Comité permanent de la justice et des droits de la personne
Étude de la haine en ligne
Mémoire

Lutter contre la haine en ligne : une autre approche

Alyssa Blank, M.A.

Alyssa Bank est titulaire d'une maîtrise ès arts en développement international et mondialisation de l'Université d'Ottawa, où elle a axé ses études sur la haine en ligne et sur les opérations psychologiques de la guerre. Elle possède une vaste expérience de l'analyse et de la neutralisation de la haine en ligne et hors ligne.

Résumé

Le présent mémoire propose des mesures d'intervention en ligne et hors ligne dans le but de lutter efficacement contre la haine en ligne. À cette fin, il met à profit les données dont disposent les réseaux sociaux et qui sont tirées de leurs conditions d'utilisation afin de proposer une autre solution pour supprimer le contenu offensant de sources en ligne. Il affirme que le gouvernement canadien devrait s'efforcer de :

- travailler avec les réseaux sociaux pour déterminer la provenance géographique du contenu offensant en ligne, lequel a été signalé par des utilisateurs;
- travailler avec des spécialistes pour déterminer les réalités sociales et/ou les tensions dans ces secteurs hors ligne;
- mettre en œuvre des programmes adaptés dans ces régions.

Cette structure permettra de créer un mécanisme d'évaluation en temps réel pour les programmes visant à contrer l'extrémisme violent – ce que les praticiens n'ont pas à l'heure actuelle – tout en encourageant la mise au point par les réseaux sociaux d'outils pratiques novateurs servant à évaluer le contenu qu'ils recueillent et diffusent actuellement.

Faire face à la haine en ligne non contrôlée

La discussion complexe sur la façon de lutter contre la haine en ligne a été ravivée au cours des derniers mois, alors que le monde entier a été témoin d'attaques contre des communautés religieuses, notamment contre des juifs qui étaient en train de prier aux États-Unis en octobre 2018. Au moment où les communautés religieuses reprenaient leur souffle, des musulmans ont été ciblés au moment de la prière en Nouvelle-Zélande, des chrétiens ont été la cible d'attaques en pleine messe au Sri Lanka, et des juifs, encore une fois, ont été ciblés pendant qu'ils priaient aux États-Unis. La plupart de ces attaques comprenaient une dimension en ligne, ce qui a amené de nombreuses personnes à tenter de trouver une façon de modérer le contenu en ligne afin de protéger la société, sans brimer la liberté d'expression.

Bon nombre de ces préoccupations concernent directement les réseaux sociaux qui, très souvent, publient involontairement ces messages violents ou haineux. Compte tenu des efforts déployés à l'échelle internationale pour tenter de supprimer ces messages, ou d'en empêcher carrément la publication, le message fondamental est que l'Internet est devenu incontrôlable et qu'il fait la promotion de la haine. On pourrait, bien sûr, prétendre que l'Internet et ses réseaux sociaux permettent une certaine liberté de haine non contrôlée. Il est également vrai que l'Internet reflète les expériences et les impressions en temps réel de ses utilisateurs dans le monde entier. Effectivement, il est essentiel de reconnaître que la haine hors ligne s'exprime en ligne, ce qui peut faciliter la planification ou faire partie d'un plan d'actes de violence dans la réalité. La haine en ligne n'existe donc pas dans un vide, mais elle est plutôt le reflet de facteurs sociaux externes, qui peuvent varier sur le plan géographique.

Le désir de supprimer les propos haineux tenus en ligne est compréhensible, mais on peut aussi voir l'expression de ces propos comme une occasion de travailler avec les plateformes de médias sociaux concernées afin d'explorer, de déterminer et d'évaluer efficacement la réalité de la haine dangereuse là où elle se trouve : hors ligne. Dans cette optique, le présent mémoire propose que pour lutter efficacement contre la haine en ligne, le gouvernement du Canada a l'occasion d'adopter une approche à plusieurs volets en collaboration avec les réseaux sociaux afin d'apprendre les uns des autres et de trouver une façon d'établir des liens entre les réalités en ligne et hors ligne pour lutter contre la haine en ligne et hors ligne.

