

Résolution 2590 (2025)¹

Réglementer la modération de contenu sur les réseaux sociaux pour sauvegarder la liberté d'expression

Assemblée parlementaire

1. Les réseaux sociaux sont devenus un espace en ligne où les utilisateurs et utilisatrices viennent exercer leur droit à la liberté d'expression et d'information de nombreuses manières. Ces plateformes permettent aux personnes notamment de publier leurs propres contenus, de profiter de contenus publiés par d'autres, de s'informer et d'informer autrui, et de communiquer avec d'autres utilisateurs et utilisatrices.
2. Le droit à la liberté d'expression n'étant pas un droit absolu, les réseaux sociaux sont légalement tenus de supprimer tout contenu illégal lorsqu'ils apprennent son existence sur leurs services ou en prennent connaissance. Il incombe en outre aux réseaux sociaux de lutter contre la diffusion de contenus préjudiciables.
3. Les entreprises de réseaux sociaux sont également porteuses de droits fondamentaux, notamment le droit de propriété et la liberté d'entreprise; elles ont donc leur mot à dire sur la manière dont les utilisateurs et utilisatrices peuvent recourir à leurs services et sur les contenus qu'ils et elles peuvent publier. Les règles de modération de contenu qui sont incluses dans leurs conditions générales permettent aux entreprises de réseaux sociaux de déclasser ou de démonétiser un contenu précis, d'en restreindre l'accès ou de le supprimer en raison de son incompatibilité avec ces conditions. Dans des cas extrêmes, les entreprises de réseaux sociaux peuvent suspendre, voire supprimer, le compte d'un utilisateur. Leurs conditions générales ont un caractère contractuel et sont «à prendre ou à laisser» pour l'utilisateur ou l'utilisatrice.
4. Les grandes entreprises de réseaux sociaux, qui sont principalement situées aux États-Unis, sont d'envergure mondiale; leurs politiques de modération de contenu et leurs décisions commerciales ou idéologiques concernant les contenus à promouvoir ou à déclasser peuvent exercer une influence considérable sur l'opinion publique et les choix de milliards de personnes. Il leur incombe néanmoins de respecter les lois du pays dans lequel elles fournissent leurs services.
5. Compte tenu de l'impact potentiel que les flux d'information et de communication qui circulent sur les réseaux sociaux peuvent avoir sur les comportements sociétaux et le bon fonctionnement des processus démocratiques, il incombe à l'État de mettre en place les principes fondamentaux et le cadre institutionnel susceptibles de corriger le déséquilibre des pouvoirs qui découle de cette relation contractuelle inégale et d'assurer la protection effective du droit à la liberté d'expression.
6. Il est cependant indispensable que la réglementation publique de la modération de contenu n'ait pas d'effet dissuasif sur la liberté d'expression et n'ait pas pour objet d'imposer les vues du pouvoir politique en place et de censurer les opinions ou les idées susceptibles d'entrer en conflit avec les intérêts acquis de la majorité au pouvoir. En outre, les réglementations nationales ne devraient pas imposer de contraintes excessives aux réseaux sociaux qui pourraient se traduire par un excès de zèle dans la suppression de contenus. Ces réglementations et leur mise en œuvre doivent respecter la liberté d'expression et évaluer soigneusement la nécessité de toute restriction.

1. *Discussion par l'Assemblée* le 30 janvier 2025 (8^e séance) (voir [Doc. 16089](#), rapport de la commission de la culture, de la science, de l'éducation et des médias, rapporteure: M^{me} Valentina Grippo). *Texte adopté par l'Assemblée* le 30 janvier 2025 (8^e séance).



7. Le risque que représentent les politiques de modération de contenu restrictives est accru par le manque de transparence de leur mise en œuvre. Les réseaux sociaux ont été accusés d'utiliser le «bannissement furtif» (en anglais *shadow banning*), une pratique qui consiste à déréférencer ou déclasser des contenus qui traitent de questions controversées sans en informer l'utilisateur ou l'utilisatrice concerné·e, ce qui rend ces contenus invisibles pour les autres utilisateurs et utilisatrices. Cette pratique sournoise et cachée devrait être interdite: elle prive les utilisateurs et utilisatrices de la possibilité de défendre efficacement leur droit à la liberté d'expression.

8. La presse et les médias en général utilisent les médias sociaux comme plateforme de diffusion d'informations au public. Il est donc essentiel que les pratiques de modération des contenus n'affectent pas indûment les contenus médiatiques et journalistiques qui respectent les règles professionnelles et le cadre réglementaire national.

9. La modération de contenu est de plus en plus souvent effectuée par des moyens automatisés. Les outils d'intelligence artificielle sont beaucoup plus efficaces que les modérateurs et modératrices humains pour traiter à grande vitesse la quantité colossale de contenus circulant sur internet et identifier les contenus interdits. Il leur manque cependant, pour l'instant, la capacité de comprendre pleinement les subtilités des échanges humains (humour, parodie, satire, etc.) et celle d'évaluer le contenu dans son contexte.

