour le progrès des communications, p. 17 ; Fédération internationale des associations de bibliothécaires et des bibliothèques, p. 4 et 5, Access Now, p. 17 ; et EDRi, p. 5.

[123] Kara Swisher et Kurt Wagner, « Here's the transcript of Recode's interview with Facebook CEO Mark Zuckerberg about the Cambridge Analytica controversy and more », Recode, 22 mars 2018.

l'entreprise et reposent sur des déclarations vagues concernant l'existence d'une communauté d'intérêts, ont créé des environnements instables, imprévisibles et peu sûrs pour les utilisateurs, dans lesquels les pouvoirs publics exercent une surveillance accrue. Les lois nationales ne sont pas adaptées aux entreprises qui souhaitent s'appuyer sur des normes communes applicables à leur communauté d'utilisateurs, dont la diversité est importante du point de vue géographique et culturel. Or, lorsqu'elles sont appliquées de manière transparente et cohérente et que les utilisateurs et la société civile contribuent véritablement à leur mise en œuvre, les normes relatives aux droits de l'homme offrent un cadre permettant de demander aux États ainsi qu'aux entreprises de rendre des comptes aux utilisateurs indépendamment des frontières nationales.

42. Le cadre des droits de l'homme met à disposition des moyens normatifs solides de contester le bien-fondé de restrictions abusives imposées par l'État − à condition que les entreprises appliquent des règles analogues. Les Principes directeurs et l'ensemble de textes juridiques non contraignants qui les accompagnent donnent des orientations sur la façon dont les entreprises peuvent prévenir les demandes de suppression de contenus excessives émanant des pouvoirs publics ou en réduire la portée. En outre, ces textes consacrent des principes de diligence raisonnable, de transparence, de responsabilité et de réparation qui limitent les atteintes aux droits de l'homme par les plateformes grâce à l'élaboration de produits et de politiques. Les entreprises qui se seront engagées à respecter les normes relatives aux droits de l'homme dans toutes leurs activités − et pas uniquement lorsqu'elles coïncident avec leurs intérêts − se sentiront plus inattaquables lorsqu'elles demanderont aux États de rendre des comptes sur la façon dont ils respectent ces mêmes normes. En outre, lorsque les entreprises auront harmonisé leurs conditions d'utilisation avec le droit des droits de l'homme, les États auront davantage de difficultés à se servir d'elles pour censurer des contenus.

43. Les principes relatifs aux droits de l'homme permettent en outre aux entreprises de créer un environnement inclusif qui tienne compte des besoins et des intérêts divers des utilisateurs tout en établissant des normes fondamentales de comportement qui soient prévisibles et cohérentes. Face au débat grandissant sur la question de savoir si les entreprises jouent à la fois le rôle d'intermédiaire et celui de rédacteur de contenus, le droit des droits de l'homme promet aux utilisateurs qu'ils peuvent compter sur les normes fondamentales pour protéger leur liberté d'expression dans une mesure allant bien au-delà des restrictions qui pourraient être prévues par les législations nationales[124]. Toutefois, le droit des droits de l'homme n'est pas rigide ou dogmatique au point d'exiger des entreprises qu'elles autorisent la publication de propos portant atteinte aux droits d'autres personnes ou nuisant à la capacité des États à garantir des intérêts légitimes liés à la sécurité nationale ou à l'ordre public. Compte tenu de la quantité d'actes malveillants susceptibles d'avoir des effets plus marqués dans l'espace numérique qu'ils ne pourraient en avoir hors ligne − comme le harcèlement misogyne ou homophobe, qui vise à réduire les femmes et les minorités sexuelles au silence, ou l'incitation à la violence sous toutes ses formes − le droit des droits de l'homme ne risque pas de priver les entreprises de moyens d'action. Au contraire, il leur offre un cadre reconnu mondialement qui leur permet de mettre au point ces moyens d'action et de trouver un langage commun pour définir leur nature, leur finalité et la façon de les appliquer aux utilisateurs et aux États.