Défis et modèles européens

Dans le monde entier, de nombreuses initiatives existent déjà ou sont envisagées pour lutter contre la haine en ligne, et certains modèles européens sont très instructifs pour le Canada. C'est le cas de l'approche générale du consentement préalable adopté par l'Union européenne, de l'approche rigoureuse et de grande portée que préconise l'Allemagne, et de la proposition globale naissante actuellement à l'étude au Royaume-Uni.

En 2016, le commissaire européen a élaboré un *code de conduite pour lutter contre les propos haineux illégaux en ligne*, qui exige des entreprises en ligne consentantes qu'elles évaluent en moins de 24 heures les signalements de contenu offensant effectués par des utilisateurs et ensuite, conformément à la législation nationale et à celle de l'UE, supprimer le contenu illégal[1]. Bien qu'elle soit assujettie à l'approche adoptée par l'UE, l'Allemagne a aussi adopté en 2018 sa Loi de conformité des réseaux (NetzDG), qui exige que les réseaux sociaux comptant plus de deux millions d'utilisateurs allemands retirent le contenu offensant en ligne dans un délai de 24 heures (ou, dans les cas les plus compliqués, en moins de sept jours). Les contrevenants à la loi seraient passibles d'une amende pouvant atteindre 50 millions d'euros[2]. Enfin, depuis avril 2019, le Royaume-Uni tente d'établir un cadre réglementaire à plusieurs facettes pour garantir la sécurité en ligne au moyen d'un processus de consultation permanent. Cette approche vise essentiellement à assurer une surveillance des réseaux sociaux et à faciliter la collaboration avec les réseaux.

La principale critique des approches qui mettent l'accent sur le retrait de contenu en ligne est la suivante : une telle intervention contrevient à la liberté d'expression, et lesrestrictions à l'égard d'un droit démocratique fondamental sont laissées à la discrétion de réseaux sociaux appartenant à des intérêts privés. C'est une préoccupation qui a été soulevée par le grand public et les réseaux sociaux. En effet, en ce qui concerne la loi NetzDG, le vice-président aux communications et aux politiques publiques de Facebook, Elliot Schrage, a affirmé que, pour l'entreprise, « l'idée du gouvernement allemand de confier à l'externe la décision de ce qui est légal ou illégal est mauvaise ».

Dans cette optique, le Canada a l'occasion de tirer des leçons de pays pairs qui ont déjà mis en œuvre des modèles s'appliquant particulièrement à leur territoire, ainsi que des défis qui se présentent pour élaborer une approche toute canadienne.

Une marche à suivre : une approche canadienne axée sur la collaboration

Pour qu'une approche permette de contrer efficacement l'extrémisme violent, il est fondamental de reconnaître que la haine en ligne est un reflet de la réalité hors ligne. Une approche fondée principalement sur la suppression des messages haineux constituerait une tâche interminable qui ne ferait pas disparaître la cause fondamentale du problème. De plus, le fait de punir les réseaux sociaux parce qu'ils nuisent au bien-être social – en leur imposant des amendes ou des lois – est une approche à courte vue qui ne règle pas efficacement la question des discours haineux.

[1] Commission européenne, 2019, Lutte contre les discours de haine illégaux en ligne – le code de conduite de l'UE permet de réagir rapidement, Commission européenne. http://europa.eu/rapid/press-release_IP-19-805_fr.htm
[2] Dodds, Laurence, 2018, British MPs call for German-style law to block hate speech on social media. The Telegraph. https://www.telegraph.co.uk/technology/2018/07/28/british-mps-call-german-style-law-block-hate-speech-social-media/

Une approche efficace pour s'attaquer à la haine en ligne et hors ligne serait, par conséquent, fondée sur une collaboration entre le gouvernement du Canada et les réseaux sociaux, qui comprendrait les efforts suivants :

