COMMISSION SUR L'ÉTAT D'URGENCE | PUBLIC ORDER EMERGENCY COMMISSION

Document commandé:

La mésinformation, la désinformation, la malinformation et le Convoi : un examen des rôles et des responsabilités des médias sociaux

Préparé par: Emily Laidlaw

Avis au lecteur

Conformément aux règles 5 à 10 des *Règles de pratique et de procédure de la phase relative aux politiques* de la Commission, le commissaire peut, à sa discrétion, faire appel à des experts externes pour produire des documents de discussion, de recherche ou d'orientation (« documents commandés »)

Les points de vue exprimés dans un document commandé sont ceux des auteurs et ne reflètent pas nécessairement ceux du commissaire. Les énoncés de faits contenus dans un document commandé ne représentent pas nécessairement le point de vue du commissaire. Les conclusions de fait du commissaire sont fondées sur la preuve présentée lors des audiences de la Commission.

Les parties et les membres du public peuvent fournir des commentaires écrits à la Commission en réponse à ce document. Des informations sur le processus de dépôt d'observations, y compris les dates limite, sont énoncées dans l'*Avis concernant la phase politique de la Commission* (disponible sur le site Web de la Commission).

La mésinformation, la désinformation, la malinformation et le Convoi : un examen des rôles et des responsabilités des médias sociaux

Emily B. Laidlaw, Chaire de recherche du Canada en droit de la cybersécurité, professeure agrégée, Faculté de droit, Université de Calgary
emily.laidlaw@ucalgary.ca

Document rédigé à l'intention de la Commission sur l'état d'urgence
Septembre 2022

Table des matières

Résumé *3*

Partie I Le Convoi et l'environnement de l'information en ligne *4*

Les médias sociaux et le Convoi **4**

Définitions de la mésinformation, de la désinformation et de la malinformation **7**

L'ABC-D de l'environnement de l'information **10**

A pour acteurs manipulateurs 11

B pour techniques behaviorales et comportementales trompeuses 12

C pour contenu nuisible 15

D pour distribution 16

La psychologie et les dangers de la manipulation de l'information **17**

Partie II Liberté d'expression et droits et responsabilités des utilisateurs *22*

Liberté d'expression **23**

Lois canadiennes sur la désinformation **28**

Partie III Le droit et la gouvernance des médias sociaux *31*

Aperçu des aspects juridiques **32**

Modération du contenu par les médias sociaux **37**

La technologie de la modération du contenu 40

Politiques de modération du contenu 43

Réforme des lois **50**

Conclusion *53*

Résumé

Le présent document soutient la Commission sur l'état d'urgence chargée d'examiner « les effets, le rôle et les sources de la désinformation et de la mésinformation, notamment l'utilisation de médias sociaux »[1]. L'expression « médias sociaux » est utilisée au sens large dans ce document pour désigner des applications conçues pour donner la possibilité d'interagir avec des tiers, de créer et de partager du contenu, notamment des messages, des vidéos, de l'audio et des images.

Ce document ne fait pas de constatations factuelles en ce qui concerne la manipulation de l'information en ligne et le Convoi. Son objectif est plutôt d'arriver à mieux comprendre l'environnement informationnel de la mésinformation, de la désinformation et de la malinformation, la façon dont il est réglementé et la manière dont il a été intimement lié au Convoi. Les médias sociaux ont été le système nerveux central du Convoi, et l'exploration de leur rôle recoupe de nombreux domaines, tels que le droit, la psychologie, l'histoire, la sociologie et les politiques publiques, pour n'en citer que quelques-uns. Même dans le domaine du droit, les lois applicables (et les lacunes importantes des lois) sont trop nombreuses pour être examinées en détail. Pour les lecteurs intéressés, je fournis autant de détails que possible dans les notes de bas de page, et je les encourage également à consulter les nombreuses sources citées dans ce document.

Voici la structure du présent document. La partie I examine les divers médias sociaux utilisés dans le Convoi, la signification des mots mésinformation, désinformation et malinformation, la façon dont l'information se répand, ses aspects psychologiques et ses effets. Les parties II et III examinent la manière dont est réglementée la manipulation de l'information dans les médias sociaux. Deux aspects de la réglementation sont pertinents. Premièrement, quelles sont les lois qui réglementent les utilisateurs et les autres entités qui consomment ou diffusent la mésinformation, la désinformation ou la malinformation? Il s'agit de la question de savoir, par exemple, si une personne commet un crime ou peut être civilement responsable lorsqu'elle diffuse des fausses informations. Un élément nécessaire de cette analyse est le droit à la liberté d'expression : sa valeur, son application et ses limites. Cet aspect de la réglementation est examiné à la partie II. Deuxièmement, quelles sont les responsabilités juridiques et les responsabilités de gouvernance des fournisseurs de médias sociaux en matière de mésinformation, de désinformation et de malinformation? Cet aspect est examiné à la partie III et nécessite une analyse des lois qui régissent les entreprises de médias sociaux et de la manière dont celles-ci s'autoréglementent via la modération du contenu[2].

[1] Voir a)(ii)(C) https://publicorderemergencycommission.ca/files/documents/Order-in-Council-De%CC%81cret-2022-0392.pdf.

[2] Je tiens à remercier mes assistants de recherche, Akinkunmi Akinwunmi et Sylvana Crosby, pour l'excellent travail qu'ils ont accompli en soutien à la rédaction du présent document.

Partie I Le Convoi et l'environnement de l'information en ligne

Les médias sociaux et le Convoi

Les médias sociaux ont fourni le réseau qui a donné une forme et une voix au Convoi, un mouvement par ailleurs « peu structuré » et « décentralisé »[3]. Comme le souligne Stephanie Carvin, le militantisme en ligne a été « l'élément vital du mouvement du Convoi »[4] [traduit par nos soins]. Il s'agissait d'un mouvement canadien, amplifié d'abord par des influenceurs canadiens sur les médias sociaux, et ensuite par des médias et des influenceurs américains, et exploité par d'autres acteurs dans le monde[5]. Les organisateurs et les influenceurs ont utilisé une variété de médias sociaux, comprenant des applications vidéo et audio et des applications de messagerie, pour diffuser leurs messages et communiquer avec leurs adeptes et abonnés. Ces applications incluent notamment Facebook, Twitter, TikTok, YouTube, Rumble, BitChute, Odysee, Telegram et Zello[6].

Certains aspects sont essentiels en ce qui concerne le rôle des médias sociaux dans le Convoi. Premièrement, le mouvement a commencé bien avant janvier 2022[7]. L'organisation initiale du Convoi s'est faite par l'intermédiaire d'un groupe Facebook, Canada Unity, qui a mené un convoi « United We Roll » en 2019. Avant l'organisation du Convoi, le contenu publié sur Canada Unity avait pour thèmes l'antivaccination et l'anticonfinement[8]. Il semblerait que de nombreux comptes et influenceurs du Convoi avaient des liens avec des groupes d'extrême droite comme les gilets jaunes canadiens et étaient porteurs de théories du complot[9]. Deuxièmement, le mouvement s'est construit « presque entièrement par le partage de liens

---

[3] CBC, « How anger, faith and conspiracy theories fuelled the trucker convoy » (24 février 2022) *The Fifth Estate*.
[4] Stephanie Carvin, « How the Freedom Convoy was fuelled by online activism » (5 mars 2022) *National Post,* en ligne : https://nationalpost.com/opinion/stephanie-carvin-how-the-freedom-convoy-was-fuelled-by-online-activism.
[5] *Ibid*. Des comptes du Convoi frauduleux et piratés ont été créés et supprimés par Facebook : Elizabeth Culliford, « Meta says it removed scammers' Canada convoy Facebook group » (7 février 2022) *Reuters*, en ligne : https://www.reuters.com/technology/meta-says-it-removed-scammers-canada-convoy-facebook-groups-2022-02-08/; Anya van Wagtendonk et al., « The hacked account and suspicious donations behind the Canadian trucker protests » (8 février 2022) *Grid*, en ligne : https://www.grid.news/story/misinformation/2022/02/08/the-hacked-account-and-suspicious-donations-behind-the-canadian-trucker-protests/.
[6] La liste ne répertorie pas tous les médias sociaux auxquels ont eu recours les organisateurs et les influenceurs du Convoi, mais elle donne une idée des principaux services utilisés. Voir ce résumé : Maggie Parkhill, « Who is who? A guide to the major players in the trucker convoy protest » (22 février 2022) *CTV News*, en ligne : https://www.ctvnews.ca/canada/who-is-who-a-guide-to-the-major-players-in-the-trucker-convoy-protest-1.5776441.
[7] Ryan Broderick, « How Facebook Twisted Canada's Trucker Convoy into an International Movement » (19 février 2022) *The Verge*, en ligne : https://www.theverge.com/2022/2/19/22941291/facebook-canada-trucker-convoy-gofundme-groups-viral-sharing.
[8] *Fifth Estate*, *supra* note 3.
[9] Broderick, *supra* note 7.

vidéo »[10]. Les plateformes de partage de vidéos utilisées comprennent YouTube, et les plateformes alternatives à YouTube Rumble, BitChute et Odysee, les plateformes de diffusion en direct sur Facebook et Twitter, ainsi que TikTok[11]. Comme nous le verrons plus loin à la section portant sur la psychologie et les dangers de la manipulation de l'information, les vidéos et les images sont des vecteurs particulièrement puissants d'influence des utilisateurs. Troisièmement, les messages étaient diffusés sur plusieurs plateformes, ce qui avait des répercussions sur la modération du contenu. Par exemple, des vidéos publiées sur Facebook ont entraîné un sociofinancement du Convoi et ont dirigé les utilisateurs vers GoFundMe. Une vidéo téléchargée sur Rumble, dont la modération du contenu est moins restrictive, a été partagée sur Facebook et sur l'application de messagerie Telegram. La modération du contenu sur ces plateformes est examinée à la partie III.

Le Convoi a d'abord été organisé par le groupe Facebook Canada Unity. Alors que l'exemption de l'obligation de vaccination pour les camionneurs devait prendre fin en janvier 2022, Canada Unity a commencé à publier des messages concernant les camionneurs. Puis un groupe Facebook « Convoi de la liberté 2022 » (« 2022 Freedom Convoy ») a été créé. Au moment de la rédaction de ce document, la page Facebook de Canada Unity comptait 79 877 abonnés et 34 161 mentions « j'aime »[12]. Sur Twitter, certains des messages les plus anciens remontaient au 12 janvier 2022[13]. Le 18 janvier 2022, la popularité du Convoi sur Facebook a pris de l'ampleur lorsqu'une vidéo sur le mouvement de protestation a été publiée sur Rumble[14]. La vidéo, intitulée « Convoi de la liberté 2022 », a été vue 60 588 fois sur Rumble[15]. Elle fournissait des liens vers des groupes sur Facebook, Telegram, GoFundMe et Change.org, ainsi que vers le site Web de Canada Unity.