A. Normes fondamentales applicables à la modération des contenus

44. L'ère du numérique se caractérise certes par la possibilité de diffuser rapidement un contenu et d'atteindre un nombre considérable de personnes, mais elle se distingue aussi par l'absence de prise en considération du contexte humain. D'après les Principes directeurs, les entreprises peuvent tenir compte de la taille, de la structure et des fonctions particulières des plateformes qu'elles offrent lorsqu'elles apprécient la nécessité et la proportionnalité des restrictions de contenu.

45. *Droits de l'homme par défaut*. Les conditions d'utilisation ne devraient pas reposer sur une approche discrétionnaire fondée sur les besoins génériques et les intérêts égoïstes de

[124] Communication de Global Partners Digital, p. 3 ; Principes directeurs, principe 11.

la « communauté ». Les entreprises devraient plutôt prendre des engagements politiques de haut niveau visant à laisser les plateformes à la disposition des utilisateurs pour qu'ils exposent leurs opinions, s'expriment librement et accèdent à toutes sortes d'informations d'une manière compatible avec le droit des droits de l'homme[125]. Ces engagements devraient être le fondement de leur approche en matière de modération de contenus et de gestion des problèmes complexes tels que la propagande informatique[126] et la collecte et le traitement des données des utilisateurs. Les entreprises devraient faire directement figurer dans leurs conditions d'utilisation et dans leurs « normes applicables à la communauté » les principes pertinents du droit des droits de l'homme selon lesquels les mesures ayant des incidences sur les contenus doivent être conformes aux principes de légalité, de nécessité et de légitimité par lesquels les États sont liés lorsqu'ils réglementent la liberté d'expression[127].

46. *« Légalité ».* Les règles des entreprises sont généralement dépourvues de la clarté et de la précision qui permettrait aux utilisateurs de prédire avec un degré de certitude raisonnable quel type de contenu leur ferait franchir la ligne rouge. Cette lacune est particulièrement évidente dans le contexte de la publication de contenus extrémistes et de discours de haine, domaines soumis à des restrictions qui risquent facilement de déboucher sur des suppressions abusives si une évaluation rigoureuse du contexte n'est pas effectuée par un être humain. La compréhension par le public des règles propres à un contexte est rendue encore plus difficile par l'émergence d'une nouvelle exception générale, « l'intérêt de l'information »[128]. La reconnaissance de l'intérêt public est certes bienvenue, mais les entreprises devraient préciser quels facteurs sont pris en considération dans la détermination de cette notion et quels facteurs autres que l'intérêt public sont utilisés comme référence pour évaluer l'intérêt potentiel d'une information. Les entreprises devraient compléter les explications détaillées qu'elles donnent pour mieux faire comprendre leurs règles par des données agrégées illustrant l'évolution de l'application des règles ainsi que par des exemples de cas concrets ou de cas hypothétiques complets et précis mettant en évidence les nuances existant dans l'interprétation et l'application de certaines règles.

47. *Nécessité et proportionnalité.* Les entreprises ne devraient pas se borner à décrire plus en détail les règles controversées et les règles propres à un contexte particulier ; elles devraient aussi fournir des données et des exemples permettant de se faire une idée des éléments pris en considération dans la détermination d'une violation, de sa gravité et des mesures prises en conséquence. S'agissant des discours de haine, expliquer la façon dont certains cas ont été réglés peut permettre aux utilisateurs de mieux comprendre comment les entreprises abordent la question de la distinction délicate entre les contenus offensants et l'incitation à la haine, ou comment des éléments tels que l'intention de l'auteur des propos ou le risque de violence sont appréciés lorsqu'il s'agit de contenus diffusés en ligne. Des données détaillées sur les mesures prises pourraient aussi servir à déterminer si les entreprises adaptent finement leurs restrictions. Il conviendrait en outre d'expliquer dans quelles circonstances les entreprises appliquent des restrictions moins intrusives (avertissements, limites d'âge ou démonétisation).

