



Assemblée générale

Distr. générale
9 octobre 2019
Français
Original : anglais

Soixante-quatorzième session

Point 70 b) de l'ordre du jour

Promotion et protection des droits de l'homme :
questions relatives aux droits de l'homme, y compris
les divers moyens de mieux assurer l'exercice effectif
des droits de l'homme et des libertés fondamentales

Promotion et protection du droit à la liberté d'opinion et d'expression*

Note du Secrétaire général

Le Secrétaire général a l'honneur de transmettre à l'Assemblée générale le rapport établi par le Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, David Kaye, soumis en application de la résolution [34/18](#) du Conseil des droits de l'homme. Dans le présent rapport, le Rapporteur spécial examine le droit des droits de l'homme qui s'applique à la réglementation du « discours haineux » en ligne.

* Le présent rapport a été soumis après la date limite afin que puissent y figurer les informations les plus récentes.



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

Résumé

Dans un monde où de plus en plus de voix s'élèvent pour exiger une limite aux discours haineux, le droit international des droits de l'homme établit des normes applicables à l'action menée par les États et les entreprises pour réglementer la liberté d'expression en ligne. Dans le présent rapport, soumis en application de la résolution [34/18](#) du Conseil des droits de l'homme, le Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression explique en quoi ces normes peuvent servir aux gouvernements pour rechercher des solutions en matière de réglementation et aux entreprises qui entendent assurer le respect des droits de l'homme en ligne. Le Rapporteur spécial fait d'abord une introduction consacrée au cadre juridique international, en mettant l'accent sur les traités des Nations Unies et les principales interprétations des dispositions relatives à la notion communément appelée « discours haineux ». Puis, il met en lumière les principales obligations des États et examine la manière dont les entreprises peuvent garantir le respect des droits de l'homme des utilisateurs et du public par la modération de leurs contenus. Pour conclure, il formule des recommandations à l'intention des États et des entreprises.

Table des matières

	<i>Page</i>
I. Introduction	4
II. La notion de discours haineux et sa réglementation en droit international des droits de l'homme	5
III. Réglementation des discours haineux en ligne	14
A. Obligations des États et réglementation des discours haineux en ligne	14
B. Modération des contenus par les entreprises et discours haineux	18
IV. Conclusions et recommandations	25

I. Introduction

1. L'expression « discours haineux » – formule qui n'a pas de définition en droit international conventionnel – est doublement ambiguë. Peu précise, elle ne fait pas l'objet de consensus quant à sa signification, si bien qu'elle peut être employée à tort pour attenter à toute une série de modes d'expression pourtant autorisés par la loi. De nombreux gouvernements invoquent le motif du « discours haineux », de la même manière qu'ils emploient le terme « infox », pour s'en prendre à leurs adversaires politiques, aux non-croyants, aux dissidents et à leurs détracteurs. Toutefois, l'expression, faiblement connotée (« ce ne sont que des mots »), semble par ailleurs peu utile lorsqu'il s'agit, pour les gouvernements et les entreprises, de s'attaquer aux véritables maux, tels que ceux découlant des discours qui incitent à la violence ou à la discrimination contre les personnes vulnérables, ou qui visent à réduire au silence les personnes marginalisées. Cette situation explique la frustration d'un public qui a, souvent, le sentiment d'être la cible de vexations généralisées sur Internet.

2. Dans un monde où de plus en plus de voix s'élèvent pour exiger une limite aux « discours haineux », le droit international des droits de l'homme établit des normes applicables à l'action menée par les États et les entreprises pour réglementer la liberté d'expression en ligne (A/HRC/38/35, par. 45)¹. Dans le présent rapport, le Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression explique en quoi ces normes peuvent servir aux gouvernements qui recherchent des solutions en matière de réglementation et aux entreprises qui entendent d'assurer le respect des droits de l'homme en ligne. Le Rapporteur spécial fait d'abord une introduction consacrée au cadre juridique international, en mettant l'accent sur les traités des Nations Unies et les principales interprétations des dispositions relatives à la notion communément appelée « discours haineux ». Puis, il met en lumière les principales obligations des États et examine la manière dont les entreprises peuvent garantir le respect des droits de l'homme des utilisateurs et du public par la modération de leurs contenus. Pour conclure, il formule des recommandations à l'intention des États et des entreprises.

3. Le présent rapport est le sixième d'une série, publiée depuis 2015, de rapports dans lesquels le Rapporteur spécial traite des normes relatives aux droits de l'homme applicables à la liberté d'opinion et d'expression dans le secteur des technologies de l'information et des communications (TIC)². Il doit être lu à l'aune des normes et recommandations proposées dans les rapports précédents, qui ne sont pas nécessairement reproduites ici. Comme dans les rapports précédents, le Rapporteur spécial s'inspire largement des normes internationales existantes et de l'importante contribution apportée en la matière par la société civile au cours de ces dernières années.

¹ Dans le présent rapport, l'expression « discours haineux » renvoie aux obligations et aux restrictions prévues par le droit des droits de l'homme où elle n'est pas expressément employée. Voir Susan Benesch, « *Proposals for improved regulation of harmful online content* », document établi pour Israel Democracy Institute, 2019. Benesch est l'auteure de l'expression apparentée – « discours dangereux » –, qui sert à désigner la « capacité d'inciter un groupe à commettre des actes de violence contre un autre groupe ». Voir également Susan Benesch, « *Dangerous speech : a proposal to prevent group violence* », 2012.

² Voir A/HRC/29/32 (chiffrement et anonymat), A/HRC/32/38 (incidences du secteur des TIC sur l'exercice des droits), A/HRC/35/22 (secteur de l'accès numérique), A/HRC/38/35 (modération des contenus en ligne) et A/73/348 (intelligence artificielle et droits de l'homme).

II. La notion de discours haineux et sa réglementation en droit international des droits de l'homme

4. En droit international des droits de l'homme, il apparaît nécessaire, pour imposer une limite aux « discours haineux », de concilier deux principes distincts, à savoir l'impératif, pour toute société démocratique, de faire une place au débat public ainsi qu'à l'autonomie et à l'épanouissement de l'individu, et l'obligation, tout aussi impérieuse, d'éviter que les populations vulnérables ne soient prises pour cible et de garantir une participation égale et non discriminatoire de tous et de toutes à la vie publique³. Les gouvernements exploitent souvent l'imprécision de ce cadre pour réprimer certaines formes d'expression légitime, comme la dissidence et la contestation politiques ou les divergences religieuses⁴. Or, la liberté d'expression, le droit à l'égalité, le droit à la vie et l'obligation de non-discrimination se renforcent mutuellement : le droit des droits de l'homme permet aux États et aux entreprises de privilégier la protection et la promotion de la liberté d'expression de toutes les personnes, en particulier de celles dont les droits sont souvent menacés, tout en s'attaquant à la discrimination dans les secteurs public et privé, qui constitue un obstacle à l'exercice de tous les droits.

Liberté d'expression

5. Le paragraphe 1 de l'article 19 du Pacte international relatif aux droits civils et politiques protège le droit de ne pas être inquiété pour ses opinions et le paragraphe 2 du même article garantit le droit à la liberté d'expression, qui comprend la liberté 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. De nombreux autres traités, internationaux ou régionaux, protègent expressément la liberté d'expression⁵. Le Comité des droits de l'homme, qui est l'organe composé d'experts chargés de surveiller l'application du Pacte, a souligné que ces libertés « sont des conditions indispensables au développement complet de l'individu [...] [et] constituent le fondement de toute société libre et démocratique ». Elles « constituent la base de l'exercice sans réserve d'un grand nombre d'autres droits de l'homme »⁶.

6. La liberté d'expression étant fondamentale pour l'exercice de tous les droits de la personne, les restrictions qui lui sont imposées doivent être exceptionnelles, assujetties à des conditions bien précises et soumises à un contrôle strict. Le Comité des droits de l'homme a souligné que les restrictions, même lorsqu'elles sont justifiées, « ne peuvent pas compromettre le droit lui-même »⁷. Le caractère exceptionnel des restrictions est souligné au paragraphe 3 de l'article 19 du Pacte, où

³ Voir, en particulier, le rapport de l'ancien Rapporteur spécial, Frank La Rue, sur le discours haineux (A/67/357).

⁴ Ibid., par. 51 à 54.

⁵ Voir, par exemple, la Convention internationale sur l'élimination de toutes les formes de discrimination raciale (art. 5) ; la Convention relative aux droits de l'enfant (art. 13) ; la Convention relative aux droits des personnes handicapées (art. 21) ; la Convention internationale sur la protection des droits de tous les travailleurs migrants et des membres de leur famille (art. 13) ; la Convention américaine relative aux droits de l'homme (art. 13) ; la Charte africaine des droits de l'homme et des peuples (art. 9) ; la Convention européenne des droits de l'homme (art. 10).

⁶ Comité des droits de l'homme, observation générale n° 34 (2011) relative à la liberté d'opinion et d'expression, par. 2 et 4 ; voir également *ibid.*, par. 5 et 6.

⁷ Ibid., par. 21. Le Comité des droits de l'homme a précisé que « les restrictions ne doivent pas porter atteinte à l'essence même du droit » et que « [l]es lois autorisant l'application de restrictions devraient être formulées selon des critères précis et ne peuvent pas conférer des pouvoirs illimités aux personnes chargées de veiller à leur application » (Comité des droits de l'homme, observation générale n° 27 (1999) relative à la liberté de circulation, par. 13).

l'on reconnaît que les États ne peuvent imposer de restrictions à la liberté d'expression garantie au paragraphe 2 que dans la mesure où celles-ci sont fixées par la loi et nécessaires au respect « des droits ou de la réputation d'autrui [ou] à la sauvegarde de la sécurité nationale, de l'ordre public, de la santé ou de la moralité publiques ». Il s'agit là d'exceptions très précises (voir, en particulier, [A/67/357](#), par. 41 et [A/HRC/29/32](#), par. 32 à 35), et il appartient à l'autorité qui impose ces restrictions à la liberté d'expression de les justifier, et non à l'individu de démontrer qu'il a droit à cette liberté⁸. Toute restriction doit remplir les trois conditions suivantes :

a) **Légalité** : la restriction doit être fixée par des lois précises, publiques et transparentes ; elle ne doit pas conférer aux autorités un pouvoir discrétionnaire illimité et doit être dûment notifiée aux personnes à qui elle s'applique. Les règles en la matière devraient être soumises aux observations du public et faire l'objet de procédures législatives ou administratives régulières. Des garanties procédurales, en particulier celles assurées par les cours ou tribunaux indépendants, devraient protéger les droits ;

b) **Légitimité** : la restriction doit être justifiée par la nécessité de protéger un ou plusieurs des intérêts visés au paragraphe 3 de l'article 19 du Pacte, à savoir le respect des droits ou de la réputation d'autrui, ou la sauvegarde de la sécurité nationale, de l'ordre public, de la santé ou de la moralité publiques.

c) **Nécessité et proportionnalité** : l'État doit démontrer que la restriction est nécessaire à la protection d'un intérêt légitime et qu'elle est le moyen le moins contraignant d'atteindre le but visé. Le Comité des droits de l'homme qualifie la nécessité et la proportionnalité de « critères stricts », selon lesquels les restrictions « doivent être appliquées exclusivement aux fins pour lesquelles elles ont été prescrites et doivent être en rapport direct avec l'objectif spécifique qui les inspire »⁹.