Outre les médias sociaux couramment utilisés, les organisateurs ont eu recours à l'application de messagerie Telegram, à l'application de walkie-talkie Zello et aux plateformes de partage de vidéos Rumble, BitChute et Odysee[16]. Telegram est une application de messagerie constituée de groupes et de canaux. Les groupes sont des espaces où les membres peuvent discuter, tandis que les canaux permettent la diffusion de messages à un large public. Les groupes et les

---

[10] Ryan Broderick explique certaines de ses recherches dans son infolettre *Garbageday* publiée sur la plateforme Substack : « Freedom Convoy Facebook Content Is Coming From YouTube» (9 février 2022), en ligne : https://www.garbageday.email/p/boomers-are-weird-and-obsessive-posters.

[11] Parkhill, *supra* note 6; Press Progress, « Meet the Extremists and Social Media Influencers at the Centre of the Far-Right Siege of Ottawa » (8 février 2022), en ligne : https://pressprogress.ca/meet-the-extremists-and-social-media-influencers-at-the-centre-of-the-far-right-siege-of-ottawa/.

[12] URL du groupe Facebook de Canada Unity : https://www.facebook.com/CanadaUnity/.

[13] Voir les publications sur Twitter de Canadian for Freedom (12 janvier 2022) : https://twitter.com/CanFreedomLover/status/1481340478247346179?s=20&t=j-FIYPbSX217iiw4Z4DKPw, Dr. Ezra Kaxah (13 janvier 2022) : https://twitter.com/EzraKahan/status/1481760964043325448?s=20&t=j-FIYPbSX217iiw4Z4DKPw, et (17 janvier 2022) : https://twitter.com/EzraKahan/status/1483134186919706626, et Fringe-Juli (13 janvier 2022) : https://twitter.com/Juliz1lb/status/1481767182577160193?s=20&t=j-FIYPbSX217iiw4Z4DKPw.

[14] Ryan Broderick, *supra* note 7.

[15] Ce chiffre date du 31 août 2022.

[16] TVO Today, « How does Social Medial Fuel Protest? » (18 février 2022), en ligne: https://www.tvo.org/video/how-does-social-media-fuel-protest>.

canaux peuvent être publics ou privés. Cependant, même les groupes privés peuvent compter jusqu'à 200 000 membres, et les canaux peuvent avoir un nombre illimité d'abonnés[17]. Telegram prend en charge ce qu'elle appelle les « discussions secrètes », qui utilisent un cryptage de bout en bout, ce qui signifie que Telegram ne voit pas le contenu de ces discussions de groupe et ne le stocke pas sur ses serveurs. Seuls les participants au groupe sont au courant des discussions et des échanges[18].

L'application Zello a été utilisée par les organisateurs du Convoi pour coordonner les lieux de rencontre[19]. Zello est une application de walkie-talkie qui permet aux utilisateurs de configurer des canaux de communication publics ou privés. Les canaux privés sont cryptés de bout en bout. Les organisateurs ont principalement utilisé les canaux publics, qui sont limités à 7 000 utilisateurs[20]. Une fois qu'ils sont dans un canal, les utilisateurs peuvent entendre toutes les autres personnes qui s'y trouvent et leur parler. L'organisateur peut mettre en place plusieurs canaux et diffuser des messages sur tous les canaux en même temps. Les utilisateurs peuvent être connectés à plusieurs canaux en même temps[21]. Ils peuvent également envoyer des textos et des images. Avec Zello, les organisateurs ont créé plusieurs canaux de communication. Par exemple, pour le barrage du pont Ambassador, un canal appelé « Windsor Convoy 2 » a été utilisé. Des contre-manifestants ont utilisé certains de ces canaux et ont perturbé les communications[22]. Tout au long de la manifestation, les organisateurs ont communiqué avec leurs partisans sur divers médias sociaux en diffusant des événements en direct, en publiant des vidéos, des mèmes et des textos, et les partisans ont participé en commentant ou en partageant des publications, ou en créant leur propre contenu.

Il n'est pas possible de désigner un seul média social qui aurait joué un rôle central dans le Convoi. Ce sont plutôt tous les médias sociaux qui ont servi de système nerveux au Convoi, comme c'est le cas pour de nombreux mouvements à l'ère numérique. Comme le montre le présent document, divers facteurs ont convergé pour donner un élan au Convoi, notamment en ce qui concerne l'amplification et l'influence. La conception des médias sociaux, notamment leurs systèmes de recommandation, de publicité et de modération du contenu, amplifie certains contenus, ce qui favorise l'apparition d'influenceurs (individus et médias) et la diffusion de leurs messages. Dans le cas du Convoi, les influenceurs ont principalement publié de courtes vidéos, qui sont des moyens efficaces de manipulation de l'information, et ont utilisé des applications de messagerie, qui sont moins susceptibles d'être modérées. La nature interplateforme des communications signifie que si l'on coupait une de leurs voies, la communication pouvait être réacheminée ailleurs. Au niveau des applications, l'un des seuls

---

[17] FAQ de Telegram, en ligne : https://telegram.org/faq#q-quelle-est-la-difference-entre-les-groupes-et-les-canaux.
[18] Politique sur la vie privée de Telegram, en ligne (en anglais) : https://telegram.org/privacy.
[19] Demar Grant, « What is Zello? Inside the app that helped organize « freedom convoy » blockades » (11 février 2022) *Toronto Star*, en ligne : https://www.thestar.com/news/canada/2022/02/11/what-is-zello-inside-the-app-that-helped-organize-freedom-convoy-blockades.html.
[20] *Ibid.*
[21] Les canaux de Zello, en ligne (en anglais) : https://zello.com/product/features/channels/.
[22] Grant, *supra* note 18.

moyens d'échapper à l'élimination est la collaboration interplateforme[23]. Au niveau de l'infrastructure, c'est la conception ouverte d'Internet[24].

Définitions de la mésinformation, de la désinformation et de la malinformation

La manipulation de l'information, quelle que soit sa portée, couvre un terrain vaste et varié et « ne date pas d'aujourd'hui, mais elle est maintenant alimentée par les nouvelles technologies »[25]. La propagande, les canulars et les campagnes de dénigrement social existent depuis les premiers écrits. Et dans la mesure où les nouvelles technologies peuvent les faciliter et les amplifier, celles-ci ont été utilisées dans ce but, comme ce fut le cas de l'imprimerie de Gutenberg, des journaux et de la radio[26]. Ce qui est différent de nos jours, ce sont le caractère abordable des réseaux sociaux, ainsi que la rapidité, la portée et la précision de la communication et de la diffusion des messages, et l'accès à des outils de rédaction et de publication bon marché[27].

L'espace global de la communication « est un bien commun de l'humanité »[28]. Dans cet espace, la mésinformation, la désinformation et la malinformation fonctionnent comme un « système

---

[23] Voir toutefois Evelyn Douek, « The Rise of Content Cartels » (11 février 2020) *Knight First Amendment Institute*, en ligne : https://knightcolumbia.org/content/the-rise-of-content-cartels.

[24] Il est important de reconnaître l'architecture originale d'Internet et de comprendre la pile de protocoles Internet. Il n'y a pas qu'un seul modèle de la pile Internet. Le modèle le plus durable et le plus complet serait le modèle OSI développé par l'Organisation internationale de normalisation. Il existe d'autres intermédiaires que ceux qui ont été explorés jusqu'à présent et qui sont de plus en plus visés par la réglementation des contenus, surtout en raison des incidents de crise liés à des contenus violents et extrémistes et à la manipulation de l'information. Au sommet de la pile Internet se trouve la couche application ou couche de contenu. La plupart des débats sur la réglementation du contenu concernent cette couche, car c'est celle des médias sociaux. Les couches supérieures dans la pile Internet reposent sur des technologies situées plus en profondeur dans la pile pour fonctionner et sont affectées par leurs actions. En descendant dans la pile Internet, on trouve des technologies comme les fournisseurs d'hébergement Web (p. ex. WordPress) et les services d'informatique en nuage (p. ex. Amazon Web Services), et en dessous, les fournisseurs d'infrastructure de réseau (p. ex. les registres de noms de domaine, les fournisseurs de services Internet et les réseaux de prestation de services en nuage). Ce qu'il faut comprendre, c'est que plus on va en profondeur dans la pile, plus les actions réglementaires sont floues, imprécises et peu visibles : Georgia Evans, « Down the Stack: Power and Accountability in Internet Intermediaries' Content Moderation Decisions » (9 juillet 2021) *Kroeger Policy Review*, en ligne : https://www.kroegerpolicyreview.com/post/down-the-stack-power-and-accountability-in-internet-intermediaries-content-moderation-decisions.

[25] Cherilyn Ireton et al., *Journalisme, fake news & désinformation* (2018) UNESCO, p. 15.

[26] *Ibid* p. 15-19. Voir Heidi J.S. Tworek, *News From Germany: The Competition to Control World Communications, 1900-1945* (Harvard University Press, 2019).

[27] Samantha Bradshaw et al., « Industrialized Disinformation – 2020 Global Inventory of Organized Social Media Manipulation » (2020) *Computational Propaganda Research Project*, en ligne : https://demtech.oii.ox.ac.uk/research/posts/industrialized-disinformation/ p. 11; Claire Wardle et Hossein Derakhshan, *Les désordres de l'information, Vers un cadre interdisciplinaire pour la recherche et l'élaboration des politiques,* Rapport du Conseil de l'Europe DGI(2017)09, en ligne : https://rm.coe.int/0900001680935bd4, p. 12-13.

[28] Reporters sans frontières, *L'espace global de l'information et de la communication : un bien commun de l'humanité*, en ligne : https://rsf.org/fr/l-espace-global-de-l-information-et-de-la-communication-un-bien-commun-de-l-humanite.

complexe », dans lequel des graines sont plantées, puis amplifiées pour atteindre un public plus large et se répandre dans le vaste écosystème de l'information[29]. Il n'existe pas de définition uniforme de la mésinformation, de la désinformation et de la malinformation. Cela reflète la complexité de ces concepts et la nature contextuelle de leur application[30]. Je recommande l'utilisation des définitions de l'UNESCO :

- La désinformation est une fausse information dont la personne qui en est à l'origine connaît la fausseté. Il s'agit d'un mensonge intentionnel, délibéré, qui vise à répandre la désinformation et dont les auteurs sont des opérateurs malveillants.
- La mésinformation est une fausse information que la personne qui en est à l'origine croit vraie.
- L'information malveillante (malinformation) est une information basée sur des faits réels, mais qui est utilisée pour nuire à une personne, à une organisation ou à un pays[31].

Essentiellement, il s'agit de types de « désordres de l'information »[32]. D'autres expressions sont utilisées, telles que tromperie virale[33], chaos informationnel, propagande[34], opérations d'influence[35] et propagande algoristique, qui désigne la combinaison de plateformes, d'algorithmes, de mégadonnées et d'intelligence artificielle qui façonnent les flux d'information et manipulent l'opinion publique[36]. Souvent, le terme désinformation est utilisé comme terme

---

[29] Service canadien du renseignement de sécurité, *Qui dit quoi? Défis sécuritaires découlant de la désinformation aujourd'hui* (5 décembre 2016), en ligne : https://www.canada.ca/content/dam/csis-scrs/documents/publications/disinformation_post-report_fra.pdf, ch. 1.

[30] *Désinformation et liberté d'opinion et d'expression*, Rapport de la Rapporteuse spéciale sur la promotion et la protection du droit à la liberté d'opinion et d'expression, Irene Khan, (13 avril 2021), A/HRC/47/25, par. 9; Ronan Ó Fathaigh et al., « The perils of legally defining disinformation » (2021) 10(4) *Internet Policy Review*, p. 3.