48. *Non-discrimination.* Pour se doter de garanties de non-discrimination dignes de ce nom, les entreprises doivent dépassent les approches formalistes selon lesquelles toutes les caractéristiques protégées sont également exposées aux violences, au harcèlement et à diverses formes de censure[129]. En effet, ces approches paraîtraient incompatibles avec l'accent que les entreprises mettent sur le contexte. En fait, lorsqu'elles mettent au point ou modifient des politiques ou des produits, les entreprises devraient chercher activement à comprendre les préoccupations des communautés qui ont toujours été exposées à des risques de censure et de discrimination et en tenir compte.

[125] Principes directeurs, principe 16.

[126] Voir Samuel Wooley et Philip Howard, *Computational Propaganda Worldwide: Executive Summary* (Computational Propaganda Research Project working paper No. 2017.11 (Oxford, 2017)).

[127] Communication de Global Partners Digital, p. 10 à 13.

[128] Voir Joel Kaplan, « Input from community and partners on our community standards », service de presse de Facebook, 21 octobre 2016 ; règles et politiques de Twitter.

[129] Voir notamment la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, art. 1 (par. 4) et 2 (par. 2).

B. Procédures de modération par l'entreprise et activités connexes

Réponses aux demandes émanant de gouvernements

49. Comme le montrent les rapports de transparence des entreprises, les gouvernements exercent des pressions sur les entreprises pour les convaincre de supprimer des contenus, de suspendre des comptes et de recueillir et divulguer des informations sur des comptes. Lorsque de telles demandes sont formulées en vertu de la législation interne d'un pays, les entreprises n'ont manifestement pas d'autre choix que de s'y plier. Toutefois, elles peuvent mettre au point des outils permettant de prévenir ou d'atténuer les risques d'atteintes aux droits de l'homme que peuvent entraîner des législations nationales ou des exigences incompatibles avec les normes internationales.

50. *Prévention et atténuation.* Les entreprises affirment souvent qu'elles accordent une grande importance aux droits de l'homme. Cependant, elles ne devraient pas se contenter de prendre ce type d'engagement au plan interne et de donner ponctuellement des assurances au public lorsque des controverses éclatent. Elles devraient aussi, au plus haut niveau de leur direction, adopter publiquement des règlements prévoyant que toutes leurs unités, y compris les filiales locales, doivent faire en sorte que toute ambiguïté dans la loi soit interprétée en faveur de la liberté d'expression et du respect du droit à la vie privée et des autres droits de l'homme. Les politiques et procédures qui interprètent et mettent en œuvre les demandes de gouvernements engageant les entreprises à définir strictement et à « limiter au minimum les restrictions appliquées aux contenus » devraient découler de ces engagements[130]. Les entreprises devraient s'assurer que ces demandes sont formulées par écrit, qu'elles citent à l'appui des dispositions juridiques précises et valables et qu'elles sont émises en bonne et due forme par une autorité publique habilitée à le faire[131].

51. Lorsqu'elles reçoivent des demandes problématiques, les entreprises devraient : demander des éclaircissements ou des modifications ; solliciter l'aide de la société civile, des autres entreprises du secteur, des organes publics compétents, d'organismes internationaux ou régionaux et d'autres parties prenantes ; et étudier tous les moyens que la loi met à leur disposition pour contester ces demandes[132]. Lorsque les entreprises reçoivent des demandes d'États au titre de leurs conditions d'utilisation ou par d'autres moyens extrajudiciaires, elles devraient soumettre ces demandes à des procédures de vérification de la légalité et en apprécier la validité au regard de la législation interne pertinente et des normes relatives aux droits de l'homme.