10. C'est pourquoi les modérateurs et modératrices humains doivent rester la pierre angulaire de tout système de modération de contenu et être chargés de prendre des décisions dans les cas où les systèmes automatisés ne sont pas à la hauteur de la tâche. Cependant, la modération humaine peut être biaisée et entraîner des incohérences entre les pays en raison de différences culturelles. Il est donc impératif d'établir des normes claires et exhaustives, et de garantir une formation appropriée, afin de s'assurer que tous les modérateurs et modératrices ont les connaissances requises à la fois de la législation applicable et des lignes directrices internes de l'entreprise, ainsi que de la langue et du contexte du pays d'où provient le contenu. Toutefois, en cas de conflit militaire entre deux pays, les modérateurs et modératrices d'un pays partie au conflit ne devraient pas modérer les contenus provenant de l'autre pays.

11. Malheureusement, les conditions de travail des modérateurs et modératrices humains sont inadaptées malgré leur rôle essentiel. Ces personnes sont surexposées à des contenus dérangeants qui peuvent leur causer de graves problèmes de santé mentale, et elles subissent des restrictions à leur liberté de s'exprimer sur les problèmes qu'elles rencontrent au travail.

12. Les outils d'intelligence artificielle générative permettent de produire des contenus synthétiques qui sont pratiquement impossibles à distinguer des contenus générés par un être humain. Ces contenus peuvent être très trompeurs, servir de moyen de désinformation et de manipulation, et inciter à la haine et à la discrimination, entre autres dangers. Il est essentiel que les utilisateurs et utilisatrices sachent que des contenus qui semblent authentiques ne le sont pas en réalité. Les techniques de tatouage numérique sont particulièrement bénéfiques à cet égard, même si elles présentent des inconvénients, notamment leur manque d'interopérabilité entre les services de réseaux sociaux.

13. Une évaluation indépendante des conditions générales, des politiques de modération de contenu et de leur mise en œuvre réalisée dans le but de définir et de promouvoir de bonnes pratiques pourrait contribuer à garantir leur conformité avec les principes qui défendent une approche de la modération de contenu fondée sur les droits humains.

14. Il est essentiel d'établir des règles claires et transparentes de règlement des conflits afin d'assurer la protection des utilisateurs et utilisatrices, et d'atténuer le risque d'être soumis à une décision de l'entreprise de réseaux sociaux qui pourrait être tendancieuse, ou d'être contraint d'intenter une action en justice coûteuse contre une multinationale disposant d'énormes ressources financières.

15. La mise en place d'organes indépendants de règlement extrajudiciaire des litiges, chargés d'évaluer les décisions de modération de contenu, peut s'avérer judicieuse pour renforcer le respect des droits fondamentaux. La collaboration entre les entreprises de réseaux sociaux pour la mise en place de tels organes pourrait également faciliter le règlement des litiges.

16. Comme l'a indiqué l'Assemblée parlementaire dans sa [Résolution 2281 \(2019\)](#) «Médias sociaux: créateurs de liens sociaux ou menaces pour les droits humains?», les entreprises de médias sociaux devraient utiliser des algorithmes qui favorisent la diversité des sources, des sujets et des points de vue, garantir la qualité des informations disponibles et réduire ainsi les risques de «bulles de filtrage» et de «chambres d'écho».

17. Au vu de ces éléments, l'Assemblée appelle les États membres du Conseil de l'Europe à réviser leur législation afin de mieux protéger le droit à la liberté d'expression sur les réseaux sociaux. À cet égard, les États devraient en particulier:

17.1. exiger que les réseaux sociaux respectent les droits fondamentaux des utilisateurs et utilisatrices, notamment la liberté d'expression, dans leur politique de modération de contenu et leurs pratiques de mise en œuvre;

17.2. exiger que les plateformes de médias sociaux justifient toute mesure prise pour modérer les contenus fournis par la presse ou les fournisseurs de services de médias avant sa mise en œuvre, et leur donner la possibilité de répondre dans un délai approprié;

17.3. en coopération avec les organes de presse ou organisations de médias, mettre en œuvre un système de vérification des comptes des médias et des journalistes, ainsi que des mécanismes solides pour les protéger contre le harcèlement en ligne, les piratages et la fraude, et élaborer des lignes directrices en matière de médias sociaux à l'attention des organes de presse ou organisations de médias sur la publication d'informations concernant des questions sensibles, en vue d'éviter des restrictions de modération inutiles pour ce type de contenu;

17.4. prévoir des normes minimales encadrant les conditions de travail des modérateurs et modératrices humains, y compris l'exigence d'une formation adéquate pour mener à bien leurs tâches souvent stressantes et l'accès à un soutien psychologique approprié et à des soins de santé mentale en cas de besoin;

17.5. signer et ratifier la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit (STCE n° 225) et adopter ou maintenir des mesures pour veiller à la mise en place d'obligations de transparence et de contrôle adaptées aux contextes et aux risques spécifiques pour relever les défis de l'identification de contenu généré par des systèmes d'intelligence artificielle;