- Repérer les discours haineux :
    - le gouvernement du Canada doit définir ce concept;
    - les réseaux sociaux doivent aider le gouvernement à assurer un suivi de ces discours, notamment au moyen d'une surveillance gouvernementale.
- Déterminer l'emplacement dans le monde réel des discours haineux en ligne :
    - le but des discours haineux peut varier d'une région à l'autre; en déterminant l'origine d'un discours haineux, on aura une meilleure idée de sa raison d'être et de la façon de l'éliminer;
    - les réseaux sociaux recueillent et échangent actuellement des renseignements similaires, tout en conservant l'anonymat des utilisateurs (voir l'Annexe 1).
- Analyser les réalités régionales des messages publiés :
    - les praticiens de la lutte à l'extrémisme violent peuvent mener une enquête sur les facteurs sociaux dans les régions déterminées à la suite des renseignements recueillis sur leur emplacement par des réseaux sociaux.
- Créer des programmes pour s'attaquer aux principaux facteurs recensés par les praticiens de la lutte à l'extrémisme violent, s'attaquant du même coup à la haine en ligne.
- Voir à ce que les praticiens de la lutte à l'extrémisme violent évaluent l'efficacité des programmes avec l'aide des données fournies par les réseaux sociaux.

Les messages haineux publiés dans les médias sociaux et leur signalement mettent en lumière des tensions – au sujet desquelles les réseaux sociaux recueillent des données et des renseignements sur leur emplacement – qui peuvent être portées à l'attention des gouvernements, de la même façon que les réseaux sociaux font part de leur préférence en matière de contenu à leurs annonceurs (voir l'Annexe 1).

En recueillant ces données, le gouvernement pourrait avoir une meilleure compréhension des réalités des régions d'où émanent les messages publiés. Il serait alors possible d'analyser hors ligne des facteurs comme les taux de chômage, la répartition des ressources, les niveaux de pauvreté, les taux de criminalité, le multiculturalisme, l'éducation, l'isolement social, etc., et de mettre en place des programmes adaptés afin de contrer la hausse de la haine en ligne et, en définitive, de prendre des mesures hors ligne.

En s'attaquant à la haine de cette façon, on permettra aux praticiens de la lutte contre l'extrémisme violent de mettre en œuvre un mécanisme d'évaluation en établissant une base de référence des tensions, définies au moyen des données fournies par les réseaux sociaux, tout en respectant la vie privée et la liberté d'expression des utilisateurs.

Annexe 1 : Sections pertinentes des politiques de confidentialité et des conditions de service des médias sociaux

À l'heure actuelle, Facebook, Twitter et Instagram disposent d'un réglage par défaut, que les utilisateurs peuvent refuser, pour recueillir des données géographiques qui sont ensuite transmises à certaines organisations, notamment des annonceurs. Ces médias sociaux possèdent aussi des conditions de service qui leur permettent de transmettre des renseignements à des administrations, s'il y a lieu, et précisent leur approche concernant l'échange de renseignements à des fins de sécurité. Les passages qui s'appliquent sont les suivants :