52. *Transparence.* Face à la censure et aux risques d'atteintes aux droits de l'homme qui y sont associés, les utilisateurs ne peuvent prendre de décisions éclairées sur l'opportunité d'être actifs sur les réseaux sociaux et sur la façon dont ils peuvent communiquer par ce moyen que si les interactions entre les entreprises et les États sont véritablement transparentes. Des pratiques optimales sur la manière de garantir une réelle transparence devraient être définies. Les informations publiées par les entreprises sur les demandes émanant d'États devraient être accompagnées de précisions sur le type de problème soulevé (diffamation, incitation à la haine, contenus liés au terrorisme) et les mesures prises (suppression partielle ou complète, suppression pour le pays concerné ou à l'échelon mondial, suspension du compte, demande de suppression accordée au titre des conditions d'utilisation). En outre, les entreprises devraient donner des exemples concrets aussi souvent que possible[133]. Les rapports de transparence devraient porter également sur les demandes de gouvernements fondées sur les conditions d'utilisation[134] et rendre compte des initiatives conjointes des secteurs public et privé visant à limiter la diffusion de contenus,

[130] Voir A/HRC/35/22, par. 66 et 67.

[131] Communications de Global Network Initiative, p. 3 et 4, et de GitHub p. 3 à 5.

[132] Voir A/HRC/35/22, par. 68.

[133] Voir notamment *Twitter Transparency Report: Removal Requests* (janvier-juin 2017).

[134] Twitter a commencé à publier des données sur des demandes non judiciaires soumises par de célèbres représentants de gouvernements concernant des contenus susceptibles de violer les règles de Twitter interdisant les comportements répréhensibles, la promotion du terrorisme et les atteintes aux droits de propriété intellectuelle. Ibid. Voir aussi Microsoft, Content Removal Requests Report (janvier-juin 2017).

dont le Code de conduite de l'Union européenne visant à combattre les discours de haine illégaux en ligne, les initiatives gouvernementales telles que les services de signalement sur Internet et les accords bilatéraux tels que ceux qui auraient été conclus entre YouTube et le Pakistan et entre Facebook et Israël. Les entreprises devraient conserver des copies des demandes formulées au titre de ces initiatives ainsi que des échanges entre l'entreprise et l'auteur de la demande, et étudier les possibilités de confier des exemplaires de ces demandes à un tiers qui en serait le dépositaire.

Définitions de règles et élaboration de produits

53. *Devoir de diligence*. Bien que plusieurs entreprises s'engagent à faire preuve de diligence raisonnable en matière de droits de l'homme dans le cadre de l'évaluation de leurs réponses aux demandes de restrictions émanant d'États, il est difficile de dire si elles prennent les mêmes précautions pour prévenir ou atténuer les risques d'atteintes à la liberté d'expression engendrés par l'élaboration et l'application de leurs propres politiques[135]. Les entreprises devraient définir des critères clairs et précis pour recenser les activités qui appellent de telles évaluations. Il conviendrait non seulement de revoir les politiques et les procédures de modération des contenus, mais aussi de procéder à des évaluations concernant l'édition des flux d'utilisateurs et autres types de diffusion de contenus, l'introduction de nouvelles caractéristiques ou de nouveaux services et la modification des caractéristiques et services existants, et l'élaboration de technologies d'automatisation et de décisions de mise sur le marché telles que les arrangements concernant la création de versions de la plateforme spécifiquement conçues pour certains pays[136]. Les rapports antérieurs permettent en outre de définir les questions sur lesquelles ces évaluations devraient porter et les procédures et la formation internes nécessaires pour intégrer les évaluations et leurs conclusions dans les activités pertinentes. De plus, ces évaluations devraient être continues et s'adapter à l'évolution des circonstances ou au contexte dans lequel les entreprises exercent leurs activités[137]. Les initiatives regroupant diverses catégories de parties prenantes, dont la Global Network Initiative, permettent aux entreprises d'élaborer et d'affiner des mécanismes d'évaluation et d'autres procédures de diligence raisonnable.