7. Très souvent, les États affirment que les restrictions imposées à la liberté d'expression sont justifiées mais n'apportent pas la preuve de leur légalité, nécessité ou proportionnalité (voir [A/71/373](#)). Pour cette raison, les règles en la matière doivent être appliquées de manière stricte et de bonne foi et s'accompagner d'une surveillance rigoureuse et transparente. Au titre de l'alinéa b) du paragraphe 3 de l'article 2 du Pacte, les États sont tenus de garantir que, lorsqu'un individu souhaitera introduire un recours pour une violation de ses droits reconnus dans le Pacte, « l'autorité compétente, judiciaire, administrative ou législative, ou toute autre autorité compétente selon la législation de l'État, statuera sur [s]es droits » (voir également [A/HRC/22/17/Add.4](#), par. 31).

Appel à la haine constitutif de l'incitation

8. Aux termes du paragraphe 2 de l'article 20 du Pacte, les États parties doivent interdire par la loi « [t]out appel à la haine nationale, raciale ou religieuse qui constitue une incitation à la discrimination, à l'hostilité ou à la violence ». Les États ne sont pas tenus d'ériger en infraction pénale cette forme d'expression. Le précédent Rapporteur spécial a expliqué que le paragraphe 2 de l'article 20 vise : a) « l'appel à la haine », b) « l'appel constitutif de l'incitation » et c) l'incitation pouvant aboutir à la discrimination, à l'hostilité ou à la violence ([A/67/357](#), para. 43).

9. Les normes des Nations Unies en matière de droits de l'homme offrent, contre la discrimination, une protection plus large que celle qu'offre le paragraphe 2 de l'article 20 du Pacte, qui met l'accent sur la haine nationale, raciale ou religieuse. Le

⁸ Comité des droits de l'homme, observation générale n° 34 (2011), par. 27.

⁹ Ibid., par. 22.

paragraphe 1 de l'article 2 du Pacte garantit des droits à tous les individus, « sans distinction aucune », et l'article 26 dispose expressément que « la loi doit interdire toute discrimination et garantir à toutes les personnes une protection égale et efficace contre toute discrimination ». Les normes internationales garantissent une protection contre les violations fondées sur la race, la couleur, le sexe, la langue, la religion, l'opinion politique ou toute autre opinion, l'origine nationale ou sociale, la fortune, la naissance ou toute autre situation, y compris l'origine ou l'identité autochtone, le handicap, le statut de migrant ou de réfugié, l'orientation sexuelle, l'identité de genre ou l'intersexualité¹⁰. Avec le temps, le champ couvert par la protection s'est élargi, si bien que d'autres catégories, comme l'âge ou l'albinisme, sont aussi expressément protégées aujourd'hui. Compte tenu de l'expansion de la protection dans le monde entier, l'interdiction de l'incitation devrait être comprise comme s'appliquant à toutes les catégories désormais visées par le droit international des droits de l'homme.

10. Il est essentiel de souligner que l'individu dont il y a lieu d'interdire l'expression en vertu du paragraphe 2 de l'article 20 du Pacte est l'auteur d'un appel qui constitue une incitation. Le fait de ne pas lancer un appel à la haine constitutif d'une incitation à la discrimination, à l'hostilité ou à la violence, comme, par exemple, celui de faire une interprétation minoritaire, voire offensante, d'un principe religieux ou d'un événement historique, ou de donner des exemples d'appel à la haine et d'incitation en vue de les signaler ou de sensibiliser à la question, n'est pas visé par l'interdiction prévue à l'article 20 (ou dans toute autre disposition du droit des droits de l'homme). Cette forme d'expression doit être protégée par l'État, même si celui-ci la désapprouve ou s'en trouve offensé¹¹. Le « musellement du contradictoire » n'existe pas en droit international des droits de l'homme¹².

11. Dans la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, adoptée l'année précédant l'entrée en vigueur du Pacte international relatif aux droits civils et politiques, les États sont invités à « éliminer toute incitation » à la discrimination raciale « ou tous actes » de discrimination raciale, en « tenant dûment compte » des autres droits protégés par le droit des droits de l'homme, y compris la liberté d'expression (voir articles 4 et 5 de la Convention). À l'article 4 de la Convention, les États parties s'engagent, entre autres, à : a) « déclarer délits punissables par la loi toute diffusion d'idées fondées sur la supériorité ou la haine raciale, toute incitation à la discrimination raciale, ainsi que tous actes de violence, ou provocation à de tels actes, dirigés contre toute race ou tout groupe de personnes d'une autre couleur ou d'une autre origine ethnique » ; « déclarer illégales et à interdire les organisations ainsi que les activités de propagande organisée et tout autre type d'activité de propagande qui incitent à la discrimination raciale et qui l'encouragent et à déclarer délit punissable par la loi la participation à ces organisations ou à ces activités ».

12. Le paragraphe 2 de l'article 20 du Pacte international relatif aux droits civils et politiques et l'article 4 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale visent des formes précises d'expression, souvent qualifiées de « discours haineux »¹³. Le libellé de ces dispositions est ambigu, comparé à celui du paragraphe 2 de l'article 19 du Pacte¹⁴. Si la définition de la liberté

¹⁰ Voir également Article 19, « "Hate Speech" Explained: A Toolkit » (Londres, 2015), p. 14. À propos de la violence en ligne à l'égard des femmes, voir [A/HRC/38/47](#).

¹¹ Comité des droits de l'homme, observation générale n° 34, par. 11.

¹² Voir Evelyn M. Aswad, « To ban or not to ban blasphemous videos », *Georgetown Journal of International Law*, vol. 44, n° 4 (2013).

¹³ Voir Jeremy Waldron, « *The Harm in Hate Speech* » (Harvard University Press, 2012).

¹⁴ Cette ambiguïté n'est pas surprenante, vu le déroulement des négociations. Voir Jacob Mchangama, « *The sordid origin of hate-speech laws* », *Policy Review* (décembre 2011 et janvier 2012).

d'expression donnée au paragraphe 2 de l'article 19 couvre des droits élargis exprimés avec des verbes actifs (rechercher, recevoir, répandre) et un champ aussi large que possible (des idées de toute espèce, sans considération de frontières, par tout autre moyen), les interdictions énoncées au paragraphe 2 de l'article 20 du Pacte et à l'article 4 de la Convention, bien que plus précises que la simple notion générale de « discours haineux », sont libellées dans des termes connotés d'émotions difficiles à cerner (haine, hostilité) et sont éminemment contextuelles (appel qui constitue une incitation). Le Comité des droits de l'homme a conclu que les articles 19 et 20 du Pacte « sont compatibles l'un avec l'autre et se complètent »¹⁵. Or, même si cette conclusion est vraie, ces articles nécessitent une interprétation.

13. Dans son observation générale n° 34 (2011), le Comité des droits de l'homme estime que, chaque fois qu'un État impose des restrictions à la liberté d'expression, y compris la forme d'expression visée au paragraphe 2 de l'article 20 du Pacte, il n'en doit pas moins « justifier les interdictions et les dispositions qui les définissent, dans le strict respect de l'article 19 »¹⁶. En 2013, un groupe de haut niveau composé d'experts en matière de droits de l'homme, constitué sous les auspices de la Haute-Commissaire des Nations Unies aux droits de l'homme, a donné une interprétation du paragraphe 2 de l'article 20¹⁷. Dans le Plan d'action de Rabat sur l'interdiction de l'appel à la haine nationale, raciale ou religieuse qui constitue une incitation à la discrimination, à l'hostilité ou à la violence, les termes clefs en sont définis comme suit :

Les termes « haine » et « hostilité » font référence à « des manifestations intenses et irrationnelles d'opprobre, d'inimitié ou de détestation envers le groupe visé » ; le terme « appel » sous-entend qu'il y a une « intention de promouvoir la haine publiquement envers le groupe visé » ; le terme « incitation » fait référence à « des déclarations sur des groupes nationaux, raciaux ou religieux qui créent un risque imminent de discrimination, d'hostilité ou de violence envers des personnes appartenant à ces groupes » (A/HRC/22/17/Add.4, annexe, note de bas de page 5)¹⁸.

14. Afin d'évaluer le niveau de gravité requis pour faire de l'incitation une infraction pénale, six facteurs ont été définis dans le Plan d'action de Rabat (ibid., par. 29) :

a) Les « contextes sociaux et politiques qui existent au moment où l'acte verbal a été émis et propagé » ;

b) Le statut de l'orateur, « en particulier la position de cette personne ou de son organisation dans le contexte de l'auditoire auquel s'adresse le discours » ;

c) L'intention, ce qui signifie que « [l]a négligence ou l'imprudence ne sont pas suffisantes pour qualifier la situation au sens de l'article 20 » du Pacte : au sens dudit article, une « simple dissémination ou circulation d'une information » ne constitue pas un appel ou une incitation ;

d) Le contenu et la forme du discours, en particulier « le degré de provocation et la manière dont ce dernier est direct, ainsi que la forme, le style, la nature des arguments utilisés » ;

¹⁵ Conseil des droits de l'homme, observation générale n° 34 (2011), par. 50.

¹⁶ Ibid., par. 52 et, dans le contexte particulier du paragraphe 2 de l'article 20 du Pacte, voir par. 50.

¹⁷ Voir, par exemple, la recommandation générale n° 35 (2013) du Comité pour l'élimination de la discrimination raciale, relative à la lutte contre les discours de haine raciale.

¹⁸ L'ancien Rapporteur spécial, Frank La Rue, a défini comme un facteur essentiel dans l'évaluation de l'incitation le fait de voir s'il existe un « danger réel et imminent de violence résultant des propos tenus » (A/67/357, par. 46). Voir également Article 19, « *Prohibiting Incitement to Discrimination, Hostility or Violence* » (Londres, 2012), p. 24 et 25.

e) L'ampleur ou la portée du discours, à savoir la prise en compte d'éléments comme « la taille de son audience » et l'examen d'autres éléments répondant à des questions telles que : par quels moyens le discours a-t-il été diffusé, par exemple au moyen d'« un seul dépliant ou diffusé dans les médias grand public ou par internet » ? Quelle était « la fréquence, la quantité et la portée de la communication » ? L'audience a-t-elle eu « les moyens d'agir à partir de l'incitation » ? ;

f) La probabilité, y compris l'imminence, qui implique la nécessité d'« identifier le niveau de risque de préjudice », notamment en faisant déterminer par les juridictions, comme prévu dans le Plan d'action, « la probabilité raisonnable que le discours ait pu inciter des actions concrètes envers un groupe visé ».

15. En 2013, le Comité pour l'élimination de la discrimination raciale, qui est l'organe composé d'experts chargés de surveiller l'application de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, a suivi la direction imprimée par le Comité des droits de l'homme et dans le Plan d'action de Rabat. Le Comité a apporté des éclaircissements concernant les termes « tenant dûment compte » contenus dans l'article 4 de la Convention en précisant qu'ils impliquent l'impératif de respecter strictement les garanties de la liberté d'expression¹⁹. Dans une volonté de faire converger les interprétations, le Comité a souligné que l'incrimination énoncée à l'article 4 est réservée à certains cas donnés ci-après :

« [S]eulement les formes graves de discours racistes [devraient être] considérées comme des infractions pénales, pouvant être prouvées au-delà de tout doute raisonnable, les formes moins graves devant être traitées par d'autres moyens que le droit pénal, compte tenu notamment de la nature et de l'étendue des conséquences pour les personnes et les groupes visés. L'imposition de sanctions pénales devrait être régie par les principes de légalité, de proportionnalité et de nécessité »²⁰.