17.6. exiger que les contenus générés par l'intelligence artificielle soient identifiés comme tels par les personnes qui les publient initialement et que les réseaux sociaux mettent en œuvre des solutions techniques permettant aux utilisateurs et utilisatrices d'identifier facilement ces contenus, et encourager la collaboration entre les entreprises de réseaux sociaux pour assurer l'interopérabilité des techniques de tatouage numérique pour les contenus générés par l'intelligence artificielle;

17.7. exiger que les organes de règlement extrajudiciaire des litiges, lorsqu'ils sont en place, soient indépendants et impartiaux, disposent de l'expertise nécessaire, soient facilement accessibles, fonctionnent selon des règles claires et équitables, et fassent en sorte que ces exigences soient certifiées par l'autorité de régulation nationale compétente;

17.8. promouvoir, au sein du Forum sur la gouvernance de l'internet et du Dialogue européen sur la gouvernance de l'internet, une réflexion sur la possibilité pour la communauté internet d'élaborer, dans le cadre d'un processus collaboratif et, s'il y a lieu, multipartite, un système d'évaluation et d'audit externe visant à déterminer si les algorithmes ne sont pas biaisés et s'ils respectent le droit à la liberté d'expression, ainsi qu'un «label de bonnes pratiques» qui pourrait être octroyé aux réseaux sociaux dont les algorithmes sont conçus pour réduire le risque de «bulles de filtrage» et de «chambres d'écho», et pour favoriser un environnement qui offre aux utilisateurs et utilisatrices une expérience pluraliste sur le plan idéologique, tout en étant sûre.

18. L'Assemblée appelle les entreprises de réseaux sociaux à éviter de prendre des mesures qui restreignent inutilement la liberté d'expression des utilisateurs et utilisatrices. Ces entreprises devraient notamment:

18.1. incorporer directement des principes juridiques fondamentaux, en particulier la liberté d'expression, dans leurs conditions générales;

18.2. faire preuve de prudence lors de la modération de contenus qui ne sont pas manifestement illégaux;

18.3. fournir aux utilisateurs et utilisatrices des conditions générales accessibles, claires et informatives sur les types de contenu autorisés sur leurs services et sur les conséquences en cas de non-conformité, et compréhensibles pour un large éventail d'utilisateurs et utilisatrices, quel que soit leur niveau d'alphabétisation numérique et de maîtrise de la lecture;

- 18.4. informer rapidement les utilisateurs et utilisatrices que des mesures de modération ont été appliquées à leur contenu, en fournissant un exposé complet des motifs de la décision, accompagné d'une référence aux règles internes qui ont été appliquées;
- 18.5. s'abstenir d'appliquer le bannissement furtif aux contenus des utilisateurs et utilisatrices, et informer ces personnes à chaque fois qu'une décision de déclasser ou de déréférencement a été prise;
- 18.6. veiller à ce que les processus automatisés de modération de contenu fassent l'objet d'une surveillance humaine et d'une évaluation rigoureuse et continue de leur fonctionnement;
- 18.7. mettre à disposition un système de traitement des réclamations qui soit facilement accessible, aisé à utiliser et qui permette aux utilisateurs et utilisatrices de déposer une réclamation précise;
- 18.8. fournir aux modérateurs et modératrices humains une formation appropriée et des conditions de travail qui tiennent compte du stress psychologique important auquel ils sont soumis, et assurer une protection adéquate de leur santé;
- 18.9. s'abstenir de supprimer définitivement le contenu (y compris ses métadonnées) qui a été retiré conformément aux obligations légales ou aux conditions générales, en particulier lorsque le contenu en question peut servir de preuve de crimes de guerre ou d'autres crimes;
- 18.10. veiller à ce que les systèmes d'intelligence artificielle qu'elles développent ou utilisent respectent les normes du Conseil de l'Europe, y compris la nouvelle Convention-cadre sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit; les algorithmes devraient être conçus pour respecter le droit à la liberté d'expression et encourager la pluralité et la diversité des points de vues et des opinions tout en garantissant une expérience d'utilisation sécurisée; leurs modalités de fonctionnement devraient être divulguées et les utilisateurs et utilisatrices dûment informés sur la façon dont ces algorithmes filtrent et promeuvent les contenus;
- 18.11. collaborer avec d'autres services en ligne afin de garantir l'interopérabilité des techniques de tatouage électronique pour les contenus générés par l'intelligence artificielle;
- 18.12. collaborer avec les journalistes et les organisations de vérification des faits pour lutter efficacement contre la désinformation, sur la base d'informations qui respectent les normes éthiques et professionnelles du journalisme;
- 18.13. promouvoir et soutenir la création d'organes indépendants de règlement extrajudiciaire des litiges, et se conformer à leurs décisions et recommandations;
- 18.14. soutenir le travail des organismes de contrôle tiers indépendants et se conformer à leurs décisions et recommandations;
- 18.15. veiller à ce que les décisions relatives à la modération de contenu soient dûment motivées et que les chercheurs aient accès à des informations complètes sur le fondement juridique et le raisonnement qui sous-tend chaque décision.