[31] Ireton, *supra* note 25, p. 54-55.

[32] Wardle et Derakhshan, *supra* note 27. Voir schéma p. 23. Ces catégories ne s'excluent pas mutuellement : Dhanaraj Thakur et DeVan L. Hankerson, *Facts and their Discontents: A Research Agenda for Online Disinformation, Race, and Gender* (2021) Center for Democracy & Technology, en ligne : https://cdt.org/insights/facts-and-their-discontents-a-research-agenda-for-online-disinformation-race-and-gender/, p. 8.

[33] Khan, *supra* note 30, par. 13; Camille François, *Actors, Behaviors, Content: A Disinformation ABC* (20 septembre 2019), Transatlantic Working Group, en ligne : https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf, p. 1.

[34] Ireton et al. affirment que la propagande est différente de la désinformation : « Le terme propagande n'est pas synonyme de désinformation, bien que la désinformation puisse servir les intérêts de la propagande. Cependant, la propagande montre souvent de manière plus ouverte son but de manipulation, généralement parce qu'elle fait davantage appel aux émotions qu'à l'information », *supra* note 25, p. 56; Voir aussi Yochai Benkler, Robert Faris et Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford University Press, 2018). Ce livre traite de notre connexion constante à un environnement riche en propagande.

[35] Cette expression semble la plus répandue dans les documents sur la sécurité nationale. Voir SCRS, *supra* note 29; Wardle et Derakhshan définissent les opérations d'influence comme suit : « Actions entreprises par des gouvernements ou des acteurs non-étatiques organisés pour orienter le sentiment politique national ou international, le plus souvent dans un but stratégique ou géopolitique », *supra* note 27, p. 18; L'expression « fake news » ou « fausses nouvelles » n'est pas utilisée dans le présent document parce qu'elle est vague et chargée politiquement et n'est donc pas utile à l'analyse : *Ibid*, p. 15-16; Alice Marwick et Rebecca Lewis, *Media Manipulation and Disinformation Online* (15 mai 2017) Data & Society Research Institute, en ligne : https://datasociety.net/library/media-manipulation-and-disinfo-online/, p. 44.

[36] Voir le projet de propagande algoristique du Oxford Internet Institute, en ligne : https://www.oii.ox.ac.uk/research/projects/computational-propaganda/.

générique, bien que ce ne soit pas le cas dans le présent document pour éviter toute confusion. Pour plus de facilité, lorsque j'utiliserai un terme général, je parlerai de manipulation de l'information.

Les définitions varient même pour les termes mésinformation, désinformation et malinformation. Par exemple, la définition de malinformation donnée ci-dessus est axée sur l'information fondée sur la réalité qui est intentionnellement diffusée pour nuire[37]. D'autres définitions ne se basent pas sur l'intention de nuire, mais simplement sur la communication d'une information exacte dans un contexte trompeur[38]. D'autres définitions encore mettent l'accent sur la diffusion intentionnelle d'informations privées[39]. Ces différences sont importantes. Si la désinformation et la malinformation supposent toutes deux une intention de nuire, quelle est la ligne de démarcation entre une information exacte mais induisant en erreur et une fausse information? La malinformation est-elle un autre terme pour désigner le « *doxing* », qui consiste essentiellement à divulguer publiquement des données personnelles? Claire Wardle et Hossein Derakhshan, par exemple, considèrent le discours haineux comme une forme de malinformation, ce qui ne correspond à aucune des définitions données ci-dessus, même s'il est logique d'inclure le discours haineux dans ce cadre[40]. Comme l'ont constaté Samantha Bradshaw et ses coauteurs, le harcèlement est l'outil de plus en plus utilisé pour réduire au silence la presse et la dissidence politique, comme c'est le cas, par exemple, dans les centres Web au Guatemala qui utilisent de faux comptes pour cibler des individus et des journalistes en les qualifiant de terroristes et d'ennemis de l'État[41].

De même, les définitions de désinformation diffèrent sur des points importants. Par exemple, la définition de fausseté varie et peut être « information fausse », « information fausse de manière vérifiable »[42], « information trompeuse »[43] ou « information inexacte »[44]. En outre, le préjudice et l'intention requis varient. Certaines définitions parlent de préjudice causé au public, tandis que d'autres parlent de préjudice causé à des personnes, à des groupes sociaux ou à des États. D'autres encore mentionnent un gain économique, ce qui signifie que la diffusion d'une fausse information à des fins commerciales correspondrait à la définition de la désinformation[45].

---

[37] Voir Ireton et al., *supra* note 25; Kate Jones utilise une définition semblable, *Online Disinformation and Political Discourse: Applying a Human Rights Framework* (novembre 2019) Chatham House The Royal Institute of International Affairs, section 2.2.

[38] Thakur et Hankerson, *supra* note 32, p. 7.

[39] Comme la définition mentionnée dans Faithagh et al., *supra* note 30, p. 4.

[40] Wardle et Derakhshan, *supra* note 27, p. 23; Aucune tentative n'est faite ici pour définir le discours haineux, car il existe une grande variété de définitions. Voir *Code criminel* (L.R.C. (1985), ch C-46), art. 319, et *Pacte international relatif aux droits civils et politiques* (PIDCP), 1966, article 20.

[41] Bradshaw, *supra* note 27, p. 13.

[42] Fathaigh et al., *supra* note 30, p. 4.

[43] Ireton et al. donnent comme exemples de contenus trompeurs le recadrage de photos ou la sélection de citations hors contexte, ce qu'on appelle la théorie du cadrage (« Framing Theory »), *supra* note 25, p. 58

[44] Fathaigh et al., *supra* note 30, p. 5.

[45] *Ibid* p. 5-7.

Comme l'indiquent Fathaigh et ses coauteurs, ces définitions sont des catégories juridiques mal adaptées, même si elles sont utiles dans le domaine des politiques[46]. Pour une bonne compréhension du Convoi et du rôle des médias sociaux, je recommande les trois catégories suivantes. Pour les termes désinformation et mésinformation, les définitions de l'UNESCO citées ci-dessus sont instructives. La désinformation désigne la diffusion intentionnelle d'une fausse information, la personne ou l'entité qui la diffuse sachant que l'information est fausse. Il s'agit d'une catégorie pour les acteurs malveillants, tels que ceux qui mènent des campagnes de désinformation parrainées par un État ou ceux qui créent des sites Web accessibles à des abonnés pour diffuser intentionnellement de fausses informations sanitaires en vue d'un gain économique. La mésinformation désigne la diffusion intentionnelle d'une fausse information par une personne ou une entité qui la croit vraie. Une grande partie des fausses informations publiées sur les médias sociaux sont de la mésinformation, et il y a certainement un chevauchement entre la mésinformation et la désinformation, surtout en ce qui concerne le contenu trompeur.

La troisième catégorie est ce que j'appelle « tout le reste ». Les discours haineux, le harcèlement, la diffamation, les contenus violents et extrémistes, les *trolls*, etc. sont des formes d'expression ou des vecteurs d'attaque qui sont alimentés par la désinformation et la mésinformation, et en constituent le fondement. Prenons l'exemple du Gamergate, l'attaque contre les développeuses de jeux vidéo. C'est un bon exemple d'une combinaison d'attaques comprenant du *doxing*, qui consiste à obtenir des données personnelles d'un individu (par piratage ou autrement) et à les divulguer au grand public, ainsi que du harcèlement en bande organisée, qui inclut des mensonges et des attaques sexistes[47]. Lorsque la malinformation est évoquée dans le présent document, elle se situe dans la catégorie « tout le reste », et j'utilise la définition de l'UNESCO selon laquelle il s'agit de l'information fondée sur des faits réels et diffusée dans le but de nuire. Les meilleurs mensonges sont proches de la vérité, et la malinformation reflète cette réalité.

L'ABC-D de l'environnement de l'information

Le cadre ABC-D est un moyen utile pour comprendre l'environnement de la manipulation de l'information[48] :

- A pour les acteurs manipulateurs qui diffusent sciemment de la désinformation;
- B pour les techniques behaviorales (aussi appelées « comportementales » ou « comportements », ci-dessous) utilisées pour diffuser de la désinformation;
- C pour les contenus nuisibles; et
- D pour les architectures numériques des médias sociaux et leur incidence sur la distribution de l'information.

Ce cadre a été créé pour que l'on puisse déterminer les solutions à privilégier, mais il est également utile pour comprendre l'espace de l'information.

---

[46] *Ibid.*
[47] Marwick et Lewis, *supra* note 35, p. 27.
[48] François, *supra* note 33.

A pour acteurs manipulateurs

Ces acteurs lancent sciemment et secrètement une campagne de désinformation. À cet égard, de nombreux experts établissent une distinction entre la désinformation soutenue par l'État et celle des autres acteurs. Les recherches indiquent que les acteurs qui produisent de la désinformation sont motivés par « l'idéologie, l'argent, et/ou le statut et l'attention »[49] [traduit par nos soins]. Il convient de se rappeler que s'il s'agit de mésinformation, l'acteur ne diffuse pas sciemment une fausse information. En effet, l'un des grands défis que présente la manipulation de l'information est que l'information finit par être transmise à des humains, qui la croient vraie et l'amplifient dans leurs réseaux. Sur le plan stratégique, de nombreuses campagnes de désinformation ciblent les principaux influenceurs en ligne, qui diffusent ensuite le contenu dans leurs réseaux. C'est ce qu'on a pu observer dans une étude sur du contenu antivaccin, qui s'est principalement propagé par l'intermédiaire de 12 influenceurs clés en ligne[50].

En outre, les médias jouent un rôle dans la diffusion de la désinformation. Alice Marwick établit un spectre de la manipulation des médias. À une extrémité se trouvent les sites Web créés intentionnellement pour tromper les lecteurs. Ces sites sont conçus de manière à ressembler à des sources fiables et les articles sont sensationnels pour attirer les lecteurs et faire de l'argent. Au milieu du spectre, on trouve des médias, souvent motivés par une idéologie, qui publient un mélange d'histoires vraies et fausses. À l'autre extrémité du spectre se trouvent les médias grand public, qui peuvent utiliser des titres de type « piège à clics », sensationnalistes et trompeurs, pour attirer plus de lecteurs, et qui peuvent relater des fausses nouvelles, en amplifiant ainsi ces dernières par inadvertance[51]. Certains médias soutenus par l'État, comme Russia Today (RT), sont bien connus pour leur rôle dans la diffusion de la désinformation. Cette situation a conduit les principales plateformes de médias sociaux à bloquer RT de leurs services pendant la guerre entre la Russie et l'Ukraine, et a incité l'Union européenne (UE) à demander aux plateformes de bloquer l'accès à RT[52]. Cependant, les sources de nouvelles partisanes jouent également un rôle dans la diffusion de la mésinformation et de la désinformation, par leurs titres et légendes induisant en erreur[53]. Comme l'explique Stephanie Carvin, les médias de la droite américaine ont eu une incidence plus importante sur les élections de 2016 aux États-

---

[49] Marwick et Lewis, *supra* note 35, p. 7-9, 27-29.
[50] The Center for Countering Digital Hate and Anti-Vax Watch, *Disinformation Dozen: the Sequel – How Big Tech is Failing to Act on Leading Anti-Vaxxers Despite Bipartisan Calls from Congress* (2021), en ligne : https://counterhate.com/research/the-disinformation-dozen/; Voir aussi, par exemple, le ciblage des influenceurs de la communauté latino-américaine avant les élections de 2020 : Thakur et Harkeson, *supra* note 32, p. 13-15.
[51] Marwick et Lewis, *supra* note 35, p. 44-45.
[52] Communiqué de presse du Conseil de l'UE, *L'UE impose des sanctions aux médias publics RT/Russia Today et Sputnik, qui diffusent dans l'UE* (2 mars 2022), en ligne : https://www.consilium.europa.eu/fr/press/press-releases/2022/03/02/eu-imposes-sanctions-on-state-owned-outlets-rt-russia-today-and-sputnik-s-broadcasting-in-the-eu/; Voir Wardle et Derakhshan, *supra* note 27, p. 14.
[53] Voir l'étude examinée dans Wardle et Derakhshan *supra* note 27, p. 41.