16. Le Comité pour l'élimination de la discrimination raciale a expliqué que les conditions énoncées à l'article 19 du Pacte international relatif aux droits civils et politiques s'appliquent également aux restrictions prévues à l'article 4 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale²¹. En ce qui concerne la qualification en délits punissables par la loi des actes de diffusion et d'incitation, le Comité pour l'élimination de la discrimination raciale a estimé que, pour déterminer si une forme d'expression entre dans les catégories interdites, les États doivent prendre en compte un ensemble de facteurs, dont le « contenu et la forme » du discours, le « climat économique, social et politique » dans lequel le discours a été prononcé, la « position et le statut de l'orateur », la « portée du discours » et ses objectifs²². Le Comité a recommandé que les États parties à la Convention tiennent compte « du risque imminent ou de la probabilité que le comportement recherché ou préconisé par l'orateur débouche sur de l'incitation ».

17. Le Comité a également estimé qu'à la lumière de la Convention, sont interdites « l'expression d'insultes, de moqueries ou de calomnies à l'égard de personnes ou de groupes, ou la justification de la haine, du mépris ou de la discrimination », soulignant

¹⁹ Comité pour l'élimination de la discrimination raciale, recommandation générale n° 35 (2013), par. 19. Le Comité considère que les termes « tenant dûment compte » revêtent une importance particulière en ce qui concerne la liberté d'expression qui, selon lui, constitue « le principe de référence le plus pertinent lorsqu'il s'agit d'apprécier la légitimité de restrictions imposées à la liberté de parole ».

²⁰ Comité pour l'élimination de la discrimination raciale, recommandation générale n° 35 (2013), par. 12.

²¹ Ibid., par. 4, 19 et 20.

²² Ibid., par. 15 et 16.

que ces actes ne peuvent être interdits que lorsqu'ils « s'apparentent clairement à de l'incitation à la haine ou à la discrimination »²³. Les termes « moqueries » et « justification » couvrent un champ extrêmement large et sont généralement exclus des restrictions prévues par le droit international des droits de l'homme, qui protège les droits d'offenser et de se moquer. Ainsi, les liens avec l'incitation et le cadre établi au paragraphe 3 de l'article 19 contribuent à limiter cette interdiction à la catégorie la plus grave.

18. Dans le Plan d'action de Rabat, on précise également que l'incrimination devrait être réservée aux formes d'incitation les plus graves visées au paragraphe 2 de l'article 20 et que, d'une manière générale, d'autres mesures devraient être envisagées dans un premier temps (A/HRC/22/17/Add.4, annexe, par. 34) : déclarations publiques de personnalités influentes de la société faisant contrepoids aux discours haineux et favorisant la tolérance et le respect entre collectivités ; sensibilisation et dialogue interculturel ; élargissement de l'accès à l'information et aux idées propres à désamorcer les messages de haine ; promotion des normes et principes des droits de l'homme et formation en la matière. Le fait que des mesures autres que l'interdiction en droit puissent être appliquées signifie que, bien souvent, de tous les moyens dont disposent les États pour faire face au phénomène des discours haineux, l'interdiction n'en sera pas le moins restrictif.

Expressions de haine ne constituant pas nécessairement un appel ou une incitation

19. Il existe d'autres types de discours comportant, par exemple, un appel à la haine, qui peuvent ne pas correspondre aux définitions ou atteindre les seuils énoncés au paragraphe 2 de l'article 20 ou à l'article 4. Il y a alors lieu de se demander si les États peuvent imposer des restrictions à l'appel à la haine qui ne constitue pas une incitation à la discrimination, à l'hostilité ou à la violence. Autrement dit, il faut se demander si les États peuvent imposer des restrictions au discours haineux qui est défini, comme il l'a récemment été dans la Stratégie et le Plan d'action des Nations Unies pour la lutte contre les discours de haine, comme un discours « constituant une atteinte ou utilisant un langage péjoratif ou discriminatoire à l'égard d'une personne ou d'un groupe en raison de leur identité, en d'autres termes, de l'appartenance religieuse, de l'origine ethnique, de la nationalité, de la race, de la couleur de peau, de l'ascendance, du genre ou d'autres facteurs constitutifs de l'identité »²⁴. De toute évidence, cette définition est en-deçà de celle de la notion d'incitation donnée au paragraphe 2 de l'article 20 et à l'article 4 ; or, si les États et les entreprises doivent combattre ces attitudes par l'éducation, l'expression d'une condamnation ou d'autres moyens, les restrictions imposées par la loi devront être conformes aux normes strictes du droit international des droits de l'homme.

20. En ce qui concerne les contenus qui comportent la forme d'expression visée dans la Stratégie des Nations Unies pour la lutte contre les discours de haine, à savoir le discours qui est haineux mais ne constitue pas une incitation, le paragraphe 3 de l'article 19 du Pacte donne des orientations utiles. Ses dispositions sont à appliquer de manière stricte, de sorte que toute restriction – et toute mesure prise contre la liberté d'expression – remplisse les critères de légalité, nécessité et proportionnalité ainsi que de légitimité. Du fait de son imprécision, un libellé comme celui de la définition donnée dans la Stratégie, s'il est censé orienter l'imposition d'interdictions

²³ Ibid., par. 13.

²⁴ Dans le Plan d'action de Rabat, il est fait référence à une forme d'expression qui n'atteint pas les seuils fixés au paragraphe 2 de l'article 20 du Pacte mais qui « peut justifier une procédure civile ou des sanctions administratives » ou à l'expression qui ne donne pas suite à des sanctions mais « pose néanmoins des problèmes en matière de tolérance, de civilité et de respect des droits d'autrui » (A/HRC/22/17/Add.4, par. 20).

en droit, poserait problème sur le plan de la légalité, même s'il peut servir de base à une action politique et sociale visant à lutter contre la discrimination et la haine. Tout État qui adopterait cette définition devrait également fixer une restriction conforme au critère de la légitimité. Dans la plupart des cas, les droits d'autrui, tels que définis au paragraphe 3 de l'article 19, peuvent constituer une base appropriée, axée sur les droits associés à la discrimination ou à l'atteinte à la vie privée, ou sur la sauvegarde de l'ordre public. En tout état de cause toutefois, l'État aurait impérativement à démontrer la nécessité et la proportionnalité de la mesure prise ; de surcroît, plus la peine est sévère, plus il devra en démontrer la stricte nécessité²⁵.

21. Certaines restrictions sont expressément découragées au regard des normes internationales relatives aux droits de l'homme. Ainsi, premièrement, le Comité des droits de l'homme a souligné que « les interdictions des manifestations de manque de respect à l'égard d'une religion ou d'un autre système de croyance, y compris les lois sur le blasphème, sont incompatibles avec le Pacte », sauf dans les cas où le blasphème peut également être défini comme un appel à la haine religieuse qui constitue un des trois types d'incitation visés²⁶. En d'autres termes, les lois sur le blasphème ne remplissent pas le critère de légitimité énoncé au paragraphe 3 de l'article 19 du Pacte puisque ledit article protège les individus ainsi que leur droit à la liberté d'expression et d'opinion : ni le paragraphe 3 de l'article 19 ni l'article 18 du Pacte ne protègent les idées ou les croyances contre les moqueries, les insultes, les critiques ou autres « atteintes » considérées comme offensantes. Plusieurs mécanismes de défense des droits de l'homme ont appelé à l'abrogation des lois sur le blasphème du fait qu'elles compromettent le débat autour de la religion et qu'elles peuvent être exploitées par les gouvernements pour faire prévaloir les idées d'une religion sur celles d'autres religions, d'autres systèmes de croyance ou de systèmes non religieux (voir, en particulier, [A/HRC/31/18](#), par. 59 à 61).

22. Deuxièmement, les lois qui « criminalisent l'expression d'opinions concernant des faits historiques sont incompatibles » avec l'article 19 du Pacte, ce qui remet en question les lois qui criminalisent la négation de la Shoah et d'autres atrocités ainsi que les lois analogues, qui sont souvent justifiées par référence au discours haineux. Le Comité des droits de l'homme a relevé que l'expression d'une opinion « erronée » ou d'une « interprétation incorrecte d'événements du passé » ne peut faire l'objet d'une interdiction générale et que les restrictions imposées à ces formes d'expression « ne devraient pas aller au-delà de ce qui est permis » par le paragraphe 3 de l'article 19 « ou exigé par l'article 20 » du Pacte²⁷. Au vu, notamment, de ces interprétations, le déni de l'historicité d'atrocités ne doit pas faire l'objet de sanctions pénales ou d'autres restrictions sans qu'une appréciation supplémentaire n'ait été faite d'après les définitions et contextes mentionnés plus haut. L'imposition de restrictions de ce genre en droit international des droits de l'homme devrait être conditionnée à l'évaluation des six facteurs énumérés dans le Plan d'action de Rabat.

23. Troisièmement, une autre forme d'expression non constitutive de l'incitation peut se produire lorsque l'orateur vise isolément une victime identifiable sans chercher à inciter autrui à s'en prendre à des individus sur la base d'une

²⁵ Il est peu probable que l'exception de la moralité publique énoncée au paragraphe 3 de l'article 19 du Pacte constitue une base en l'espèce, mais on notera que le Comité des droits de l'homme a précisé que « les restrictions [...] pour protéger la morale doivent être fondées sur des principes qui ne procèdent pas d'une tradition unique » (Comité des droits de l'homme, observation générale n° 34 (2011), par. 32, citant l'observation générale n° 22 (1993) relative au droit à la liberté de pensée, de conscience et de religion, par. 8).

²⁶ Comité des droits de l'homme, observation générale n° 34 (2011), par. 48. En pareil cas, ce ne serait pas le blasphème qui serait en cause mais plutôt l'appel constitutif d'une incitation.

²⁷ Comité des droits de l'homme, observation générale n° 34 (2011), par. 49. Voir Sarah Cleveland, « *Hate Speech at Home and Abroad* », dans Lee C. Bollinger et Geoffrey R. Stone, dir., *The Free Speech Century* (New York, Oxford University Press, 2019). Voir également [A/67/357](#), par. 55.

caractéristique protégée²⁸. Là encore, par référence au paragraphe 3 de l'article 19 du Pacte, il est possible de restreindre cette forme d'expression pour assurer le respect des droits d'autrui ou la sauvegarde de l'ordre public. Souvent, les États imposent des restrictions à cette forme d'expression en la classant dans la catégorie générale des « crimes haineux », qui fait de la motivation haineuse, aux fins de la peine, une circonstance aggravante de l'atteinte à l'intégrité physique d'une personne ou du préjudice porté à un bien.