54. *Participation et mobilisation du public*. Les participants aux consultations ont systématiquement exprimé des préoccupations concernant le fait que les entreprises ne collaboraient pas de manière satisfaisante avec les utilisateurs et la société civile, en particulier dans les pays du Sud. L'apport des titulaires de droits touchés (ou de leurs représentants) et des experts locaux ou des spécialistes du domaine, ainsi que l'existence de processus décisionnels internes prenant véritablement en considération les retours d'information font partie intégrante de la diligence raisonnable[138]. Les consultations − en particulier celles de grande ampleur telles que les appels à commentaires adressés au public − permettent aux entreprises d'envisager les incidences de leurs activités sur les droits de l'homme selon différents angles de vue, tout en les encourageant à examiner de plus près la façon dont des règles apparemment inoffensives ou « respectueuses de la communauté » peuvent avoir d'importantes incidences « hyperlocales » sur certaines collectivités[139]. Par exemple, le dialogue avec des groupes autochtones issus de zones géographiques très diverses peut aider les entreprises à élaborer de meilleurs indicateurs permettant de tenir compte du contexte culturel et artistique pour l'évaluation de contenus où la nudité est présente.

55. *Transparence de l'élaboration des règles*. Trop souvent, les entreprises introduisent des produits et modifient des règles sans vérifier préalablement leur conformité aux droits de l'homme ni évaluer leur impact dans des cas concrets. Elles devraient au moins

[135] Communication de Ranking Digital Rights, p. 12 ; Principes directeurs, principe 17.

[136] Voir A/HRC/35/22, par. 53.

[137] Ibid., par. 54 à 58.

[138] Voir les Principes directeurs, principe 18, et le document A/HRC/35/22, par. 57.

[139] Chinmayi Arun, « Rebalancing regulation of speech: hyper-local content on global web-based platforms », Berkman Klein Center for Internet and Society Medium Collection, Harvard University, 2018; *Pretoria News*, "Protest at Google, Facebook 'bullying' of bare-breasted maidens", 14 décembre 2017.

demander aux utilisateurs et aux experts concernés de commenter leurs évaluations d'impact, et ce, dans des contextes qui garantissent la confidentialité de ces évaluations, si besoin est. Elles devraient aussi donner des explications claires au public sur les règles et les procédures appliquées dans le cadre de ces évaluations.

Application des règles

56. *Automatisation et évaluation par l'homme*. La modération automatisée des contenus, liée à l'ampleur et à la portée considérables des contenus générés par les utilisateurs, pose des problèmes distincts, à savoir le risque que des mesures concernant les contenus soient incompatibles avec le droit des droits de l'homme. Dans le cadre de la responsabilité qui leur incombe de prévenir et d'atténuer l'impact sur les droits de l'homme, les entreprises devraient tenir compte des limites importantes de l'automatisation, qui sont notamment liées à la difficulté de comprendre le contexte, aux variations considérables des expressions et des significations, ainsi qu'aux particularités linguistiques et culturelles. L'automatisation fondée sur des notions créées dans le pays d'origine de l'entreprise risque d'être source de graves discriminations à l'égard des communautés d'utilisateurs des autres pays. Au minimum, les technologies mises au point pour prendre en compte les questions d'échelle devraient être soumises à des vérifications rigoureuses et élaborées dans le cadre d'une large consultation des usagers et de la société civile.

57. La responsabilité de promouvoir des pratiques en matière de modération qui soient bien conçues et adaptées au contexte et qui respectent la liberté d'expression suppose que les entreprises renforcent et garantissent la professionnalisation de leurs procédures d'évaluation par l'homme des contenus signalés. À cette fin, les entreprises devraient faire bénéficier les modérateurs de garanties compatibles avec les normes relatives aux droits de l'homme applicables aux droits du travail et s'engager fermement à faire appel à des experts ayant des compétences culturelles, linguistiques et autres dans tous les marchés sur lesquels elles sont présentes. La direction des entreprises et les équipes stratégiques devraient aussi être diversifiées afin que des compétences locales ou spécifiques à une question puissent être utilisées lorsque des problèmes de contenu se posent.