Unis que la désinformation[54]. Par conséquent, les journalistes sont souvent l'une des principales cibles des producteurs de désinformation[55].

Une question que la Commission pourrait se poser est celle de savoir qui étaient les principaux acteurs des médias sociaux – individus ou médias – qui ont déclenché le mouvement du Convoi, et qui l'ont amplifié. Divers reportages ont identifié certains des principaux acteurs, ici et à l'étranger[56]. Une autre question est de savoir s'il y a eu une influence étrangère par l'intermédiaire des médias sociaux – ou quelle a été son ampleur – qu'elle soit étatique, médiatique ou individuelle[57]. Par exemple, 88 % des fonds donnés par l'intermédiaire de GoFundMe seraient venus du Canada[58]. Facebook a également supprimé certains groupes, pages et comptes du Convoi, qui attiraient les utilisateurs vers des sites Web hors plateforme au moyen de publicités payantes, et vers des groupes haineux et complotistes[59].

B pour techniques behaviorales et comportementales trompeuses

Le comportement est défini par les techniques utilisées par les acteurs pour diffuser l'information. Parmi ces techniques, on peut mentionner les suivantes.

- Les outils automatisés, comme les *bots*, qui sont des algorithmes qui extraient des données sur Internet (*web scraping*), puis diffusent des messages par l'intermédiaire de réseaux et contribuent à renforcer la viralité du contenu[60]. Tous les *bots* ne sont pas mauvais. Un *bot* du Vatican publie des réflexions. Les organismes de presse utilisent des *bots* pour publier des informations de dernière minute, et ainsi de suite. Toutefois, les *bots* peuvent également être des faux-nez (fausses identités en ligne), créés pour incarner une personne et manipuler l'opinion. Ils ne sont pas aussi faciles à détecter qu'on pourrait l'imaginer et sont conçus pour « passer sous le radar »[61]. Les outils de

---

[54] Stephanie Carvin, *Stand on Guard: Reassessing Threats to Canada's National Security* (University of Toronto Press, 2020) p. 223.
[55] Thakur et Hankerson, *supra* note 32, p. 8.
[56] Voir Parkhill, *supra* note 6; Par exemple, l'une des vidéos de Russell Brand a été vue 1 252 343 fois en date du 9 septembre 2022, « Truckers Convoy: Why The Mainstream Media Blackout?! » (27 janvier 2022) *YouTube*, en ligne : https://www.youtube.com/watch?v=itbSIqY4Nnw; Tucker Carlson a également attiré l'attention sur le Convoi : « What's happening to truckers in Canada reveals the future of the United States » (21 février 2022) *Fox News*, en ligne : https://www.foxnews.com/opinion/tucker-carlson-truckers-canada-future-united-states.
[57] Carvin, *supra* note 4.
[58] Sarah Turnbull, « GoFundMe head testifies over Freedom Convoy fundraising, says most donors were Canadian » (3 mars 2022) *CTV*, en ligne : https://www.ctvnews.ca/politics/gofundme-head-testifies-over-freedom-convoy-fundraising-says-most-donors-were-canadian-1.5804094.
[59] Culliford, *supra* note 5.
[60] « Les *bots* sont des protocoles de publication automatique utilisés pour relayer du contenu de manière programmée » [traduit par nos soins], citation de Marco Bastos et Dan Mercea, « The Accountability of Social Platforms: Lessons from a Study of Bots and Trolls in the Brexit Campaign » (2018) 376(2128) *Philosophical Transactions*.
[61] Samuel Woolley, « The Business of Computational Propaganda Needs to End » (20 septembre 2021) *Centre for International Governance Innovation*, en ligne : https://www.cigionline.org/articles/the-business-of-computational-propaganda-needs-to-end/; Bastos et Mercea, *supra* note 56.

prédiction de texte s'améliorent et peuvent produire du contenu à grande échelle. Un exemple de texte reflétant la perspective idéologique que l'auteur souhaite diffuser est utilisé pour générer un nombre illimité d'articles du même genre, qui semblent tous être des originaux[62].

- La tromperie par l'image, par la création de *deepfakes*, qui sont des documents audio, vidéo ou des images modifiés et hyperréalistes[63]. Par exemple, une fausse vidéo montrant le président ukrainien Zelensky capitulant a circulé au début du conflit avec la Russie, mais a été rapidement démentie[64]. Le plus souvent, la tromperie par l'image est une tactique beaucoup plus simple qui consiste à utiliser des images hors contexte. Par exemple, d'anciennes photos sont présentées comme preuves d'un nouvel événement, comme ces photos publiées après une manifestation portant sur le réchauffement climatique à Londres pour montrer la présence de déchets, mais certaines étaient des photos de Mumbai[65]. Les mèmes sont de puissants outils de tromperie, parce que les images visuelles combinées à des énoncés courts et émotivement chargés sont particulièrement efficaces pour influencer l'opinion publique[66]. Les mèmes sont des outils de persuasion tellement efficaces que l'expression « guerre mémétique » a été inventée pour décrire le rôle clé que jouent les mèmes dans les stratégies d'influence[67].
- Le trucage manuel par lequel des êtres humains interviennent en ligne pour façonner les flux Internet. Il peut s'agir d'une usine ou ferme à *trolls*, qui sont payés ou organisés d'une autre manière[68]. L'objectif peut être le harcèlement, la modification des opinions politiques ou, plus généralement, la mise en place d'une méfiance à l'égard des institutions et de la démocratie. Certaines de ces usines à *trolls* sont devenues célèbres, comme celles de Russie utilisées pour perturber les élections américaines de 2016[69]. Cependant, des sociétés privées sont aussi régulièrement engagées par des entreprises pour lancer des campagnes d'influence pour leurs produits et services, ce qui transforme la désinformation à but lucratif en une industrie[70]. Les humains sont plus

---

[62] Voir Sarah Kreps, « The Role of Technology in Online Misinformation » (juin 2020) *Foreign Policy at Brookings*, en ligne : https://www.brookings.edu/wp-content/uploads/2020/06/The-role-of-technology-in-online-misinformation.pdf, p. 4.

[63] Bobby Chesney et Danielle Citron, « Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security » (2019) *California Law Review,* Vol. 107 : 1753.

[64] Tom Simonite, « A Zelensky Deepfake Was Quickly Defeated. The Next One Might Not Be » (17 mars 2022) *Wired,* en ligne : https://www.wired.com/story/zelensky-deepfake-facebook-twitter-playbook/.

[65] Lisa Fazio, « Out-of-context photos are a powerful low-tech form of misinformation » (4 février 2020) *The Conversation*, en ligne : https://theconversation.com/out-of-context-photos-are-a-powerful-low-tech-form-of-misinformation-129959.

[66] *Ibid.*

[67] Voir SCRS, *supra* note 29, p. 23.

[68] Bradshaw, *supra* note 27, fait référence aux cybertroupes : « acteurs du gouvernement ou des partis politiques chargés de manipuler l'opinion publique en ligne » [traduit par nos soins], p. 2.

[69] Rapport de la commission du Sénat des États-Unis sur le renseignement concernant les campagnes de mesures actives russes et l'ingérence de la Russie dans les élections américaines de 2016.

[70] Bradshaw, *supra* note 27, p. 9; Samuel Woolley parle de ce phénomène comme d'un élément clé de la propagande algoristique. Une question qui se pose est celle de savoir quelle est la limite entre cette pratique et le marketing. Woolley décrit cette pratique comme la « fabrication d'un consensus », et bien que moins

efficaces au début d'une campagne de désinformation pour cibler et affaiblir les sceptiques qui pourraient douter de la véracité de l'information[71].

- Des hybrides des techniques susmentionnées comme des campagnes de désinformation produites par des humains qui utilisent des outils automatisés. Une campagne fait souvent appel à de multiples techniques, telles que la création de faux comptes (usurpation d'identité ou piratage ou vol d'informations d'identification), l'utilisation de *bots*, de publications manuelles, de publicités payantes pour faire du microciblage, etc.[72]. Par exemple, ceux qui cherchent à perturber analysent d'abord les points de division sociale et politique, les groupes qui occupent des points de vue particuliers dans les débats et les types de contenu qui seraient les plus polarisants. Ils choisissent ensuite des outils pour générer et distribuer ce contenu polarisant, souvent en utilisant des outils d'intelligence artificielle (IA) qui peuvent fonctionner à grande échelle[73].
- La réalité virtuelle. Le mode d'expression artistique des courtes vidéos sur TikTok est devenu un vecteur efficace de propagation de la désinformation[74]. Le nouveau champ de bataille pour la manipulation de l'information est la réalité virtuelle. Le monde immersif du métavers a déjà été le théâtre de harcèlement, de mésinformation, de désinformation et de discours haineux[75].

Le défi que posent la mésinformation et la désinformation est qu'elles englobent souvent des activités sur plusieurs plateformes. Par exemple, le comportement en ligne après des tueries de masse suit un schéma d'ensemencement, d'amplification et de propagation. Les théories sont implantées dans des forums de moindre visibilité comme Reddit, 4chan et Discord, puis amplifiées sur des plateformes plus grand public comme Twitter et Facebook, avant d'être propagées dans l'écosystème plus vaste qui interagit avec ces plateformes[76]. Ce même schéma d'exploitation des influenceurs est évident dans toute campagne de désinformation; il se manifeste par exemple par la recherche d'un alignement sur des groupes aux idéologies similaires en vue du soutien à une cause et de son amplification[77], ou par la décontextualisation

---

dommageable, elle inclut par exemple les faux « J'aime » sur la publication d'un client. Plus insidieuses sont les pratiques consistant à payer des *trolls* pour harceler des journalistes ou à utiliser des *bots* pour renforcer certains contenus tels que les contenus antivaccin ou à payer des influenceurs pour diffuser des messages politiques : Woolley, *supra* note 61, p. 2.

[71] Wardle et Derakhshan, *supra* note 27, p. 35-36.

[72] Bradshaw, *supra* note 27, p. 2, 11; Wardle et Derakhshan, *supra* note 27, p. 43.

[73] Kreps, *supra* note 62, p. 4.