24. Quatrièmement, il importe de souligner que l'expression qui peut être offensante ou empreinte de préjugés et qui peut faire fortement penser à une manifestation d'intolérance peut souvent ne pas atteindre un seuil de gravité suffisant pour justifier une restriction, quelle qu'elle soit. Il existe différentes formes d'expression de haine, aussi déplorables soient-elles, qui ne constituent ni une incitation ni une menace directe, comme, par exemple, l'emploi d'un langage stigmatisant à l'encontre de groupes protégés. L'expression de tels sentiments ne devrait pas faire l'objet de l'interdiction visée dans le Pacte international relatif aux droits civils et politiques ou dans la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, et l'imposition d'autres restrictions ou mesures répressives devrait nécessairement s'appuyer sur une appréciation des conditions prévues au paragraphe 3 de l'article 19 du Pacte. Les six facteurs énoncés dans le Plan d'action de Rabat pour ériger l'incitation en infraction pénale constituent également un cadre utile pour évaluer la réaction des autorités publiques face à ces formes d'expression. De fait, l'absence de restriction n'est pas synonyme d'inaction : les États peuvent (et devraient, conformément à la résolution 16/18 du Conseil des droits de l'homme) prendre des mesures énergiques, telles que la condamnation des préjugés par les pouvoirs publics, la sensibilisation, la formation, les annonces d'intérêt public et les projets d'intérêt collectif, pour lutter contre ce type d'intolérance et veiller à ce que les autorités publiques protègent les individus contre la discrimination fondée sur cette forme d'expression de haine.

25. Enfin, au titre de la Convention pour la prévention et la répression du crime de génocide, les États sont tenus de criminaliser l'incitation à commettre le génocide. Dans certains cas, comme au Myanmar, l'inaction de l'État face à l'incitation au génocide peut expliquer en partie les conséquences très graves qu'endurent les populations vulnérables. Une telle inaction en soi est condamnable et, tout comme l'incitation elle-même, doit être combattue et punie²⁹.

Normes régionales relatives aux droits de l'homme

26. En Europe, aux Amériques et en Afrique, les systèmes de protection des droits de l'homme définissent également des normes applicables aux discours haineux. La Cour européenne des droits de l'homme a souligné que la liberté d'expression protège les formes de discours qui « heurtent, choquent ou inquiètent »³⁰. Toutefois, la Cour a adopté une attitude relativement déférente à l'égard d'États qui continuent, dans leur législation, d'interdire le blasphème en en faisant une forme de discours haineux ou qui continuent de criminaliser la négation du génocide, contrairement aux tendances observées à l'échelle mondiale³¹. Souvent, la Cour élude purement et

²⁸ Article 19, « *Hate Speech* » Explained », p. 22.

²⁹ Voir, en particulier, A/HRC/39/64, par. 73. Au titre du point c) de l'article III de la Convention pour la prévention et la répression du crime de génocide, « l'incitation directe et publique à commettre le génocide » doit être punie par la loi.

³⁰ Cour européenne des droits de l'homme, *Handyside c. Royaume-Uni*, requête n° 5493/72, arrêt du 7 décembre 1976, par. 49. Voir Sejal Parmar, « *The legal framework for addressing 'hate speech' in Europe* », exposé fait lors de la conférence internationale sur la lutte contre le discours haineux dans les médias, Zagreb, novembre 2018.

³¹ Voir Conseil de l'Europe, « Discours de haine », fiche thématique, octobre 2019, et Evelyn M. Aswad, « *The future of freedom of expression online* », *Duke Law and Technology Review*, vol. 17 (août 2018).

simplement la question du discours haineux, s'appuyant non pas sur la liberté d'expression, mais plutôt sur l'« abus de droit » pour conclure à l'irrecevabilité des allégations de violation³². Les normes européennes peuvent fluctuer lorsqu'il s'agit d'engager la responsabilité des intermédiaires pour les discours haineux diffusés sur leurs plateformes³³. En revanche, les normes de la Commission interaméricaine des droits de l'homme présentent généralement des similitudes avec les normes internationales énoncées plus haut, tandis que la définition de normes dans le système africain n'en est qu'à un stade relativement précoce³⁴. Quoiqu'il en soit, on ne saurait invoquer des normes régionales pour justifier une dérogation aux normes du système international de protection des droits de l'homme.

27. Le Comité des droits de l'homme a expressément rejeté la doctrine de la marge d'appréciation énoncée par la Cour européenne, précisant que « l'État partie doit, dans une affaire donnée, démontrer de manière spécifique la nature précise de la menace qui pèse sur l'un quelconque des éléments énoncés au paragraphe 3 de l'article 19, et qui l'a conduit à restreindre la liberté d'expression »³⁵. Le Comité n'accorde pas de pouvoir discrétionnaire à l'État du seul fait que les autorités nationales affirment qu'elles sont généralement mieux placées pour comprendre le contexte de leur pays.

Synthèse des instruments de l'ONU relatifs au discours haineux

28. Ces dernières années, le cadre international relatif aux droits de l'homme a connu des changements induits par une volonté de rationaliser ce qui apparaît, à première vue, comme des normes concurrentes. En résumé, la liberté d'expression est un droit reconnu par la loi, primordial pour les sociétés démocratiques, qui a un lien d'interdépendance avec tous les autres droits consacrés dans le droit des droits de l'homme, qu'il renforce. Dans le même temps, la lutte contre la discrimination, l'égalité et la participation égale et effective à la vie publique sous-tendent l'ensemble du corpus du droit des droits de l'homme. La forme d'expression visée à l'article 20 du Pacte international relatif aux droits civils et politiques et à l'article 4 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale est problématique au regard des deux ensembles de normes, ce dont tous les participants à la vie publique doivent être conscients. Ainsi, les restrictions du droit à la liberté d'expression doivent être exceptionnelles, et il incombe à l'État qui les impose de démontrer qu'elles sont en conformité avec le droit international ; les interdictions prévues à l'article 20 du Pacte et à l'article 4 de la Convention doivent être soumises aux conditions strictes et précises énoncées au paragraphe 3 de l'article 19 du Pacte, et les États devraient généralement user de tous les moyens autres que la criminalisation et l'interdiction dont ils disposent, dont la sensibilisation, les contre-discours et la promotion du pluralisme, pour combattre toutes les formes de discours haineux.

³² Pour un aperçu de la pratique, voir Conseil de l'Europe, « *Guide on article 17 of the European Convention on Human Rights: prohibition of abuse of rights* », mise à jour du 31 août 2019.

³³ Comparer Cour européenne des droits de l'homme, Grande Chambre, *Delfi AS c. Estonie*, requête n° 64569/09, arrêt du 16 juin 2015, et Cour européenne des droits de l'homme, Section IV, *Magyar Tartalomszolgáltatók Egyesülete et Index.hu Zrt c. Hongrie*, requête n° 22947/13, arrêt du 2 février 2016. Voir également Article 19, « *Responding to 'hate speech': comparative overview of six EU countries* », 2018.

³⁴ Voir Commission interaméricaine des droits de l'homme, « *Hate speech and incitement to violence* », dans Commission interaméricaine des droits de l'homme, *Violence against Lesbian, Gay, Bisexual, Trans and Intersex Persons in the Americas* (2015).

³⁵ Comité des droits de l'homme, observation générale n° 34 (2011), par. 36.

III. Réglementation des discours haineux en ligne

A. Obligations des États et réglementation des discours haineux en ligne

29. Le strict respect des normes relatives au droit international des droits de l'homme protège contre les excès des pouvoirs publics. D'emblée, les États ne devraient pas se servir des entreprises de l'Internet pour imposer des limites aux formes d'expression que le droit international des droits de l'homme leur interdit de restreindre. Les mesures qu'ils exigent des entreprises, soit dans le cadre de la réglementation, soit sous la menace d'une réglementation, doivent être justifiées en droit international et conformes à celui-ci. Certains types de mesures visant des contenus ne sont manifestement pas en conformité avec le paragraphe 3 de l'article 19 du Pacte international relatif aux droits civils et politiques, comme le blocage de l'accès à Internet ou la criminalisation de la dissidence politique ou de la critique envers les autorités publiques en ligne (voir [A/HRC/35/22](#)). Les peines imposées pour diffusion de discours haineux tombant sous le coup de la loi ne devraient pas être alourdies simplement parce que les propos en question ont été tenus en ligne.

30. Imaginons, aux fins de notre étude, le cas d'un État qui envisage d'adopter une loi engageant la responsabilité des intermédiaires en ligne qui ne prendraient pas de mesures précises pour lutter contre les propos haineux. Ce genre de loi sur la « responsabilité des intermédiaires » vise généralement à restreindre la liberté d'expression – des utilisateurs d'une plate-forme donnée ou de la plate-forme elle-même –, parfois dans le but de respecter l'obligation prévue au paragraphe 2 de l'article 20 du Pacte. Toute évaluation juridique d'un tel projet doit tenir compte des conditions cumulatives énoncées au paragraphe 3 de l'article 19 afin d'assurer sa conformité avec les normes internationales relatives à la liberté d'expression³⁶.

Légalité

31. Conformément aux dispositions énoncées au paragraphe 3 de l'article 19 du Pacte, il importe, pour attribuer la responsabilité d'un discours haineux tenu sur une plateforme, que l'expression elle-même et les facteurs permettant de déterminer les cas constitutifs du discours haineux soient définis. Lorsque l'on entend engager une responsabilité pour défaut d'élimination de l'« incitation », il faut définir le contenu de l'incitation à l'aune du paragraphe 2 de l'article 20 du Pacte et de l'article 4 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, au moyen, notamment, des termes clefs définis dans le Plan d'action de Rabat, mentionnés plus haut. Un État qui souhaite réglementer les discours haineux pour des motifs autres que ceux prévus à l'article 20 du Pacte et à l'article 4 de la Convention doit préciser les contenus qui sont effectivement punissables par la loi³⁷ : le fait

³⁶ Pour une déclaration sur les principes qui devraient s'appliquer dans le contexte de la responsabilité des intermédiaires, voir Electronic Frontier Foundation, « *Manila principles on intermediary liability* », 2015.

³⁷ La plupart des États ont fait la distinction entre les contenus terroristes ou « extrémistes » et le « discours haineux », mais les mêmes principes de légalité doivent également s'appliquer à ces notions. Voir, par exemple, [A/HRC/40/52](#), par. 75 e). Le terme « extrémisme » est souvent utilisé à la place de « discours haineux », même si ce terme n'est pas ancré dans la loi. L'expression « extrémisme violent » n'apporte pas beaucoup de clarté. Les gouvernements qui emploient le terme « extrémisme » de bonne foi dans un contexte en ligne semblent mettre l'accent sur le problème de la viralité des « idéologies du terrorisme et de l'extrémisme violent » et semblent avoir pour objectif de contrer les discours « extrémistes » et de « prévenir l'utilisation d'Internet à des fins néfastes » (Christchurch Call to Eliminate Terrorist and Violent Extremist Content Online).

d'exiger précision et clarté au paragraphe 3 de l'article 19 du Pacte signifie que la législation nationale doit limiter le pouvoir discrétionnaire excessif des acteurs gouvernementaux s'agissant de faire respecter les règles, ou celui des acteurs privés s'agissant d'utiliser ces règles pour réprimer des expressions légitimes, et prévoir que les individus visés seront dûment informés de la réglementation concernant leurs activités³⁸. Si les définitions ne sont pas claires et précises, on risque fort d'assister à des excès, de voir imposées des restrictions à des contenus légitimes et de laisser sans solution les problèmes suscités. Les États qui s'attaquent au phénomène du discours haineux devraient établir un lien étroit entre leurs définitions et les normes relatives au droit international des droits de l'homme, comme celles établies au paragraphe 2 de l'article 20 du Pacte.