58. *Notification et recours.* Les utilisateurs et les experts issus de la société civile expriment souvent des préoccupations quant aux informations limitées dont disposent les personnes dont les publications ont été supprimées ou dont le compte a été suspendu ou désactivé, ou celles qui signalent des actes répréhensibles tels que le harcèlement à motivation misogyne ou le « *doxing* ». Le manque d'informations crée un contexte dans lequel les normes sont secrètes, qui est contraire aux principes de clarté, de spécificité et de prévisibilité et qui nuit à la capacité des particuliers de contester les mesures prises à l'égard de contenus ou de connaître la suite donnée aux plaintes se rapportant aux contenus. Dans la pratique, toutefois, l'absence de mécanismes de recours solides contre la suppression de contenus favorise les utilisateurs qui font des signalements. D'aucuns objecteront qu'il serait très coûteux en temps et en argent de permettre que chaque mesure prise à l'égard d'un contenu fasse l'objet d'un recours. Toutefois, les entreprises pourraient collaborer entre elles et avec la société civile pour réfléchir à des solutions modulables telles que des programmes de médiation conçus pour une entreprise en particulier ou pour l'ensemble du secteur. L'une des meilleures idées à cet égard consiste dans la mise en place d'un « conseil des médias sociaux » indépendant, établi sur le modèle des conseils de la presse, lesquels prévoient des mécanismes de plainte concernant l'ensemble des médias et la promotion de réparation en cas de violation[140]. Ce mécanisme pourrait examiner les plaintes émanant d'utilisateurs qui remplissent certains critères et recueillir l'avis du public sur les problèmes récurrents posés par la modération des contenus tels que la censure abusive dans certains domaines. Les États devraient soutenir les mécanismes de plainte modulables qui mènent leurs activités dans le respect des normes relatives aux droits de l'homme.

59. *Réparation.* Les Principes directeurs mettent l'accent sur la responsabilité de remédier aux « incidences néfastes » (principe 22). Toutefois, très peu d'entreprises offrent

[140] Voir ARTICLE 19, *Self-regulation and 'Hate Speech' on Social Media Platforms* (London, 2018), p. 20 à 22.

des mesures de réparation. Les entreprises devraient se doter de programmes solidement structurés prévoyant des mesures de réparation allant du rétablissement du contenu et de la reconnaissance des torts aux règlements des litiges relatifs aux atteintes à la réputation ou à d'autres préjudices. Il existe une certaine convergence de vues entre plusieurs entreprises en ce qui concerne les règles relatives aux contenus, ce qui ouvre la perspective d'une coopération entre entreprises visant à offrir des réparations dans le cadre d'un conseil des médias sociaux, de programmes de médiation ou d'arbitrages. Si les préjudices ne peuvent pas être réparés par ces moyens, des mesures législatives et judiciaires pourront être nécessaires.

60. *Autonomie des utilisateurs.* Les entreprises ont mis au point des outils permettant aux utilisateurs de façonner leur propre environnement en ligne et qui incluent le blocage automatique et le blocage d'autres utilisateurs ou de certains types de contenu. De même, les plateformes permettent souvent aux utilisateurs de créer des groupes fermés ou privés, modérés par les utilisateurs eux-mêmes. Les règles applicables aux contenus publiés par les membres de groupes fermés devraient certes être compatibles avec les normes fondamentales relatives aux droits de l'homme, mais ces groupes, qui sont formés sur la base d'affinités, devraient être encouragés par les plateformes car ils contribuent à protéger la liberté d'opinion, élargissent les possibilités des communautés vulnérables de s'exprimer et sont un lieu où les idées controversées ou impopulaires peuvent être mises à l'épreuve. Il conviendrait de supprimer les clauses exigeant des utilisateurs qu'ils s'inscrivent sous leur nom réel, compte tenu des incidences que cela peut avoir sur la vie privée et sur la sécurité des personnes vulnérables[141].