[74] On a récemment parlé de « guerre sur TikTok » en raison de la désinformation diffusée au sujet de la guerre entre la Russie et l'Ukraine : Alex Cadier et al., « La guerre sur TikTok : l'app expose ses utilisateurs à de la désinformation sur la guerre en quelques minutes – même s'ils ne cherchent pas de contenus liés à l'Ukraine » (mars 2022) *NewsGuard*, en ligne : https://www.newsguardtech.com/fr/misinformation-monitor/mars-2022/.

[75] Jillian Deutsh et al., « Misinformation Has Already Made Its Way to the Metaverse » (15 décembre 2021) *Bloomberg*, en ligne : https://www.bloomberg.com/news/articles/2021-12-15/misinformation-has-already-made-its-way-to-facebook-s-metaverse; voir cette étude : Adrian Verhulst et al., « Impact of Fake News in VR compared to Fake News on Social Media, a pilot study » (mai 2020) *IEEE Xplore*, en ligne: https://ieeexplore.ieee.org/document/9090558.

[76] SCRS, *supra* note 29, p. 18.

[77] Bradshaw, *supra* note 27, p. 9.

d'un message lors de sa diffusion sur différentes plateformes[78]. Largement répandus, les contenus extrémistes et autres actes préjudiciables en ligne sont couramment diffusés sur diverses plateformes, ce qui crée un environnement difficile à réglementer. Par exemple, l'auteur de la fusillade de Buffalo a exploré des contenus extrémistes sur 8chan, a rédigé un journal sur un serveur privé Discord sur lequel il a ensuite invité des utilisateurs, a publié un manifeste sur Google Docs, puis sur 8chan « moe » et 4chan, et a diffusé l'attaque en direct sur Twitch; la vidéo a ensuite été copiée et publiée ou liée à d'autres médias sociaux.

Le mouvement du Convoi était également diffusé sur de multiples plateformes et la campagne était menée par des personnes. Ce que l'on ne sait pas, c'est la mesure dans laquelle des outils automatisés ont été utilisés, des images ou des vidéos manipulées, et des publicités achetées et ciblées; on ne sait pas non plus dans quelle mesure les influenceurs clés ont bénéficié financièrement de l'amplification de leur contenu.

C pour contenu nuisible

Le contenu est le plus souvent la cible de la réglementation, en partie parce qu'il est plus difficile de réglementer les acteurs et les comportements trompeurs et parce que le contenu nuisible est ce qui est visible. Camille François soutient que pour être efficace, la réglementation doit se concentrer davantage sur AB et moins sur C[79]. Il s'agit d'un argument de poids en faveur de solutions techniques, juridiques et politiques qui devraient mieux cibler les acteurs, les comportements et la distribution. Cependant, on ne peut jamais échapper à une analyse du contenu parce qu'ultimement, une publication est générée et cela implique le droit à la liberté d'expression. En d'autres termes, la réglementation de toute activité sur les médias sociaux comporte toujours un élément de liberté d'expression. Les questions soulevées sont les suivantes :

- Les utilisateurs ont-ils droit à la liberté d'expression sur les médias sociaux? Cela englobe le droit de rechercher, de recevoir et de diffuser des informations et des idées.
- Les médias sociaux ont-ils un droit à la liberté d'expression implicite?
- Le contenu publié est-il potentiellement *illégal* sous une certaine forme, qu'il s'agisse de propagande haineuse, de propagande terroriste, de diffamation, d'atteinte à la vie privée, etc.? J'utilise cette formulation au sens large, car elle ne reflète pas les étapes d'une analyse juridique. Je la propose plutôt ici pour rappeler au lecteur qu'en fin de compte, il est possible qu'un contenu publié soit illégal, que ce soit sur le plan criminel ou civil. La question de savoir qui pourrait être responsable du contenu, en particulier si celui-ci est généré par un algorithme, peut être une tout autre histoire.

[78] Thakur et Hankerson, *supra* note 32, p. 9.
[79] François, *supra* note 33.

- Si la réglementation se concentre sur les acteurs, les comportements ou les méthodes de distribution, cela ne concerne-t-il pas indirectement l'expression, et quelle en est la légalité[80]?

Les lois et la gouvernance actuelles sont principalement axées sur la réglementation du contenu et sont examinées aux parties II et III.

D pour distribution

Alexandre Alaphilippe a ajouté un D au cadre ABC pour désigner la distribution. Comme il l'explique, la distribution de la désinformation dépend de la conception architecturale des plateformes. Les systèmes de recommandation et la publicité payante jouent un rôle important dans la diffusion et la monétisation de la mésinformation et de la désinformation[81]. Cela correspond au concept de propagande algoristique évoqué plus haut, à savoir que si l'on peut agir dans le système de distribution en utilisant l'automatisation et les algorithmes, cela constitue un vecteur d'attaque efficace pour manipuler l'opinion publique[82]. C'est également la raison pour laquelle certaines personnes affirment que la clé de la lutte contre la désinformation réside dans la réglementation des modèles d'affaires[83].

La conception des plateformes qui déterminent les publications qui sont recommandées et publicisées aux utilisateurs, et le contenu qui est ainsi amplifié, a une importance critique pour la propagation de la désinformation[84]. Nathalie Maréchal et ses coauteurs de la New America Foundation distinguent deux types d'algorithmes : ceux qui façonnent le contenu et ceux qui le modèrent[85]. Les algorithmes qui façonnent le contenu déterminent le contenu que les utilisateurs voient lorsqu'ils utilisent les services d'une entreprise. Ce peut être le fil d'actualité de Facebook, le système de recommandation de YouTube ou la page ForYou de TikTok. Cela comprend également la publicité microciblée. Certaines propositions de réforme législative

---

[80] Cela a été un point central de ma recherche récemment; voir aussi Daphne Keller, « Amplification and its Discontents » (8 juin 2021) *Knights First Amendments Institute at Columbia University*, en ligne : https://knightcolumbia.org/content/amplification-and-its-discontents.

[81] Alexandre Alaphilippe, « Adding a D to the ABC disinformation framework » (27 avril 2020) *TechStream,* Brookings Institute, en ligne : https://www.brookings.edu/techstream/adding-a-d-to-the-abc-disinformation-framework/.

[82] Voir Bradshaw, *supra* note 27; Woolley, *supra* note 61.

[83] Nathalie Maréchal et al., *Getting to the Source of Infodemics: It's the Business Model* (Mai 2020) New America Foundation, en ligne : https://www.newamerica.org/oti/reports/getting-to-the-source-of-infodemics-its-the-business-model/; Mais voir aussi l'entrevue avec Jonathan Stray réalisée par Evelyn Douek et Quinta Jurecic, « What We Talk About When We Talk About Algorithms » *The Lawfare Podcast*, en ligne : https://www.lawfareblog.com/lawfare-podcast-what-we-talk-about-when-we-talk-about-algorithms.

[84] Maréchal et al., *supra* note 83 : « Nous ne pouvons pas nettoyer les polluants en aval tels que la mésinformation ou les discours dangereux sans nous attaquer aux processus en amont – publicité ciblée et systèmes algorithmiques – qui rendent ces discours si dommageables pour notre environnement informationnel. » [traduit par nos soins], p. 10.

[85] *Ibid.*

portent sur l'obligation de neutralité des algorithmes de façonnement du contenu[86]. La neutralité est un faux-fuyant. La curation de contenu est essentielle pour gérer la réponse aux incidents et déclasser ou supprimer les messages nuisibles, tels que les publications haineuses ou les contenus portant sur les troubles alimentaires, et pour cibler la publicité selon les préférences des utilisateurs[87]. En revanche, la reddition de comptes algorithmique passant par la production de rapports transparents, l'accès des chercheurs aux données en vue d'un contrôle de la conformité, des audits obligatoirement effectués par une tierce partie, et la création d'un organisme de réglementation, sont autant de meilleures voies pour améliorer la conduite, la transparence et la confiance dans les systèmes[88].

L'autre type d'algorithmes sert à modérer le contenu. Comme l'explique Tarleton Gillespie, la modération du contenu n'est pas un élément accessoire au fonctionnement des plateformes, mais est plutôt une caractéristique qui les définit[89]. Les algorithmes de modération de contenu analysent le contenu pour déterminer si une publication enfreint les conditions générales du service en question et décident s'il faut y donner suite, avec ou sans l'intervention d'examinateurs humains[90]. La modération de contenu sera examinée à la partie III.

À propos du Convoi, on se pose la question de savoir comment les systèmes de recommandation et d'autres caractéristiques de conception des médias sociaux ont façonné ce que les utilisateurs voyaient, et on se demande dans quelle mesure la publicité était achetée pour des sujets liés au Convoi, par qui elle était achetée, et quels paramètres ont été utilisés pour le microciblage des utilisateurs. Bien que la modération du contenu soit examinée à la partie III, les questions abordées ici portent sur l'étendue de la modération automatisée et humaine, le nombre et les types de contenus traités, la rapidité avec laquelle ils l'ont été, le nombre de plaintes et les raisons des décisions prises.

La psychologie et les dangers de la manipulation de l'information

L'étude des effets de la mésinformation et de la désinformation, et donc de leurs dangers, est difficile. Sans consensus en ce qui concerne les définitions ou la terminologie, ou le problème précis à étudier, il est difficile de mesurer les effets. Les ensembles de données sont généralement ponctuels et de petite taille, ou bien les données sont dispersées en divers endroits, ou encore elles proviennent de diverses disciplines pour lesquelles il est difficile de

---

[86] Par exemple, la loi américaine *Protecting Americans from Dangerous Algorithms Act*, H.R. 2154, 117th Cong. (2021-2022) propose une exception à l'immunité en matière de responsabilité pour les médias sociaux (voir la discussion à la partie III, Aperçu des aspects juridiques) si des algorithmes sont utilisés pour la curation de contenu, à moins que cela ne soit fait d'une manière évidente, compréhensible et transparente.

[87] Voir Keller, *supra* note 80. La publicité et le marketing reposent sur le principe qu'il faut connaître les préférences des consommateurs et faire de la publicité en fonction de ces préférences : entrevue de Douek et Jurecic, *supra* note 83.

[88] Voir partie III, Réforme des lois.

[89] Tarleton Gillespie, *Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media* (Yale University Press, 2018), p. 21.

[90] Maréchal et al., *supra* note 83.

faire des rapprochements[91]. Il en résulte que « la désinformation est souvent liée à des objectifs larges, les effets peuvent être diffus et non ciblés, ce qui complique la recherche de preuves de préjudices »[92] [traduit par nos soins].

Dans l'étude des effets, il est important d'éviter de tomber dans le piège de la théorie de l'aiguille hypodermique, selon laquelle les utilisateurs sont des récepteurs passifs des messages injectés par les médias[93]. Cette théorie a été réfutée, mais elle est toujours d'actualité. Les recherches continuent, mais des études ont montré que les fausses informations peuvent avoir des conséquences sur la santé mentale, entraînant stress, fatigue, colère et panique[94]. L'exposition à la désinformation peut entraîner des échos de croyance, ce qui signifie qu'une personne sait que l'information est fausse, mais que ses attitudes sont néanmoins façonnées par celle-ci[95]. Et les utilisateurs dont les idées sont conservatrices sont plus susceptibles de suivre des comptes de désinformation que les utilisateurs libéraux[96]. Dans la recherche sur l'extrémisme, même si aucun lien de causalité ou lien direct ne peut être établi entre la radicalisation en ligne et la violence dans le monde réel, on constate néanmoins l'existence d'un lien, ce que les chercheurs appellent le façonnement de la décision et non la prise de décision[97].