32. Plusieurs États ont adopté ou envisagent d'adopter des lois obligeant les entreprises Internet à supprimer les discours « manifestement contraires à la loi » dans un délai donné, généralement dans les 24 heures, voire rien qu'en une heure, ou alors à éliminer, dans un délai plus long, les contenus interdits par la loi. La plus connue de ces lois – la loi allemande relative à la réglementation des réseaux sociaux, dite « NetzDG » – oblige les entreprises à retirer de leurs plateformes les discours interdits par un certain nombre de dispositions bien définies du Code pénal allemand³⁹. Ainsi, l'article 130 du Code pénal dispose, entre autres, qu'est passible de sanctions tout individu qui, d'une manière pouvant troubler l'ordre public, incite à la haine contre un groupe national, racial ou religieux ou un groupe défini par son origine ethnique, contre des catégories de la population ou des individus en raison de leur appartenance à un des groupes ou catégories précédemment mentionnés, ou réclame l'adoption de mesures violentes ou arbitraires à leur encontre⁴⁰. Manifestement, la législation ne donne pas une définition des termes clefs qu'elle emploie (en particulier, les termes « inciter » et « haine »)⁴¹, alors que, par la loi « NetzDG », elle impose des amendes importantes aux entreprises qui ne se conforment pas à ses dispositions. L'imprécision du texte sur lequel reposent ces dispositions pose problème. Même si la loi « NetzDG » doit être comprise comme un effort entrepris de bonne foi pour répondre aux inquiétudes générales que suscitent les discours haineux en ligne et leurs conséquences hors ligne, l'absence de définition des termes clefs qui y sont utilisés fait qu'il est hasardeux d'affirmer que les obligations découlant de ses dispositions sont conformes au droit international relatif aux droits de l'homme.

33. Rares sont les États qui ont fait participer leurs autorités judiciaires à l'évaluation des plateformes diffusant des discours haineux interdits par leur droit interne ; or, les États ne devraient permettre l'attribution de la responsabilité que sur ordre de tribunaux indépendants et avec la possibilité, pour l'intermédiaire ou toute autre partie intéressée (comme l'utilisateur concerné), de faire appel de la décision en question⁴². Les pouvoirs publics font de plus en plus pression sur les entreprises pour

³⁸ Ce qui précède n'exclut pas la possibilité d'actions civiles qu'un individu peut intenter contre un autre pour des délits classiques commis en ligne plutôt que dans le monde hors ligne. Toutefois, l'article 19 du Pacte exige que soit définie l'expression qui peut causer un préjudice juridiquement réparable.

³⁹ Allemagne, *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* (Loi relative à la réglementation des réseaux sociaux, dite « NetzDG ») (2017), art. 1 3).

⁴⁰ Des références analogues sont faites dans la proposition de loi française visant à lutter contre les contenus haineux sur Internet. Voir la communication FRA 6/2019 et la réponse du Gouvernement français, accessible à l'adresse suivante : <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>.

⁴¹ Voir, par ailleurs, Allemagne, Cour fédérale de justice, arrêt du 3 avril 2008, affaire n° 3 StR 394/07.

⁴² L'ancien Rapporteur spécial avait souligné que « [l]es restrictions imposées doivent par ailleurs être appliquées par un organe indépendant de toute pression politique ou commerciale ou de toute

qu'elles jouent un rôle d'arbitre face aux discours haineux. L'adoption des lois est un processus qui devrait lui aussi être soumis à des normes rigoureuses en matière de légalité, prévoyant la possibilité, pour les citoyennes et citoyens, de donner leur avis, de participer à des débats publics et d'évaluer d'autres solutions ainsi que les effets dudit processus sur les droits de l'homme⁴³.

Nécessité et proportionnalité

34. Les mesures législatives visant à encourager le retrait des discours haineux en ligne et à engager la responsabilité des entreprises qui ne font pas le nécessaire doivent respecter les critères de nécessité et de proportionnalité énoncés plus haut. Depuis quelques années, les États font pression sur les entreprises pour les amener à supprimer pratiquement d'office des contenus, leur exigeant de filtrer le téléchargement de contenus jugés préjudiciables. Le but ainsi recherché est la mise en place d'outils automatisés qui, en quelque sorte, censureraient les contenus avant même leur publication. Une telle exigence pose problème en ce qu'elle permettrait de bloquer les contenus, sans garantie au niveau du respect de la légalité, avant même leur publication, inversant ainsi le principe bien établi selon lequel il incombe à l'État, et non à l'individu, de justifier les restrictions imposées à la liberté d'expression⁴⁴. Notoirement incapables d'analyser le type de langage naturel habituellement constitutif d'un contenu haineux, ces filtres peuvent être extrêmement lourds de conséquences⁴⁵. Qui plus est, d'après certaines recherches, ils lèsent de façon disproportionnée les communautés traditionnellement sous-représentées⁴⁶.

35. L'exigence tendant à filtrer le téléchargement de discours haineux (et d'autres types de contenus) est à déconseiller, car elle incite les plateformes à réglementer et à retirer des contenus légitimes. Ces filtres renforcent le pouvoir des entreprises et réduisent considérablement, voire éliminent, les possibilités de surveillance ou les voies de recours. Les États devraient, en fait, adopter des lois et des politiques qui amènent les entreprises à protéger la liberté d'expression et à lutter contre les formes de discours haineux légalement restreintes, en exigeant la transparence requise pour permettre au public d'exercer un contrôle, en veillant à l'application du droit interne par des autorités judiciaires indépendantes et en prenant toute autre mesure dans les domaines social et éducatif, conformément aux orientations proposées dans le Plan d'action de Rabat et dans la résolution 16/18 du Conseil des droits de l'homme.

36. Certains États ont voulu réprimer autrement les discours haineux interdits par la loi en recourant à des moyens originaux et qui semblent proportionnés. Alors que l'Inde a pris une mesure discutable en choisissant, dans certains cas, de bloquer l'accès à Internet pour régler des problèmes liés aux contenus, entravant ainsi de

autre influence injustifiée, d'une manière qui ne soit ni arbitraire ni discriminatoire, et des garanties suffisantes doivent être prévues contre les abus » (A/67/357, par. 42).

⁴³ Voir la communication AUS 5/2019 et la réponse de la Mission permanente de l'Australie auprès de l'Office des Nations Unies et des autres organisations internationales à Genève, accessible à l'adresse suivante : <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>.

⁴⁴ Communication OTH 71/2018, accessible à l'adresse suivante : <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>. Voir, également, Commission européenne, recommandation (UE) 2018/334 du 1^{er} mars 2018 sur les mesures destinées à lutter, de manière efficace, contre les contenus illicites en ligne, qui préconise « des mesures proactives proportionnées et spécifiques, faisant notamment appel à des procédés automatisés propres à détecter, circonscrire et retirer rapidement tout contenu à caractère terroriste » (par. 36).

⁴⁵ Voir Center for Democracy and Technology, « *Mixed messages? The limits of automated social media content analysis* », 28 novembre 2017.

⁴⁶ En ce qui concerne les graves préoccupations relatives à la liberté d'expression soulevées à propos des filtres de téléchargement, voir Daphne Keller, « *Dolphins in the Net: Internet content filters and the Advocate General's Glawischnig-Pieczek v. Facebook Ireland opinion* », Stanford Center for Internet and Society, 4 septembre 2019.

manière disproportionnée l'accès de la population aux communications⁴⁷, certains États du pays ont adopté quant à eux une démarche différente. Dans un cas, il s'est agi d'ouvrir des lignes directes pour permettre aux individus de signaler aux forces de l'ordre des contenus diffusés via WhatsApp ; ailleurs, on a mis en place des « laboratoires des médias sociaux » en guise d'observatoire des discours haineux en ligne. Si ces manières de procéder nécessitent une élaboration minutieuse afin d'assurer leur conformité aux normes relatives aux droits de l'homme, elles témoignent d'une démarche originale, sortant des schémas habituels, qui permet de lutter contre les discours haineux sans confier la surveillance des contenus à des entreprises tierces⁴⁸.

37. En 2019, une mission gouvernementale, en France, a proposé une approche de la réglementation des contenus en ligne qui devrait protéger la liberté d'expression tout en permettant de lutter contre les discours haineux interdits par la loi. Même si des informations précises sur l'état d'avancement de ses travaux n'étaient pas disponibles au moment de la rédaction du présent rapport, la mission propose, entre autres choses, la prise en charge par les autorités judiciaires du problème des discours haineux et la promotion d'initiatives multipartites visant à surveiller les politiques des entreprises. La mission a formulé la conclusion suivante :

Une intervention publique pour que les plus grands acteurs adoptent une attitude plus responsable et plus protectrice de la cohésion de nos sociétés apparaît donc légitime. Compte tenu des enjeux de libertés publiques, cette intervention doit faire l'objet de précautions particulières. Elle doit 1) respecter la diversité des modèles de réseaux sociaux, qui forment un ensemble particulièrement hétéroclite, 2) faire preuve de transparence, notamment en y associant la société civile, 3) viser un objectif d'intervention minimum conformément aux principes de nécessité et de proportionnalité et 4) s'en remettre aux juridictions pour la qualification de la licéité des contenus pris individuellement⁴⁹.

38. Cette approche mérite d'être développée et examinée plus avant, car elle aborde les questions de liberté d'expression et de cohésion sociale d'une manière qui semble garantir le respect du droit international des droits de l'homme.

Légitimité

39. La réglementation gouvernementale relative aux intermédiaires en ligne devrait être soumise aux mêmes directives en matière de légitimité que celles prévues en droit des droits de l'homme, qui s'appliquent à toute restriction imposée par les États à la liberté d'expression. Comme indiqué plus haut, certains types d'expression que les États peuvent qualifier de « discours haineux » ne devraient pas faire l'objet des interdictions prévues à l'article 19 ou au paragraphe 2 de l'article 20 du Pacte. En outre, les dispositions légales ayant pour effet de restreindre l'expression qui constitue, par exemple, une incitation à la « haine contre le régime » ou à la « subversion contre l'autorité de l'État » obéissent à des critères de restriction qui sont contraires aux dispositions du paragraphe 3 de l'article 19 du Pacte (A/67/357, par. 51 à 55). Des définitions trop larges du discours haineux interdisant, par exemple,

⁴⁷ Voir les communications IND 7/2017 et IND 5/2016, accessibles à l'adresse suivante : <https://spcommreports.ohchr.org/Tmsearch/TMDocuments> ; voir, également, Haut-Commissariat des Nations Unies aux droits de l'homme, « *United Nations rights experts urge India to end communications shutdown in Kashmir* », communiqué de presse, 22 août 2019.

⁴⁸ Chinmayi Arun et Nakul Nayak, « *Preliminary findings on online hate speech and the law in India* », 8 décembre 2016, p. 11.

⁴⁹ France, « *Créer un cadre français de responsabilisation des réseaux sociaux : agir en France avec une ambition européenne* », rapport de mission remis au Secrétaire d'État en charge du numérique, mai 2019.

l'incitation à la « discorde religieuse » ou une rhétorique risquant d'exposer un pays à des actes violents⁵⁰, se traduisent généralement par une imposition de restrictions à la liberté d'expression à des fins illégitimes ou, dans le cas de la réglementation gouvernementale relative aux intermédiaires en ligne, par l'imposition à ceux-ci d'exigences qui ne sont pas conformes au droit des droits de l'homme.