61. Les préoccupations grandissantes exprimées au sujet des possibilités de vérification, de la pertinence et de l'utilité des informations publiées en ligne soulèvent des questions complexes concernant la manière dont les entreprises devraient respecter le droit d'accès à l'information. Elles devraient à tout le moins expliquer leurs méthodes en matière d'édition de contenus. Si les entreprises classent les contenus des flux des médias sociaux en fonction des interactions entre utilisateurs, elles devraient donner des explications sur les données recueillies sur ces interactions et sur la manière dont les données sont utilisées pour définir les critères de classement. Les entreprises devraient offrir à tous les utilisateurs des possibilités accessibles et réelles de refuser que leurs contenus soient édités par la plateforme[142].

Transparence dans la prise de décisions

62. Malgré les progrès réalisés en ce qui concerne la transparence globale des demandes de suppression de contenus émanant de gouvernements, une très grande partie des mesures prises au titre des conditions d'utilisation ne sont pas signalées. Les entreprises ne publient pas de données sur le volume et le type de demandes émanant de particuliers qu'elles reçoivent au titre des conditions d'utilisation, sans parler des taux d'application. Les entreprises devraient élaborer des initiatives de promotion de la transparence contenant des explications sur les incidences de l'automatisation, de la modération par l'homme et du signalement par les utilisateurs ou par des entités de confiance sur les mesures prises au titre des conditions d'utilisation. Quelques entreprises commencent à fournir des informations sur ces mesures, mais le secteur devrait s'efforcer de donner davantage de précisions sur des cas précis et représentatifs et sur les évolutions importantes dans l'interprétation et l'application de leurs politiques.

63. Les entreprises appliquent le « droit des plateformes » et prennent des mesures pour régler les problèmes liés à des contenus sans communiquer beaucoup d'informations à ce sujet. Dans l'idéal, elles devraient établir une forme de jurisprudence, ce qui permettrait aux utilisateurs, à la société civile et aux États de comprendre comment elles interprètent et appliquent leurs normes. Dans un tel système « jurisprudentiel », on ne disposerait pas d'un

[141] Voir par. 30.

[142] Par exemple, Facebook permet à ses utilisateurs de faire apparaître les publications sur leur fil d'actualités dans l'ordre chronologique inverse, mais il les avertit que les paramètres d'édition par défaut « finiront par être » rétablis. Espace d'assistance de Facebook, « Quelle est la différence entre les actualités les plus récentes et les actualités à la une dans le fil d'actualité ? ».

compte rendu des décisions tel que celui que le public attend des tribunaux et des organes administratifs, mais d'une compilation détaillée de cas et d'exemples, ce qui permettrait de clarifier les règles aussi efficacement que les comptes rendus de décisions judiciaires[143]. Un mécanisme crédible et indépendant tel qu'un conseil des médias sociaux habilité à examiner les plaintes concernant le secteur des TIC dans son ensemble pourrait contribuer à garantir la transparence dans ce domaine.

V. Recommandations

64. **Des forces opaques influent sur la capacité des individus du monde entier à exercer leur liberté d'expression. En l'état actuel des choses, une transparence absolue, une véritable responsabilisation et la volonté de réparer les préjudices sont indispensables pour protéger la capacité des particuliers à utiliser les plateformes en ligne pour s'exprimer librement, accéder à l'information et participer à la vie publique. On trouvera ci-après une série de mesures qui pourraient être prises à cette fin.**

Recommandations destinées aux États

65. **Les États devraient abroger toute loi incriminant ou limitant indûment l'expression, en ligne ou hors ligne.**

66. **Une réglementation judicieusement conçue, plutôt qu'une réglementation lourde fondée sur un point de vue, devrait être la norme. Cette réglementation devrait avoir pour objectif de garantir la transparence de l'entreprise et prévoir des mesures de réparation afin que le public puisse choisir la manière dont il souhaite participer à des forums en ligne et décider d'y participer ou non. Les États ne devraient 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 légitimité. Les États devraient s'abstenir d'imposer des sanctions disproportionnées − lourdes amendes ou peines d'emprisonnement − aux intermédiaires Internet, compte tenu de leur effet dissuasif sur l'exercice de la liberté d'expression.**