Certaines études soulignent les effets perturbateurs de la désinformation, qui mine la confiance dans les institutions et la démocratie et sème la désillusion et le doute. Dans une étude commandée par le département thématique des droits des citoyens et des affaires constitutionnelles du Parlement européen, les chercheurs ont constaté que le degré de répercussions de la désinformation dépendait de la pluralité des médias et de l'organisation de la campagne de désinformation[98]. Une autre étude a mis en évidence la manière dont le doute peut être utilisé stratégiquement. Spencer McKay et Chris Tenove ont étudié les élections

---

[91] Eleni Kapantai et al., « A systematic literature review on disinformation: Toward a unified taxonomical framework », (2020) 23(5) *New Media & Society*; Duncan J. Watts et al., « Measuring the news and its impact on democracy », (2021) 118(15) PNAS, p. 2-5.

[92] Thakur et Hankerson, *supra* note 32 p. 8.

[93] Ahmed Al-Rawi et al., « What the Fake? Assessing the extent of networked political spamming and bots in the propagation of #fakenews on Twitter » (2019) 43(1) *Online Information Review*, p. 65.

[94] Yasmin Mendes Rocha et al., « The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review », (2021) 43(2) *Journal of Public Health: From Theory to Practice*.

[95] Emily Thorson, « Belief Echoes: The Persistent Effects of Corrected Misinformation », (2016) 33(3) *Political Communication*.

[96] Frederik Hjorth et Rebecca Adler-Nissen, « Ideological Asymmetry in the Reach of Pro-Russian Digital Disinformation to United States Audiences », (2019) 69(2) *Journal of Communication*, en ligne : https://doi.org/10.1093/joc/jqz006, p. 169-170.

[97] Ghayda Hassan et al., « Exposure to Extremist Online Content Could Lead to Violent Radicalization: A Systematic Review of Empirical Evidence », (juillet 2018) 12(7) *International Journal of Developmental Sciences 1,* en ligne : https://www.researchgate.net/publication/326384034_Exposure_to_Extremist_Online_Content_Could_Lead_to_Violent_Radicalization_A_Systematic_Review_of_Empirical_Evidence; voir aussi Craig Forcese et Kent Roach, « Criminalizing Terrorist Babble: Canada's Dubious New Terrorist Speech Crime », (2015) 53(1) *Alberta L Rev* 35.

[98] Judit Bayer et al., « Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States » (2019), Département thématique des droits des citoyens et des affaires constitutionnelles du Parlement européen.

américaines de 2016 et ont expliqué que des agents russes se sont employés à saper la crédibilité des institutions, en partie par la création de fausses institutions et de récits contradictoires[99]. Souvent, les dangers de la mésinformation et de la désinformation sont liés à des grandes valeurs des droits de la personne, et à l'atteinte portée à la dignité humaine, à l'autonomie, à la liberté d'expression et d'opinion, et à la vie privée[100].

Deux biais psychologiques ont été reconnus comme essentiels à la croyance et à la diffusion de fausses informations. Premièrement, le biais de confirmation consiste à rechercher de l'information qui renforce des croyances existantes, et à interpréter l'information de façon à ce qu'elle corresponde à ce que l'on croit. Deuxièmement, la cognition motivée et le traitement motivé de l'information consistent à avoir tendance à rechercher de l'information qui renforce la vision culturelle que l'on a. Il en résulte que les sources qui renforcent notre vision du monde préexistante sont interprétées comme étant les plus fiables[101].

Wardle et Derakhshan déterminent quatre caractéristiques qui font qu'un message trouve le plus d'écho auprès des utilisateurs : 1) il provoque une réponse émotionnelle, 2) il est visuel, 3) il contient un récit puissant, et 4) il est répété[102]. On retrouve ces quatre caractéristiques dans les médias sociaux. Ils permettent de faire partager le type de contenu émotionnel qui séduit tant les utilisateurs. Il est bien connu que la réaction à nos publications agit comme une décharge de dopamine, ce qui encourage le partage de publications qui vont dans le sens de la majorité et qui seront plus susceptibles d'être « aimées » et partagées[103].
En outre, l'obtention de nouvelles par l'intermédiaire des médias sociaux alimente l'acte rituel de la communication; nous lisons les nouvelles non pas pour obtenir de l'information nouvelle, mais parce que nous aimons le rituel, en quelque sorte « comparable à aller à l'église ». Nous ne sommes pas là pour recueillir de nouveaux renseignements, mais pour renforcer nos croyances[104]. Les échanges en ligne sont relationnels. L'une des raisons pour lesquelles on croit aux messages est le fait que les histoires publiées en ligne proviennent de sources multiples, de sorte que les lecteurs ne se concentrent pas sur la source pour évaluer la crédibilité, mais plutôt sur les histoires elles-mêmes, et la crédibilité est ensuite déterminée par les réseaux qui approuvent les histoires[105]. En outre, la répétition est facile en ligne et plus le message est répétitif, plus il est probable qu'il soit cru[106].

---

[99] Spencer McKay et Chris Tenove, « Disinformation as a Threat to Deliberative Democracy » (2020) 74(3) *Political Research Quarterly*, en ligne : https://doi.org/10.1177/1065912920938143.
[100] Bayer et al. vont dans cette direction, *supra* note 98, p. 76, et c'est évident dans l'analyse de la partie II sur la liberté d'expression.
[101] Rebecca K Helm et Hitoshi Nasu, « Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation » (février 2021) 21 *Human Rights Law Review* 302, p. 305-306.
[102] Wardle, *supra* note 27, p. 44.
[103] *Ibid* p. 14.
[104] *Ibid* p. 16.
[105] *Ibid* p. 14. Wardle discute également plus loin de la recherche qui montre que les gens sont plus susceptibles de croire qu'un message est vrai s'il provient d'une personne qu'ils connaissent, p. 57.
[106] Voir discussion *ibid* p. 51-55. Les auteurs discutent des six raccourcis mentaux utilisés pour évaluer la crédibilité d'un message : 1) réputation (familiarité); 2) approbation; 3) cohérence (répétition); 4) violation des attentes (si

Les visuels sont une forme efficace de manipulation de l'information. Les gens sont plus enclins à croire qu'une déclaration est vraie si elle est accompagnée d'une image, car celle-ci provoque une réaction émotionnelle[107], et elle peut modifier la mémoire de la nouvelle. Ainsi, par exemple, l'annonce d'un événement accompagnée d'une image choc aura un effet sur la mémoire de l'événement[108].
Les mèmes sont particulièrement efficaces, car en plus de l'image et du message court et facilement assimilable, ils sont souvent humoristiques. L'humour est la stratégie utilisée pour partager du contenu extrémiste en ligne et éviter le modérateur, car il « se fait passer pour une parodie propre au média »[109] [traduit par nos soins]. Par exemple, le guide du style du Daily Stormer recommandait l'utilisation de mots codés et d'humour pour diffuser les messages[110].
Les mèmes sont un moyen de créer une appartenance sociale et de se moquer des autres qui prennent les choses trop au sérieux. Blyth Crawford explique leur utilisation de la manière suivante :

> Ainsi, ces mèmes profitent de l'ambiguïté inhérente aux interactions en ligne, comme le souligne la loi de Poe, créant ce que Milner a appelé une « logique du lulz » selon laquelle il n'est jamais possible de discerner avec certitude le ton voulu d'une publication en ligne, ce qui rend tous les participants d'un espace en ligne perpétuellement vulnérables aux *trolls*. Ainsi, les opinions extrêmes peuvent se développer sous forme de mèmes, enrichis par une culture ambiante de sensibilité aux *trolls* et d'ambiguïté[111]. [traduit par nos soins]

L'ambiguïté est stratégique. Comme l'explique Alice Marwick, « l'ambiguïté est, en soi, une stratégie; elle permet aux participants de se dissocier d'éléments particulièrement peu attrayants tout en faisant la promotion du mouvement global »[112] [traduit par nos soins]. Ainsi, les messages de suprématie blanche sont transmis par le prisme de l'ironie et de la culture des mèmes, tant par les médias alternatifs que par les groupes en ligne. La culture des mèmes s'est infiltrée dans le domaine de la guerre, avec des discussions sur la « guerre mémétique », faisant allusion au « champ de bataille des médias sociaux » où se déroule une compétition pour « les

---

l'apparence et le comportement du site Web sont conformes aux attentes); 5) autoconfirmation (confirmation des convictions); et 6) intention de convaincre (intention du créateur du message), p. 51.

[107] Eryn J Newman et al., « Nonprobative photographs (or words) inflate truthiness » (août 2012) 19 *Psychonomic Bulletin & Review* 969, comme on en discute dans Fazio, *supra* note 61.

[108] Fazio, *supra* note 65. L'idée est que les visuels sont plus faciles à récupérer dans la mémoire, facilitent l'imagination d'un événement, servent de preuve de l'événement et attirent notre attention.

[109] Blyth Crawford, « The Influence of Memes on Far-Right Radicalisation » (9 juin 2020), Centre for Analysis of the Radical Right, en ligne : https://www.radicalrightanalysis.com/2020/06/09/the-influence-of-memes-on-far-right-radicalisation/.

[110] Andrew Marantz, « Inside the Daily Stormer's Style Guide » (15 janvier 2018) *New Yorker*, en ligne : https://www.newyorker.com/magazine/2018/01/15/inside-the-daily-stormers-style-guide.

[111] Crawford, *supra* note 109; voir aussi Ryan M. Milner, « FCJ-156 Hacking the Social: Internet Memes, Identity Antagonism, and the Logic of Lulz », *The Fiberculture Journal*, en ligne : http://twentytwo.fibreculturejournal.org/fcj-156-hacking-the-social-internet-memes-identity-antagonism-and-the-logic-of-lulz/. ; Selon la loi de Poe, il est impossible de savoir si quelque chose est une blague.

[112] Marwick et Lewis, *supra* note 35, p. 11.

récits, les idées et le contrôle social »[113] [traduit par nos soins]. Les mèmes font donc partie intégrante des opérations d'information visant à obtenir un avantage concurrentiel sur l'adversaire. Les mèmes et les *trolls* sont essentiellement les nouvelles formes de propagande dans la guerre[114].

Une question abondamment étudiée est l'existence des bulles de filtres et des chambres d'écho et leur effet sur les utilisateurs[115]. Il s'agit du concept selon lequel nos expériences en ligne sont maintenant personnalisées et nous enferment dans une chambre où nous entendons et lisons les mêmes personnes et les mêmes opinions, et interagissons avec elles. Notre fil d'actualité sur Facebook est personnalisé. Nous sélectionnons les salles à rejoindre sur l'application audio Clubhouse. Nous sélectionnons les personnes et les entités que nous suivons sur Twitter. Nous échangeons des messages avec les personnes et les groupes de notre choix sur Telegram. Nous vivons donc dans une bulle que nous avons créée et nos opinions ne sont jamais remises en question ni élargies. Cependant, les études sur les bulles de filtres sont mitigées. Plusieurs chercheurs soutiennent aujourd'hui que le phénomène a été sérieusement exagéré[116]. Axel Bruns, par exemple, affirme que le filtre ne vient pas du fait que nous ne voyons pas les contenus qui s'opposent à notre vision du monde, mais qu'il s'agit plutôt d'un filtre dans notre tête, qui nous amène à adopter une position d'opposition à l'information[117].