B. Modération des contenus par les entreprises et discours haineux

40. En ligne, c'est sur les plateformes créées par les entreprises Internet que se propagent les contenus haineux, semblant prospérer grâce à un modèle d'activité qui repose sur la captation de l'attention des utilisateurs et le caractère viral des publications⁵¹. Les plus grandes entreprises ont recours à des « classificateurs » reposant sur l'intelligence artificielle, qui traquent, avec plus ou moins de succès, les contenus interdits en s'appuyant sur des résultats d'analyse et en recherchant des mots particuliers. Les plateformes de ces entreprises sont accessibles depuis diverses juridictions, dans lesquelles un même contenu n'a pas forcément des effets identiques. Sur Internet, les propos haineux sont souvent proférés sous couvert d'anonymat, à la faveur de messages de menaces envoyés de manière coordonnée à l'aide de robots, par des personnes qui ont recours à la désinformation et aux « deep fakes » (vidéos truquées hyperréalistes) et se livrent à des lynchages en ligne⁵².

41. Les entreprises Internet fixent les conditions d'utilisation de leurs plateformes et modèlent la façon dont elles sont perçues du public (autrement dit, leur image de marque)⁵³. Elles ont une influence énorme sur les droits de la personne, en particulier, mais pas seulement, là où leurs plateformes sont le principal moyen d'expression publique ou privée. Dans ces espaces, limiter les moyens d'expression revient donc à réduire le public au silence, de la même manière que manquer de lutter contre les incitations à commettre des infractions peut encourager la violence et la discrimination hors ligne (A/HRC/42/50, par. 70 à 75). La diffusion incontrôlée de propos haineux en ligne peut avoir des conséquences tragiques, comme on a pu le voir lorsque Facebook a manqué de réprimer les propos incitant à la violence contre la communauté musulmane des Rohingya, au Myanmar. Si les entreprises ne sont pas soumises aux mêmes devoirs que les gouvernements, leur influence est néanmoins telle qu'elle doit les inciter à se poser les mêmes questions qu'eux quant à la protection de la liberté d'expression de leurs utilisateurs⁵⁴.

42. Dans de précédents rapports, il a été avancé que toutes les entreprises du secteur des TIC devraient respecter les Principes directeurs relatifs aux entreprises et aux droits de l'homme et tenir compte des droits de l'homme par défaut ainsi que dans la conception de leurs produits. Pourtant, les entreprises gèrent le problème des discours haineux sur leurs plateformes, pour ainsi dire, sans prendre en considération les

⁵⁰ Voir communication JOR 3/2018, disponible à l'adresse suivante : <https://spcommreports.ohchr.org/Tmsearch/TMDocuments>.

⁵¹ Voir Tim Wu, « The Attention Merchants : The Epic Scramble to Get Inside Our Heads » (New York, Vintage Books, 2016).

⁵² Voir Gayathri Venkiteswaran, « “Let the Mob Do the Job” : How Proponents of Hatred are Threatening Freedom of Expression and Religion Online in Asia » (Association pour le progrès des communications, octobre 2017).

⁵³ Voir Kate Klonick, « The new governors: the people, rules, and processes governing online speech », Harvard Law Review, vol. 131, n° 6 (avril 2018) et David Kaye, « Speech Police: The Global Struggle to Govern the Internet » (New York, Columbia Global Reports, 2019).

⁵⁴ Voir A/HRC/32/38, par. 87 et 88 ; voir également Business for Social Responsibility et le Forum économique mondial, « Responsible use of technology », livre blanc, août 2019.

incidences de leurs pratiques sur les droits de la personne⁵⁵. Cette approche est problématique en ce qu'elle prive les entreprises d'un cadre auquel se référer pour prendre des décisions conformes au droit et pour justifier ces décisions auprès des gouvernements et du public, outre qu'elle dénie à celui-ci l'accès à des voies de recours, faute de vocabulaire universellement reconnu. Le Rapporteur spécial appelle une nouvelle fois les entreprises à adopter des politiques respectueuses des droits de l'homme et à mettre en place des dispositifs pour :

- a) Examiner périodiquement les effets de leurs produits sur les droits de l'homme ;
- b) Éviter d'avoir des incidences négatives sur les droits de l'homme, prévenir ces incidences ou les atténuer lorsqu'elles se produisent ;
- c) Appliquer une procédure de diligence raisonnable pour « identifier leurs incidences sur les droits de l'homme, prévenir ces incidences et en atténuer les effets, et rendre compte de la manière dont elles y remédient » et établir des procédures permettant de remédier aux incidences négatives en la matière que pourraient avoir leurs activités⁵⁶.

43. La façon d'appliquer les normes des Nations Unies en matière de droits de l'homme à des contenus très différents suscitera toujours des questions difficiles, tout comme il s'en pose dans le cadre du droit national et du droit régional des droits de l'homme⁵⁷. Pour autant, les textes cités ci-dessus peuvent aider à orienter les mesures prises par les entreprises pour protéger les droits de l'homme à chaque étape du processus de modération – élaboration du produit, définition, repérage, intervention et réparation. Les droits de l'homme internationalement reconnus, que les entreprises sont appelées à respecter dans les Principes directeurs relatifs aux entreprises et aux droits de l'homme (principe 12), constituent un socle solide sur lequel les entreprises à vocation planétaire – de par leurs utilisateurs et leur mode de communication – peuvent appuyer leur réflexion⁵⁸.

Diligence raisonnable en matière de droits de l'homme et examen

44. La lutte contre les discours haineux doit commencer par la prise de mesures de diligence raisonnable, dès la conception des produits. Hélas, il semblerait que, parmi les grandes entreprises Internet, aucun examen des produits axés sur le respect des droits de l'homme dans le contexte des discours haineux n'ait pratiquement été mené et, s'il l'a été, les résultats n'en ont pas été rendus publics. Or, les produits des entreprises du secteur des TIC sont sans cesse mis à jour et revus : il est donc essentiel pour ces entreprises d'évaluer régulièrement leurs produits afin de déterminer s'ils nuisent à l'exercice des droits de l'homme. D'après les Principes directeurs relatifs aux entreprises et aux droits de l'homme, les entreprises devraient, entre autres, évaluer en permanence les effets que les discours haineux tenus sur leurs plateformes peuvent avoir sur les droits de l'homme (principe 17), notamment en utilisant des algorithmes (voir [A/73/348](#)). Elles devraient recourir à des compétences internes ou

⁵⁵ Au moment de l'établissement du présent rapport, Facebook venait d'amender sa déclaration de valeurs pour y indiquer qu'il « se référerait aux normes internationales relatives aux droits de l'homme » pour prendre certaines décisions concernant ses standards de la communauté. Voir Monika Bickert, « Updating the values that inform our community standards », Facebook, 12 septembre 2019.

⁵⁶ Principes directeurs relatifs aux entreprises et aux droits de l'homme : mise en œuvre du cadre de référence « protéger, respecter et réparer » des Nations Unies ([A/HRC/17/31](#), annexe), principes 12 (avec commentaire), 13 et 15.

⁵⁷ Voir Benesch, « Proposals for improved regulation ».

⁵⁸ Voir Business for Social Responsibility, *Human Rights Impact Assessment: Facebook in Myanmar* (octobre 2018).

indépendantes externes dans le domaine des droits de l'homme et mener de véritables consultations avec des groupes et autres acteurs concernés susceptibles d'être touchés (principe 18). Elles devraient aussi évaluer régulièrement l'efficacité des mesures qu'elles ont prises pour remédier aux incidences négatives de leurs activités sur les droits de l'homme (principe 20).

45. Les processus de modération des entreprises manquent tous cruellement de transparence. L'examen par des entités externes (appartenant aux milieux universitaires, judiciaires et autres) des politiques en matière de lutte contre les discours haineux, tel que préconisé par le principe 21, est très difficile : si les règles sont consultables par le public, la manière dont elles sont appliquées, au niveau de l'ensemble des utilisateurs comme dans des cas précis, reste un mystère presque entier. En outre, les entreprises devraient former leurs employés chargés d'établir les politiques relatives aux contenus, leurs conseillers juridiques et, en particulier, leurs modérateurs, qui sont directement responsables de l'application des mesures de restriction (principe 16, commentaire). Dans le cadre de cette formation, il y aura lieu de préciser quelles normes en matière de droit des droits de l'homme la modération sert à protéger et à promouvoir. Les entreprises doivent, surtout, déterminer si leurs règles relatives aux discours haineux portent atteinte à la liberté d'expression, à l'aune des principes de légalité, de nécessité et de légitimité mentionnés précédemment.

Critère de légalité

46. La définition que les entreprises donnent du discours haineux est le plus souvent difficile à comprendre, même si les cas varient. Dans certaines entreprises, cette définition est vague ; dans d'autres, elle est inexistante. Ainsi, le réseau social russe VK interdit les contenus qui font l'apologie de la haine ou de la violence raciale, religieuse et ethnique ou y contribuent, les contenus qui véhiculent des idéologies fascistes ou prônent une supériorité raciale, ou les contenus extrémistes⁵⁹. L'application de messagerie chinoise WeChat interdit les contenus qui, objectivement ou de l'avis de l'entreprise, constituent une forme de harcèlement ou sont haineux, injurieux, offensants envers certaines races ou ethnies, diffamatoires, humiliants pour un tiers (qu'ils soient publics ou non), menaçants, vulgaires ou autrement répréhensibles⁶⁰. D'autres entreprises donnent des définitions très précises, qui témoignent de l'effort sérieux entrepris pour décrire en détail les types de contenus constituant des discours haineux et, de ce fait, soumis à restriction. Paradoxalement, le caractère exhaustif de ces définitions peut les rendre peu claires et difficiles à comprendre. Les conditions d'utilisation de trois entreprises américaines dominantes (YouTube, Facebook et Twitter) se sont améliorées au fil des ans, évoluant indépendamment mais dans une même direction, pour finalement aboutir à des résultats similaires⁶¹. Cependant, ces entreprises, qui emploient des termes différents pour désigner le fait de restreindre la diffusion de contenus incitant à la violence ou à la haine contre certains groupes protégés, ne décrivent pas clairement ce qu'elles considèrent comme un acte de promotion ou d'incitation, ou comme un acte ciblant un groupe particulier, par exemple. Entre autres problèmes, certains concepts peuvent être difficiles à expliciter dans les règlements, comme celui de l'intention de l'utilisateur ou des répercussions de ses actions (voir [A/HRC/38/35](#), par. 26).

47. Les entreprises doivent revoir leurs règlements ou en adopter de nouveaux, en gardant le critère de légalité à l'esprit. Pour être conformes aux droits de l'homme,

⁵⁹ Voir <https://vk.com/terms>.

⁶⁰ Voir www.wechat.com/en/acceptable_use_policy.html.