67. **Les États et les organisations intergouvernementales devraient s'abstenir d'élaborer des lois ou des accords prévoyant l'obligation de contrôler « en amont » ou de filtrer des contenus, ce qui est contraire au droit au respect de la vie privée et pourrait constituer une forme de censure préalable à la publication.**

68. **Les États devraient s'abstenir d'adopter des réglementations investissant des organismes publics, plutôt que les autorités judiciaires, du rôle d'arbitre de la légalité d'une forme d'expression. Ils devraient éviter de déléguer aux entreprises la responsabilité de prendre des décisions sur des contenus car, dans ce cas, le point de vue des entreprises l'emporte sur les principes relatifs aux droits de l'homme, au détriment des utilisateurs.**

69. **Les États devraient publier des rapports de transparence détaillés sur toutes les demandes relatives à des contenus adressées à des intermédiaires et tenir véritablement compte des contributions du public pour toutes les questions relatives à la réglementation.**

Recommandations aux entreprises du secteur des TCI

70. **Les entreprises devraient reconnaître que la norme mondiale qui fait autorité pour ce qui est de garantir la liberté d'expression sur leurs plateformes est le droit des droits de l'homme et non les lois nationales, qui varient d'un pays à l'autre, ou leurs intérêts propres, et elles devraient réexaminer leurs normes relatives aux contenus en conséquence. Le droit des droits de l'homme donne des moyens aux entreprises de**

[143] Voir notamment Madeleine Varner *et al.*, « What does Facebook consider hate speech ? », ProPublica, 28 décembre 2017.

définir et d'élaborer des politiques et des procédures qui sont conformes aux principes démocratiques et qui résistent aux exigences autoritaires. Selon cette approche, l'entreprise commence par établir des règles fondées sur les droits, puis réalise des études d'impact rigoureuses sur les droits de l'homme en vue de l'élaboration de produits et de politiques, et enfin lance ses activités, en procédant régulièrement à des évaluations et à des réévaluations et en consultant véritablement le public et la société civile. Les Principes directeurs relatifs aux entreprises et aux droits de l'homme ainsi que des directives sectorielles élaborées par la société civile, des organes intergouvernementaux, la Global Network Initiative et d'autres acteurs énoncent des principes de référence que toutes les sociétés Internet devraient adopter.

71. **Les entreprises doivent suivre des approches radicalement différentes en matière de transparence à toutes les étapes de leurs activités, depuis l'élaboration et l'application des règles jusqu'au développement d'une « jurisprudence » offrant un cadre pour l'interprétation des règles privées. La transparence nécessite une collaboration plus étroite avec les organisations spécialisées dans le droit numérique et les autres secteurs concernés de la société civile et suppose l'absence d'accords secrets avec des États sur des normes relatives aux contenus et leur mise en œuvre.**

72. **Étant donné leur impact sur la sphère publique, les entreprises doivent accepter de rendre des comptes. Dans le monde entier, des conseils de la presse efficaces et respectueux des droits sont une référence pour ce qui est d'instaurer un minimum de cohérence, de transparence et de responsabilité en matière de modération des contenus commerciaux. Les approches d'autres acteurs non gouvernementaux, si elles sont fondées sur les droits de l'homme, pourraient offrir des mécanismes de recours et de réparation sans que cela ait un coût prohibitif susceptible de décourager les petites entités ou les nouveaux venus sur le marché. Tous les acteurs du secteur des TIC qui modèrent des contenus ou jouent le rôle de contrôleurs d'accès devraient ériger en priorité absolue la mise en place de mécanismes de responsabilité pour l'ensemble du secteur (tels qu'un conseil des médias sociaux).**