Bien que les effets de la manipulation de l'information dans le cadre du Convoi soient difficiles à évaluer, la nature du contenu qui influençait les partisans peut être analysée. Il y avait déjà un public pour de nombreux influenceurs du Convoi qui tenaient un discours antivaccination et anticonfinement[118]. On peut se demander dans quelle mesure de fausses informations étaient échangées et si certaines d'entre elles étaient intentionnellement communiquées (désinformation) à des publics qui les croyaient et les partageaient à nouveau (mésinformation). On peut également se demander dans quelle mesure il y avait du contenu dans la catégorie « tout le reste » englobant la haine, l'extrémisme, le *doxing*, le harcèlement,

---

[113] Jeff Giesea, « It's Time to Embrace Memetic Warfare », en ligne : https://stratcomcoe.org/pdfjs/?file=/publications/download/jeff_gisea.pdf?zoom=page-fit, p. 69.

[114] *Ibid :* « La guerre mémétique peut être utile au niveau du discours général, au niveau de la bataille, ou dans des circonstances particulières. Elle peut être offensive, défensive ou prédictive. Elle peut être déclenchée de manière indépendante ou en conjonction avec des forces cybernétiques, hybrides ou conventionnelles. » [traduit par nos soins], p. 69.

[115] Axel Bruns définit les chambres d'écho comme suit : « lorsqu'un groupe de participants choisit de *se connecter* de préférence les uns aux autres, en excluant les personnes de l'extérieur »; et il définit les bulles de filtres comme suit : « lorsqu'un groupe de participants choisit de *communiquer* de préférence les uns avec les autres, en excluant les personnes de l'extérieur » [traduit par nos soins] : « Filter Bubble » (29 novembre 2021) *Internet Policy Review*, en ligne : https://policyreview.info/concepts/filter-bubble.

[116] Voir, par exemple, Axel Bruns, *ibid*; Amy Ross Arguedas et al., *Echo chambers, filter bubbles, and polarisation: a literature review* (19 janvier 2022) Reuters Institute et Université d'Oxford, en ligne : https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2022-01/Echo_Chambers_Filter_Bubbles_and_Polarisation_A_Literature_Review.pdf.

[117] Bruns, *supra* note 115. La question qu'il pose est la suivante : « Pourquoi et comment différents groupes de la société en viennent à faire des lectures personnelles aussi divergentes d'une même information? » [traduit par nos soins].

[118] *Fifth Estate*, *supra* note 3.

etc.[119] Certains des comptes et des influenceurs du Convoi ont été signalés comme ayant des liens avec des groupes d'extrême droite[120].

Existe-t-il une solution à la manipulation de l'information en ligne? Les lois et la gouvernance sont complexes, et sont brièvement examinées aux parties II et III. Il semble y avoir un consensus sur le fait que la solution est multidimensionnelle et suppose de nombreuses stratégies qui, ensemble, permettent de contrecarrer et de gérer les risques de préjudice que comporte la manipulation de l'information[121]. Par exemple, l'éducation est une « inoculation »[122] essentielle qui aide le public à détecter les fausses informations, les sources douteuses, les faux comptes, etc.[123] La transparence est également importante. Les rapports des médias sociaux sur la publicité, la modération du contenu et la vie privée, par exemple, permettent aux utilisateurs d'évaluer la véracité de ce qu'ils consomment. L'éducation et la transparence ne fonctionnent que s'il existe des sources médiatiques fiables. Il est donc essentiel de soutenir des médias diversifiés et durables pour combattre les effets de la désinformation[124]. La technologie est également indispensable à la lutte contre la manipulation de l'information[125], qui consiste notamment à déterminer, signaler, déclasser ou limiter de toute autre manière la visibilité ou la viralité du contenu, même si elle n'est pas (et ne sera jamais) un instrument de réglementation parfait. Les solutions techniques sont examinées à la partie III.

Partie II Liberté d'expression et droits et responsabilités des utilisateurs

La suite de ce document explorera la question de la réglementation de la manipulation de l'information sous trois angles juridiques et de gouvernance. Le premier angle est celui du droit à la liberté d'expression. Comme nous l'avons vu plus haut, les politiques ne se concentrent plus uniquement sur la réglementation du contenu, mais aussi sur la réglementation des

---

[119] Le *doxing* est problématique d'un point de vue juridique, éthique et de cybersécurité, et il a été utilisé par les partisans et les opposants du Convoi. Il y a eu une fuite de données sur GiveSendGo et les donateurs ont été « *doxés* » : Tanya Basu, « Online activists are doxxing Ottawa's anti-vax protesters » (11 février 2022) *MIT Technology Review,* en ligne : https://www.technologyreview.com/2022/02/11/1045281/ottawa-antivax-protests-doxxing/; voir « Letter sent to parliamentarians warning of doxing ahead of trucker convoy: 'Go somewhere safe' » (28 janvier 2022) *City News Ottawa*, en ligne : https://ottawa.citynews.ca/local-news/letter-sent-to-parliamentarians-warning-of-doxing-ahead-of-trucker-convoy-go-somewhere-safe-5002917.

[120] Broderick, *supra* note 7.

[121] Rapport final du groupe d'experts de haut niveau sur les fausses nouvelles et la désinformation en ligne, *A multi-dimensional approach to disinformation* (2018) Commission européenne, en ligne : https://coinform.eu/wp-content/uploads/2019/02/EU-High-Level-Group-on-Disinformation-A-multi-dimensionalapproachtodisinformation.pdf, p. 4; Le groupe consultatif d'experts sur la sécurité en ligne, nommé par Patrimoine canadien, a discuté de l'importance d'adopter une approche multidimensionnelle pour aborder la manipulation de l'information : voir feuilles de travail en ligne : https://www.canada.ca/fr/patrimoine-canadien/campagnes/contenu-prejudiciable-en-ligne.html.

[122] Helm, *supra* note 101, p. 318.

[123] Kreps, *supra* note 62, p. 6; Commission européenne, *supra* note 121, p. 25-27.

[124] La Commission européenne reconnaît la nécessité pour l'État de protéger la liberté d'expression, la liberté de la presse et le pluralisme des médias pour relever les défis auxquels sont confrontés les médias : *ibid* p. 29.

[125] Voir Kreps, *supra* note 62, p. 6-7.

acteurs, des comportements et des méthodes de distribution. Ce champ d'application plus large est important, mais la liberté d'expression est une énigme juridique pertinente, quelle que soit la stratégie réglementaire adoptée. S'éloigner de la réglementation du contenu ne fait que rendre l'analyse indirecte plutôt que directe. Le deuxième angle concerne les responsabilités des utilisateurs dans la diffusion de la mésinformation, de la désinformation et de la malinformation. Le troisième angle est celui des obligations juridiques des médias sociaux et de la modération du contenu. Cette dernière fait l'objet de changements juridiques considérables, tant à l'échelle mondiale qu'au Canada.

Liberté d'expression

La liberté d'expression est au cœur de toute discussion sur la manipulation de l'information. Elle est examinée de près ici, car elle est liée à la réglementation du contenu. Deux problèmes émergent. Premièrement, la réglementation de la désinformation exige qu'un tribunal ou un organisme de décision qualifie un message de désinformation. Les tribunaux sont des organes chargés d'établir des faits, et en principe, qualifier un contenu de vrai ou de faux n'est pas un obstacle, et de nombreux domaines du droit qui touchent à la liberté d'expression peuvent comporter un constat de vérité. Par exemple, la vérité est un moyen de défense contre une plainte pour diffamation et un tribunal aurait pour tâche de déterminer si le défendeur s'est acquitté du fardeau d'établir la vérité selon la prépondérance des probabilités, en cas de plaidoirie.

Toutefois, la ligne de démarcation entre vérité et fausseté peut être plus compliquée dans la sphère de la manipulation de l'information. La classification des désordres de l'information de Wardle et Derakhshan témoigne de ce dilemme. Dans leur classification des types de mésinformation et de désinformation, ils énumèrent la satire ou la parodie, le contenu trompeur, le contenu falsifié, le contenu contrefait, les fausses connexions, le faux contexte et le contenu manipulé[126]. Albert Zhang et ses coauteurs donnent l'exemple d'un message publié par un porte-parole du ministère chinois des Affaires étrangères montrant un soldat australien qui brandit un couteau devant un enfant, soi-disant pour commenter l'enquête australienne sur les crimes de guerre en 2020. Le ministre australien des Affaires étrangères a dénoncé cette image comme étant de la désinformation. L'image était une œuvre d'art et elle n'a pas nécessairement été créée pour tromper[127].

Cela mène à un deuxième problème lié à la désignation d'un contenu comme étant de la désinformation. Toutes les définitions de la désinformation reposent sur l'intention de nuire (contrairement à la mésinformation). Prouver l'intention de nuire (et de nuire à quoi?) est un exercice difficile, car il faut pouvoir mettre le doigt sur les motivations des acteurs, ce qui est particulièrement difficile compte tenu de l'ambiguïté stratégique de nombreuses

[126] Wardle et Derakhshan, *supra* note 27, p. 19.
[127] Albert Zhang et al., *Submission to the UN Special Rapporteur on disinformation and freedom of opinion and expression*, en ligne : https://www.ohchr.org/sites/default/files/Documents/Issues/Expression/disinformation/2-Civil-society-organisations/Australian-Strategic-Policy-Institute.pdf, p. 3-4.

publications[128]. De plus, le cheminement qui mène à la création d'une campagne de désinformation peut être basé sur des informations vraies et des opinions honnêtes[129]. Une information qualifiée de fausse peut s'avérer ultérieurement contenir un grain de vérité ou une possibilité de vérité, comme la théorie de la fuite de la COVID-19 d'un laboratoire, qui a d'abord été rejetée par la communauté scientifique. Les messages à ce sujet ont été supprimés par de nombreuses plateformes de médias sociaux, avant que la théorie ne finisse par faire l'objet d'une enquête par les services secrets des États-Unis, à la demande du président Biden[130].