⁶¹ Voir <https://support.google.com/youtube/answer/2801939?hl=fr>, https://www.facebook.com/communitystandards/hate_speech et <https://help.twitter.com/fr/rules-and-policies/hateful-conduct-policy>.

les règles relatives aux discours haineux en ligne doivent s'inspirer des définitions données dans les instruments mentionnés ci-dessus et répondre aux questions suivantes :

a) Que recoupe la notion de personnes ou groupes protégés ? Le droit des droits de l'homme désigne les groupes de population qui ont besoin d'une protection particulière. Les entreprises du secteur des TIC doivent tâcher de leur offrir la plus large protection possible, en suivant l'évolution des lois et normes dans le domaine. Elles doivent indiquer clairement qu'elles ne restreindront pas la promotion d'une signification positive de l'identité de groupe, en particulier dans le cas des groupes historiquement défavorisés, tout en reconnaissant que dans certains cas, par exemple chez les suprémacistes blancs, l'expression d'un sentiment d'identité de groupe peut s'apparenter à un discours haineux⁶² ;

b) Quelles formes de discours haineux violent les règles de l'entreprise ? Les entreprises doivent mettre au point des règles applicables aux discours haineux en gardant à l'esprit les types de nuisances dont les utilisateurs peuvent faire l'objet sur la plateforme. Le droit des droits de l'homme, qui pose qu'il est légitime de limiter les droits de l'individu pour protéger les droits d'autrui, peut aider à orienter la rédaction de ces règles. Les entreprises doivent notamment tenir compte de la manière dont les discours haineux en ligne peuvent inciter à commettre des actes de violence qui menacent la vie d'autrui ou l'empêchent, entre autres, d'exercer son droit à la liberté d'expression et à l'accès à l'information, son droit à la vie privée ou son droit de vote. Contrairement aux États, les entreprises ne sont pas en mesure de déterminer si des contenus menacent la sécurité nationale ou l'ordre public et doivent donc s'abstenir de prendre de leur propre chef la décision de restreindre les discours haineux pour ces raisons. Elles ne doivent prendre une telle mesure que lorsqu'elles en reçoivent l'ordre de la part des États, eux-mêmes soumis aux strictes dispositions du paragraphe 3 de l'article 19 du Pacte ;

c) L'entreprise restreint-elle la publication de certains contenus haineux en particulier ? Si les entreprises imposent des restrictions aux formes d'expression visées au paragraphe 2 de l'article 20 du Pacte et à l'article 4 de la Convention, elles doivent en décrire les modalités. Elles doivent s'inspirer de ces instruments pour établir quels types de contenus sont interdits. Les propos incitant à commettre une infraction ne sont qu'un des types de contenus problématiques pouvant constituer un discours haineux. Les entreprises doivent définir clairement tous les types de propos qui entrent dans cette catégorie, comme certaines d'entre elles l'ont déjà fait au gré de l'évolution de leurs règlements. Elles ne doivent pas s'en tenir à une simple définition mais plutôt montrer, en établissant une sorte de jurisprudence, comment les catégories qu'elles ont instaurées servent à orienter les mesures de modération (A/HRC/38/35, par. 71) ;

d) Existe-t-il des catégories d'utilisateurs non tenues de suivre les règles relatives aux discours haineux ? Les normes internationales disposent clairement que les journalistes et autres personnes qui signalent des discours haineux doivent être exempts de toute restriction ou autre mesure défavorable visant leur compte. Par ailleurs, l'application des critères du Plan d'action de Rabat relatifs au contexte dans lesquels les propos haineux sont proférés permettrait de protéger ces personnes. Les politiciens, fonctionnaires, militaires et autres personnalités ne sont pas dans le même cas. L'importance de leur rôle et leur capacité de mobilisation font qu'ils doivent être tenus aux règles relatives aux discours haineux qui figurent dans les grands instruments internationaux. Lorsqu'il est question de politiques relatives aux discours haineux, les figures influentes devraient, par défaut, être assujetties aux mêmes

⁶² Article 19, Principes de Camden sur la liberté d'expression et l'égalité, principe 12.

obligations que les autres utilisateurs. Une exception est possible, dans certains cas, après une évaluation plus approfondie du contexte, lorsque le contenu visé doit être protégé, par exemple dans le cas d'un discours politique. Cela dit, les incitations à commettre une infraction sont toujours plus néfastes lorsqu'elles proviennent de dirigeants plutôt que d'utilisateurs ordinaires, ce dont il faut tenir compte lors de l'évaluation des contenus diffusés sur les plateformes.

48. Lorsque les entreprises adoptent des règles qui diffèrent des normes internationales, elles doivent donner préalablement une explication motivée de ces différences, en les définissant clairement. Ainsi, si une entreprise venait à interdire l'utilisation d'un terme péjoratif désignant des personnes d'une certaine nationalité, race ou religion qui, à lui seul, ne serait pas sujet à restriction au regard du droit des droits de l'homme, elle devrait clarifier sa prise de position, conformément à ce même droit. Les entreprises doivent par ailleurs faire particulièrement attention à certaines formes d'utilisation abusive de leurs plateformes, notamment celles qui sont faites pour propager une désinformation qualifiable de discours haineux, en particulier dans les environnements sujets à une montée des tensions. Elles doivent, en pareil cas, communiquer clairement au sujet de leurs politiques, dialoguer avec la population et les experts pour sensibiliser l'opinion et lutter activement contre de tels actes d'incitation. Les normes internationales en matière de droits de l'homme peuvent servir d'orientation en la matière. Par ailleurs, le caractère viral que peuvent prendre les contenus haineux dans un tel environnement peut forcer les entreprises à réagir et à sonner l'alerte rapidement afin de protéger les droits fondamentaux de chacun.

49. Les entreprises doivent expliquer comment elles parviennent à la conclusion qu'un utilisateur a violé leurs règles relatives aux discours haineux. À l'heure actuelle, il est difficile de savoir quand ces règles sont violées ou non. Leur application semble manquer cruellement de cohérence, un problème dont participe l'absence de transparence qui caractérise le processus de modération. Les critères définis dans le Plan d'action de Rabat pour considérer comme criminels des actes d'incitation à commettre une infraction tels que ceux visés au paragraphe 2 de l'article 20 du Pacte devraient également être pris en compte par les entreprises lorsqu'elles entendent limiter la liberté d'expression de leurs utilisateurs. S'ils n'ont pas à être respectés à la lettre, comme ils le seraient dans le cadre d'une procédure pénale, ils peuvent néanmoins orienter la réflexion des entreprises lorsqu'elles cherchent à définir s'il est justifié de restreindre la diffusion d'un contenu bien précis – qu'il s'agisse du message publié ou des mots ou images qui le composent.

50. Les entreprises remarqueront toutefois qu'une analyse poussée du contexte dans lequel un message est diffusé est difficile et demande beaucoup de ressources. C'est pourquoi les plus grandes entreprises font une place importante à l'automatisation, du moins pour effectuer un premier tri des contenus afin d'y repérer les propos haineux, ce qui nécessite la création de règles régissant la classification en catégories (« à ignorer » ou « à supprimer »). Leurs systèmes de tri automatique reposent sur l'intelligence artificielle mais sont notoirement mauvais lorsqu'il s'agit de comprendre le contexte dans lequel des propos ont été tenus (voir [A/73/348](#)). Si les entreprises veulent réellement protéger les droits fondamentaux de leurs utilisateurs, elles doivent fixer des règles claires et se fier au jugement d'humains. L'évaluation par des personnes réelles ne doit pas se limiter à définir si certains mots appartiennent ou non à certaines catégories : elle doit reposer sur une connaissance approfondie de la communauté dans laquelle les propos haineux peuvent proliférer. Autrement dit, les modérateurs doivent pouvoir comprendre les codes du langage des utilisateurs, qui peut être employé pour dissimuler les appels à la violence, apprécier les intentions de l'auteur du message, prendre en compte la nature de cette personne et de son auditoire et évaluer les environnements dans lesquels les discours haineux peuvent mener à la violence. Rien de tout cela n'est possible si les entreprises n'ont recours qu'à

l'intelligence artificielle : les définitions qu'elles donnent de certains concepts et leurs stratégies doivent tenir compte des différents aspects du problème. Les plus grandes entreprises doivent investir les ressources nécessaires et diffuser largement leurs connaissances et leurs outils, sous forme de logiciels libres, afin que les plus petites entreprises et celles qui opèrent dans de plus petits marchés aient accès à ces technologies.

Nécessité et proportionnalité

51. Les entreprises ont les moyens de gérer les contenus dans le respect des droits de l'homme et, à certains égards, disposent d'un éventail d'outils plus large que les États. Ces divers outils leur permettent d'adapter leur réponse aux contenus problématiques qu'elles rencontrent, en fonction de la gravité de l'infraction et d'autres critères. Elles peuvent supprimer des contenus, limiter la vitesse à laquelle ils se propagent, indiquer d'où ils proviennent, suspendre le compte de l'utilisateur responsable ou de toute organisation promouvant lesdits contenus, mettre en place un système de notes pour alerter sur la diffusion de contenus interdits, restreindre temporairement l'accès à des contenus pendant qu'une équipe les évalue, empêcher les utilisateurs de monétiser les contenus qu'ils diffusent, rendre plus difficile le partage de contenus, apposer des mises en garde aux contenus, permettre aux utilisateurs de bloquer d'autres personnes plus facilement, réduire l'amplification du contenu, entraver le fonctionnement des robots et empêcher les lynchages en ligne, mettre en place des restrictions par géolocalisation ou même lancer des campagnes de contre-propagande. Ces outils ne sont pas tous appropriés en toutes circonstances et il peut être nécessaire d'en limiter l'utilisation, mais leur diversité montre que, selon les cas, les entreprises ont bien d'autres options que la seule suppression de contenus. En d'autres termes, les entreprises doivent suivre l'exemple des États et évaluer, dans chaque cas, si la méthode de restriction de la liberté d'expression choisie est la solution la moins contraignante possible. Dans le cadre de ce travail, les entreprises doivent s'astreindre à démontrer publiquement que leur intervention était nécessaire et proportionnelle à la gravité du problème lorsqu'un utilisateur visé en fait la demande, qu'il s'agisse de la personne à l'origine du contenu, de la victime présumée, d'une tierce personne ayant consulté le contenu ou d'un membre du public.

52. Evelyn Aswad a divisé en trois étapes le processus qu'une entreprise doit suivre pour rendre son intervention aussi peu contraignante que possible : déterminer si elle dispose d'outils permettant d'atteindre un objectif légitime sans empêcher l'utilisateur de s'exprimer ; parmi les outils qui nécessitent de restreindre la liberté d'expression de l'utilisateur, choisir le moins restrictif ; évaluer les mesures prises pour voir si elles permettent d'atteindre l'objectif fixé, et en apporter la preuve⁶³. Les évaluations de ce type s'inscrivent dans la lignée des Principes directeurs relatifs aux entreprises et aux droits de l'homme, dans lesquels les entreprises sont appelées à faire en sorte de prévenir ou de limiter autant que possible les atteintes aux droits, en particulier en ce qu'elles permettent aux entreprises de mesurer les deux types d'atteintes potentielles : la restriction de la liberté d'expression causée par l'application de leurs règles et la restriction découlant des propos haineux tenus par des utilisateurs visant d'autres utilisateurs ou des membres du public. Les approches qui s'inspirent de ce système permettent aux entreprises de décider de leur réaction, non seulement aux réelles incitations à commettre des infractions, mais aussi aux propos, courants en ligne, qui sont à la limite du discours haineux sans constituer une forme d'incitation.

⁶³ Aswad, « The future of freedom », pages 47 à 52.