Facebook et Instagram

- Selon la Politique d'utilisation des données de Facebook, ils utilisent « [...] des informations géographiques (comme votre position actuelle, le lieu où vous résidez, les endroits où vous aimez aller et les entreprises et personnes à proximité) afin de proposer, de personnaliser et d'améliorer nos Produits, notamment les publicités, pour vous et les autres. Ces informations géographiques peuvent provenir notamment de l'emplacement précis des appareils (si vous nous avez autorisés à les recueillir), des adresses IP et des informations concernant votre utilisation (et celle des autres) des Produits Facebook (notamment vos visites ou les évènements auxquels vous participez)[3]. »
- Facebook indique que ces renseignements sont utilisés pour améliorer la publicité, promouvoir la sécurité, et effectuer de la recherche et de l'innovation dans le but d'améliorer le réseau social. On précise :
    - « Nous utilisons les informations à notre disposition pour vérifier les comptes et les activités, pour lutter contre les comportements dangereux, pour détecter et prévenir le contenu indésirable et toutes autres expériences négatives, pour préserver l'intégrité de nos Produits et pour favoriser la sûreté et la sécurité sur les Produits Facebook et en dehors de ceux-ci. Par exemple, nous utilisons les données à notre disposition pour examiner toute activité suspecte ou toute violation de nos conditions ou de nos règlements, ou pour détecter si une personne a besoin d'aide. »
    - « Nous utilisons les informations à notre disposition (notamment celles de partenaires de recherches avec lesquels nous collaborons) pour orienter et appuyer la recherche et l'innovation sur des sujets de bien-être social général, d'avancement technologique, d'intérêt public, de santé et de bien-être. Par exemple, nous analysons les informations à notre disposition concernant les schémas de migration pendant les crises afin de faciliter les opérations de secours. »

Twitter

- Dans ses Conditions d'utilisation, Twitter précise que : « Nous nous réservons également le droit de consulter, de lire, de conserver et de divulguer toute information dans la mesure où nous l'estimons nécessaire aux fins de : (i) satisfaire à toute obligation légale ou réglementaire, procédure juridique ou demande administrative applicable ; (ii) faire respecter les Conditions, y compris en facilitant les investigations sur les éventuelles violations des présentes ; (iii) détecter, prévenir ou de quelque autre façon traiter tout problème de nature frauduleuse, sécuritaire ou technique ; (iv) répondre aux demandes d'assistance des utilisateurs ; (v) protéger les droits, les biens et la sécurité de Twitter, de ses utilisateurs et du public. Twitter ne divulgue aucune donnée personnelle à des tierces parties, à moins que notre politique de confidentialité ne le permette[4]. »

[3] Politique d'utilisation des données de Facebook, https://www.facebook.com/about/privacy/update; Conditions d'utilisation d'Instagram, https://help.instagram.com/581066165581870

[4] Twitter Conditions d'utilisation, https://twitter.com/fr/tos

- Selon sa Politique de confidentialité, Twitter recueille des renseignements géographiques en fonction des réglages de confidentialité de l'utilisateur et peut transmettre ces renseignements à n'importe quelle partie intéressée. On indique notamment ce qui suit : « Sous réserve de vos paramètres, nous pouvons recueillir, utiliser et stocker des informations supplémentaires sur votre emplacement, telles que votre position actuelle précise ou les endroits où vous avez utilisé Twitter par le passé, afin d'exploiter ou de personnaliser nos services avec un contenu plus pertinent, tel que les tendances locales, des histoires, des publicités et des suggestions de personnes à suivre[5]. »
- La politique de confidentialité explique également à quel moment des renseignements personnels seront transmis, comme suit : « Nonobstant tout élément contraire dans la présente Politique de confidentialité ou dans les contrôles que nous sommes susceptibles de vous proposer, nous pouvons préserver, utiliser ou divulguer vos données à caractère personnel si nous estimons que cela est raisonnablement nécessaire pour nous conformer à une loi, une réglementation, une demande émanant des autorités judiciaires ou gouvernementales, pour protéger la sécurité d'une personne, pour protéger la sécurité ou l'intégrité de notre plateforme, y compris pour éviter le spam, l'abus, les acteurs malveillants sur nos services, ou pour expliquer la raison pour laquelle nous avons éliminé un contenu ou des comptes de nos services, pour résoudre des problèmes de fraude, de sécurité ou d'ordre technique, ou pour protéger nos droits et notre propriété ou les droits et la propriété des utilisateurs de nos services. Toutefois, aucun élément de la présente Politique de confidentialité ne saurait limiter toute défense ou objection légale que vous pourriez avoir quant à une demande de divulgation de vos données à caractère personnel émanant d'un tiers, y compris des autorités publiques. »

[5] Twitter Conditions d'utilisation, https://twitter.com/fr/tos