Réparation

53. Le corpus du droit international des droits de l'homme contient une profusion d'idées sur la façon d'offrir réparation aux victimes de propos haineux en ligne. L'article 2 du Pacte international relatif aux droits civils et politiques et l'article 6 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale disposent que les personnes dont les droits énoncés dans ces textes ont été violés doivent disposer d'une voie de recours. Les Principes directeurs relatifs aux entreprises et aux droits de l'homme disposent eux aussi que les victimes doivent avoir accès à des mesures de réparation. Dans son rapport sur la modération de contenus datant de 2018, le Rapporteur spécial a souligné que, d'après les Principes directeurs, les entreprises ont la responsabilité d'offrir de telles mesures lorsque leurs activités ont eu des incidences néfastes sur les droits de l'homme (A/HRC/38/35, par. 59). Cette question n'a donc pas à être abordée en détail dans le présent rapport. En résumé, la première étape du processus de réparation doit consister à donner aux utilisateurs les moyens de signaler de potentielles violations des règles relatives aux discours haineux ; il importe également de prévenir toute utilisation abusive du système de signalement, qui peut être assimilée à un discours haineux. Les mécanismes de contestation des décisions des plateformes doivent être transparents et faciles d'utilisation et les entreprises doivent publier une réponse détaillant les raisons ayant motivé leur intervention.

54. Les entreprises doivent, à tout le moins, décrire publiquement les sanctions auxquelles s'expose quiconque viole leurs règles relatives aux discours haineux. Suspendre le compte d'un utilisateur peut se révéler insuffisant. Elles doivent graduer leur réponse en fonction de la gravité des violations, traitant avec plus de fermeté les cas de récidive. Elles doivent concevoir des produits de qualité, qui protègent l'autonomie des utilisateurs, leur sécurité et leur droit à la liberté d'expression afin de compenser les incidences de violations passées. Elles peuvent, notamment, empêcher l'amplification ou la monétisation des contenus problématiques que, pour quelque motif que ce soit, elles ne souhaitent pas bannir, mais elles doivent au préalable expliquer clairement leurs règles à tous leurs utilisateurs en donnant des définitions faciles à comprendre, prévenir les personnes concernées et leur donner la possibilité de retirer le contenu offensant. Le cas échéant, elles doivent également offrir réparation aux personnes lésées. Elles peuvent également mettre en place des programmes destinés à contraindre les utilisateurs dont le compte a été suspendu et qui souhaitent utiliser de nouveau la plateforme à offrir divers types de réparations, soit, par exemple, à présenter des excuses aux personnes qu'elles ont offensées, soit à instaurer d'autres formes de dialogue. Elles devraient mettre en place des politiques de réparation faisant appel à l'éducation, au contre-discours, au signalement et à la formation. Dans les cas les plus graves, les mesures de réparation devraient aussi comprendre une évaluation des incidences des violations et l'élaboration de politiques visant à mettre fin à ces violations.

55. Le Plan d'action de Rabat et la résolution 16/18 du Conseil des droits de l'homme donnent également des pistes de réflexion dont les entreprises peuvent s'inspirer pour offrir réparation aux victimes de contenus haineux. Il est indiqué, dans le Plan d'action, que « les États devraient garantir que les personnes ayant été victimes d'un préjudice réel résultant d'incitation à la haine puissent bénéficier d'un recours effectif, y compris d'un recours civil ou non judiciaire, pour des réparations ». Les sanctions et recours à envisager comprennent les dommages-intérêts financiers, le droit de rectification et le droit de réponse (A/HRC/22/17/Add.4, appendice, par. 33 et 34). Dans sa résolution 16/18, le Conseil des droits de l'homme a proposé plusieurs solutions, par exemple la formation des agents de l'État et la promotion du droit des communautés minoritaires de manifester leur religion. Le précédent Rapporteur spécial a encouragé la mise en place de recours de procédure (par exemple, garantir

l'accès à la justice et veiller au bon fonctionnement des institutions nationales) et de recours quant au fond (par exemple, prévoir des réparations qui soient suffisantes, rapides et proportionnées à la gravité de l'expression, pouvant aller de la restauration de la réputation à des mesures visant à empêcher une récidive et l'octroi d'une indemnisation financière) (A/67/357, par. 48). Pour autant, il a également recommandé la mise en œuvre d'une série de mesures autres que juridiques, que les entreprises devraient évaluer et adopter, sachant que celles-ci ont une responsabilité en tant que créatrices des plateformes sur lesquelles prolifèrent les contenus haineux. Il pourrait s'agir, par exemple, de campagnes d'information sur les méfaits des discours haineux et sur la façon dont ils servent souvent à faire fuir les communautés vulnérables des plateformes et, de fait, à les réduire au silence ; d'activités visant à promouvoir et à faire mieux connaître les dispositifs de lutte contre les discours haineux ; de dénonciations publiques des discours haineux, notamment dans le cadre de messages d'intérêt public ou de déclarations faites par des personnalités influentes ; d'une collaboration plus étroite avec des chercheurs en sciences sociales visant à évaluer l'ampleur du problème et à sélectionner les outils les plus efficaces pour lutter contre la prolifération des discours haineux (ibid., par. 56 à 74).

IV. Conclusions et recommandations

56. **Le droit international des droits de l'homme devrait être considéré comme un cadre essentiel pour la protection et le respect des droits de l'homme dans le contexte de la lutte contre les propos haineux, offensants, dangereux ou dénigrants. L'expression « discours haineux en ligne » recouvre de nombreuses formes d'expression décrites dans le présent rapport, qui peuvent avoir des effets pernicioeux. Cela dit, l'utilisation abusive de ce terme peut permettre aux États mal intentionnés de punir et de restreindre des formes d'expression tout à fait légitimes, et même nécessaires, dans des sociétés respectueuses des droits de chacun. Cependant, certains types de discours peuvent bel et bien avoir des incidences négatives. Ils peuvent intimider les communautés vulnérables et les réduire au silence, en particulier dans le cas des appels à la haine qui incitent à l'hostilité, à la discrimination ou à la violence. Si rien n'est fait pour les réprimer et si on les laisse se propager à un rythme effréné, ils peuvent créer un climat peu propice au débat public et même faire tort à des personnes qui n'utilisent pas la plateforme concernée. Il importe donc que les États et les entreprises s'attaquent au problème des discours haineux avec détermination, afin de protéger les personnes à risque et de promouvoir des débats ouverts et approfondis sur toutes les questions d'intérêt général, même les plus sensibles.**

Recommandations à l'intention des États

57. **Les États devraient aborder le problème des discours haineux sous deux angles. Premièrement, ils devraient faire en sorte que les droits de l'homme soient protégés dans le cadre des échanges en ligne, comme ils le sont lors d'échanges en personne. Les propos haineux, quels qu'ils soient, ne doivent en aucun cas être punis plus durement lorsqu'ils sont tenus en ligne que lorsqu'ils sont tenus en personne. Deuxièmement, les gouvernements ne devraient pas exiger d'intermédiaires qu'ils prennent des mesures, sous menace de sanctions judiciaires ou extrajudiciaires, si le droit international des droits de l'homme interdit aux États de prendre ces mêmes mesures. Conformément à ces principes et aux règles mentionnées ci-dessus, les États devraient au moins prendre les mesures qui suivent pour lutter contre les discours haineux en ligne :**

a) **Définir rigoureusement dans leurs lois les formes d'expression interdites au titre du paragraphe 2 de l'article 20 du Pacte international relatif**

aux droits civils et politiques et de l'article 4 de la Convention internationale sur l'élimination de toutes les formes de discrimination raciale, s'abstenir d'ériger en infraction pénale l'emploi de ces expressions, sauf dans les cas les plus graves (comme celui des appels à la haine nationale, raciale ou religieuse constitutifs d'une incitation à la discrimination, à l'hostilité ou à la violence) et relayer les interprétations du droit des droits de l'homme présentées dans le Plan d'action de Rabat ;

b) Revoir leurs lois sur les discours haineux ou en adopter de nouvelles afin de les mettre en conformité avec les critères de légalité, de nécessité et de proportionnalité ainsi que de légitimité, en veillant à ce que le public participe activement à ce processus ;

c) Prendre, ou envisager activement de prendre, des mesures de bonne gouvernance, notamment celles qui font l'objet de recommandations dans la résolution 16/18 du Conseil des droits de l'homme et dans le Plan d'action de Rabat, afin de lutter contre les discours haineux et d'atténuer ainsi le sentiment qu'il est nécessaire de limiter la liberté d'expression ;

d) Revoir les textes régissant la responsabilité des intermédiaires, ou en adopter de nouveaux, pour les mettre en totale conformité avec les normes relatives aux droits de l'homme, et s'abstenir d'exiger des entreprises qu'elles restreignent la liberté d'expression de leurs utilisateurs s'ils ne sont pas eux-mêmes en droit d'imposer directement ces restrictions par voie législative ;

e) Mettre en place des mécanismes judiciaires indépendants, ou renforcer les mécanismes existants, chargés de garantir que chaque personne puisse avoir accès à la justice et à des voies de recours lorsqu'elle subit les préjudices donnant lieu à réparation visés au paragraphe 2 de l'article 20 du Pacte ou à l'article 4 de la Convention ;

f) Adopter des lois astreignant les entreprises à donner publiquement une définition détaillée de ce qu'elles considèrent comme des discours haineux, à décrire en pareils termes la manière dont elles appliquent leurs règles relatives à ces discours et à garder trace, dans des bases de données, des mesures qu'elles prennent contre quiconque tient de tels propos, et encourager par ailleurs les entreprises à tenir compte des normes en matière de droits de l'homme dans leurs règlements ;

g) Participer activement aux manifestations internationales visant à mieux faire connaître les moyens de lutter contre les discours haineux.

Recommandations à l'intention des entreprises

58. Depuis trop longtemps, les entreprises ne font guère cas du droit des droits de l'homme pour orienter leurs règlements, ce qui ne les empêche pas d'avoir une influence remarquable sur les droits fondamentaux de leurs utilisateurs et du public. Conscientes des principes adoptés dans les précédents rapports, et dans la logique des Principes directeurs relatifs aux entreprises et aux droits de l'homme, toutes les entreprises du secteur des TIC devraient, en outre :

a) Évaluer, dans le cadre d'études d'impact sur les droits de l'homme publiées régulièrement et publiquement, la manière dont leurs produits et services lèsent les droits fondamentaux de leurs utilisateurs et du public ;

b) Adopter des politiques de contenu assorties de règles relatives aux discours haineux qui tirent leur fondement du droit international des droits de l'homme, indiquer que ces règles seront appliquées

conformément aux dispositions du droit international des droits de l'homme, dont les traités pertinents de l'Organisation des Nations Unies et les textes dans lesquels figurent l'interprétation qu'en ont donnée les organes conventionnels, les titulaires de mandats au titre de procédures spéciales et d'autres experts, tels que le Plan d'action de Rabat ;

c) Définir clairement quels types de contenu elles considèrent comme un discours haineux en donnant une explication motivée à leurs utilisateurs et au public, et adopter une démarche cohérente dans toutes les juridictions ;

d) S'assurer que l'application des règles relatives aux discours haineux se fasse compte systématiquement tenu du contexte et du préjudice porté aux utilisateurs et au public, notamment en garantissant que toute décision prise par un outil d'automatisation ou une intelligence artificielle soit validée par un être humain ;

e) Veiller à ce que les groupes de personnes les plus touchés par les contenus qualifiés de discours haineux participent aux analyses contextuelles et à ce qu'ils contribuent à recenser les outils permettant de lutter le plus efficacement contre les méfaits de tels discours sur les plateformes ;

f) Dans le cadre d'un effort global de lutte contre les discours haineux, mettre au point des outils de promotion de l'autonomie, de la sécurité et de la liberté d'expression de l'individu et, selon qu'il conviendra, préférer la limitation de l'amplification, la démonétisation, l'éducation, le contre-discours, le signalement et la formation à la suspension de comptes d'utilisateurs et à la suppression de contenus.