Le deuxième problème est plus fondamental pour la liberté d'expression elle-même; en effet, que signifie, en droit, le fait de dire que nous valorisons et protégeons la liberté d'expression et comment cela cadre-t-il avec la manipulation de l'information[131]? En vertu du droit international des droits de la personne, la liberté d'expression est protégée par l'article 19 du Pacte international relatif aux droits civils et politiques (PIDCP)[132], auquel le Canada est partie[133] :

> Article 19
> 1. Nul ne peut être inquiété pour ses opinions.
> 2. Toute personne a droit à la liberté d'expression; ce droit comprend la liberté de rechercher, de recevoir et de répandre des informations et des idées de toute espèce, sans considération de frontières, sous une forme orale, écrite, imprimée ou artistique, ou par tout autre moyen de son choix.
> 3. L'exercice des libertés prévues au paragraphe 2 du présent article comporte des devoirs spéciaux et des responsabilités spéciales. Il peut en conséquence être soumis à certaines restrictions qui doivent toutefois être expressément fixées par la loi et qui sont nécessaires :
> (a) au respect des droits ou de la réputation d'autrui;
> (b) à la sauvegarde de la sécurité nationale, de l'ordre public, de la santé ou de la moralité publiques.[134]

La liberté d'expression est protégée par la *Charte canadienne des droits et libertés* (la *Charte*)[135], en vertu de l'alinéa 2b) :

> 2 Chacun a les libertés fondamentales suivantes :

---

[128] Voir la discussion ci-dessus concernant les mèmes dans Marwick et Lewis, *supra* note 235, p. 11.
[129] Ryan Calo et al., « How do you solve a problem like misinformation? » (2021) 7(5) *Science Advances*, p. 1.
[130] Voir Stephan Lewandowsky, « The Lab-Leak Hypothesis Made It Harder for Scientists to Seek the Truth » (1er mars 2022) *Scientific American,* en ligne : https://www.scientificamerican.com/article/the-lab-leak-hypothesis-made-it-harder-for-scientists-to-seek-the-truth/.
[131] Voir Khan, *supra* note 30, partie B.
[132] *Supra* note 40. La Déclaration universelle des droits de l'homme (1948) est le point d'ancrage des droits internationaux de la personne et son article 19 énonce ceci : « Tout individu a droit à la liberté d'opinion et d'expression, ce qui implique le droit de ne pas être inquiété pour ses opinions et celui de chercher, de recevoir et de répandre, sans considérations de frontières, les informations et les idées par quelque moyen d'expression que ce soit. »
[133] Voir gouvernement du Canada, « Rapports sur les traités des Nations Unies relatifs aux droits de la personne », en ligne : https://www.canada.ca/fr/patrimoine-canadien/services/systeme-canada-nations-unies/rapports-traites-nations-unies.html.
[134] PIDCP, *supra* note 40, art. 19.
[135] Partie 1 de la *Loi constitutionnelle de 1982*, constituant l'annexe B de la *Loi de 1982 sur le Canada* (R.-U.), 1982, ch. 11, art. 8.

---

b. liberté de pensée, de croyance, d'opinion et d'expression, y compris la liberté de la presse et des autres moyens de communication[136].

La liberté d'expression est un droit fondamental dans une société démocratique et un élément central de notre quête de vérité, de démocratie, de découverte de soi et d'épanouissement[137]. C'est un droit complet, qui comprend le droit de rechercher, de recevoir et de répandre des informations et des idées sans considération de frontières[138], ainsi que le droit de ne pas s'exprimer[139]. La *Déclaration conjointe sur la liberté d'expression et les fausses nouvelles (« fake news »), la désinformation et la propagande* (la *Déclaration conjointe*)[140] confirme que le droit ne se limite pas à des déclarations correctes et qu'il comprend le droit de « choquer, offenser et déranger »[141]. Toute limitation du droit à la liberté d'expression doit répondre aux critères du droit international et de l'article 1 de la *Charte*, selon lesquels toute restriction doit être prévue par une règle de droit, servir un objectif légitime et être nécessaire et proportionnelle à cet intérêt.

Quelques éléments de l'article 19 sont particulièrement intéressants en ce qui concerne la manipulation de l'information. Premièrement, le droit à la liberté d'expression est le seul droit de la personne dans le PIDCP qui comporte des « responsabilités et des devoirs spéciaux »[142]. Alors que les débats se concentrent souvent sur les limites de ce droit, il est important de souligner que le PIDCP met l'accent sur les responsabilités et les droits particuliers qui constituent les fondements du droit à la liberté d'expression. Deuxièmement, le droit d'avoir des opinions est un droit absolu dans le PIDCP[143]. En pratique, nos idées et nos opinions sont

---

[136] *Ibid* alinéa 2b). Les limites justifiables de l'alinéa 2b) sont établies à l'article 1.

[137] La Cour suprême du Canada a déclaré : « La liberté d'expression vaut la peine d'être sauvegardée en raison de sa valeur intrinsèque » : *R. c. Keegstra*, [1990] 3 RCS 697, p. 881. La signification et la valeur de la liberté d'expression font l'objet d'un débat important, mais ces questions ne sont pas abordées dans le présent document.

[138] PIDCP, *supra* note 40, art. 19.

[139] Khan, *supra* note 30, par. 35.

[140] https://www.law-democracy.org/live/wp-content/uploads/2018/11/mandates.decl_.2017.French.pdf.

[141] *Ibid*. La Déclaration s'inspire de l'arrêt *Handyside c. UK*, [1976] ECHR 5 souvent cité (et cité dans *Irwin Toy Ltd. c. Québec (Procureur général)*, [1989] 1 RCS 927). Il convient de noter que dans *Handyside*, la Cour poursuit en déclarant : « telles sont les exigences de ce pluralisme, de cette tolérance et de cette largeur d'esprit sans lesquels il n'est pas de "société démocratique". Cela signifie, entre autres, que toute "formalité", "condition", "restriction" ou "sanction" imposée dans ce domaine doit être proportionnée à l'objectif légitime poursuivi. » [traduit en partie par nos soins]. Au Canada, la citation est souvent prise dans l'arrêt *Irwin Toy* : la liberté d'expression fait en sorte que « chacun puisse manifester ses pensées, ses opinions, ses croyances, en fait, toutes les expressions du cœur ou de l'esprit, aussi impopulaires, déplaisantes ou contestataires soient-elles », p. 968.

[142] PIDCP, *supra* note 40, art. 19. Francesca Klug a éclairé ce point dans un discours qu'elle a prononcé devant Intelligence Squared et lors du débat public du Centre culturel juif de Londres, Royal Geographical Society, « Freedom of Expression Must Include the Licence to Offend » (juin 2016), *Religion and Human Rights* 225.

[143] PIDCP, *supra* note 40, art. 19(1). Susie Alegre identifie trois éléments liés au droit d'avoir des opinions : le droit de ne pas révéler ses pensées ou ses opinions, le droit de ne pas les faire manipuler et le droit de ne pas être pénalisé pour ses idées : Susie Alegre, « Rethinking Freedom of Thought for the 21st Century » (2017) *European Human Rights Law Review* nº 3, p. 221-225; Evelyn Marie Aswad, « Losing the Freedom to be Human » (2020) 52(1) *Columbia Human Rights Law Review* 306; Khan, *supra* note 24.

influencées par toutes sortes de personnes et de médias. Le marketing et la publicité, par exemple, sont conçus pour influencer notre comportement de consommateur. La question qui se pose est celle de la limite entre les formes légitimes et les formes illégales de manipulation[144]. On peut soutenir que les campagnes de désinformation portent atteinte de manière injustifiée à l'autonomie d'une personne qui souhaite se forger une opinion sans être manipulée, et que la surveillance et le profilage des médias sociaux compromettent le droit de ne pas révéler ses pensées[145]. Troisièmement, les droits de chaque personne, y compris des personnes qui croient à la mésinformation et la diffusent, sont fragilisés par les campagnes de désinformation. Comme l'affirme la *Déclaration conjointe*, la désinformation porte atteinte à divers aspects du droit à la liberté d'expression, notamment le droit de savoir, de rechercher, de recevoir et de répandre des informations et des idées[146]. La désinformation peut nuire à la réputation et à la vie privée des individus, ainsi qu'à la sécurité nationale, ce qui peut donner lieu à une restriction légitime du droit à la liberté d'expression[147]. La désinformation peut également prôner la haine qui incite à la violence, à la discrimination ou à l'hostilité, ce qui est interdit en vertu de l'article 20 du PIDCP[148].

La mésinformation comme cible de la réglementation est particulièrement problématique. Les rumeurs et les potins font partie des rituels de l'interaction humaine[149]. Les personnes qui reçoivent et diffusent ces informations de manière innocente sont sans doute engagées dans la recherche de la vérité, l'une des valeurs philosophiques qui sous-tendent le droit à la liberté d'expression[150]. Il existe de nombreuses raisons de remettre en question la recherche de la vérité comme fondement suffisant sur lequel s'appuyer pour protéger la liberté d'expression dans ces circonstances. Elle repose sur l'idée qu'en laissant faire le marché des idées, la vérité fera surface[151]. Elle ne tient pas compte non plus du fardeau inégal que portent les groupes marginalisés et racialisés. Cette théorie suppose également que nous avons un accès égal à la liberté d'expression et que nous en faisons la même expérience. Or, des études montrent que les femmes, les personnes racialisées et les personnes LGBTQ+, en particulier les personnes qui se trouvent dans l'intersectionnalité, sont les principales cibles des abus et sont chassées de la participation en ligne[152]. Bref, le droit à la liberté d'expression est un droit dont souvent seuls les groupes privilégiés jouissent pleinement.

---

[144] Alegre, *supra* note 143, p. 227.
[145] Khan, *supra* note 30, par. 34-36; Alegre, *supra* note 143, p. 225.
[146] *Déclaration conjointe*, *supra* note 140.
[147] Voir PIDCP, *supra* note 40, art. 19(2).
[148] Voir *Déclaration conjointe*, *supra* note 140. Comme le dit Jones, *supra* note 37, l'article 20 montre que la désinformation n'est pas un phénomène nouveau et que les préoccupations relatives à son utilisation répandue au cours de la Seconde Guerre mondiale ont été prises en compte dans le PIDCP, p. 41.
[149] Robert Post, « The Social Foundations of Defamation Law: Reputation and the Constitution » (1986) 74 *Cali L Rev* 691.
[150] Les autres principales théories adoptées par la Cour suprême du Canada sont l'épanouissement personnel et la démocratie : voir, par exemple, *Irwin Toy Ltd. c. Québec (Procureur général)*, [1989] 1 RCS 927.
[151] Voir John Stuart Mills, *On Liberty* (1859) et la dissidence du juge Holmes dans l'affaire *Abrams c. US* (1919) 250 U.S. 616 : « la meilleure épreuve de la vérité est le pouvoir de la pensée de se faire accepter dans la concurrence du marché » [traduit par nos soins].
[152] Voir Jon Penney et Danielle Citron, « When Law Frees Us to Speak » (2019) 87 *Fordham Law Review*.

Cela ne veut pas dire que la liberté d'expression doit être affaiblie pour protéger les individus contre des offenses. Un système de liberté d'expression bien géré attend de nous que nous acceptions beaucoup de choses en fonction de l'idéal de la liberté d'expression. Et la libre circulation de l'information est un élément central de ce droit[153]. Cependant, l'analyse ne se limite pas à la tolérance des offenses. Les préjudices facilités par la désinformation et les médias sociaux ont des répercussions sur le droit à l'égalité, y compris l'égalité d'expression[154]. Par exemple, des chercheurs ont constaté l'utilisation de campagnes de désinformation ciblant les minorités raciales aux États-Unis afin d'empêcher la participation électorale des communautés de couleur[155]. Un autre exemple est l'utilisation de mèmes pour répandre des idéologies extrémistes en jouant sur l'humour et les clichés racistes familiers, ce qui déplace alors les limites du discours acceptable et permet de normaliser et d'accepter le racisme[156].

Il est difficile de concevoir une loi conforme aux droits de la personne qui cible les créateurs de désinformation. Tout droit doit être exercé à grande échelle et les exceptions doivent être interprétées rigoureusement[157]. Plusieurs lois adoptées dans d'autres pays illustrent les risques liés à l'adoption de lois de vaste portée interdisant la désinformation[158]. Elles risquent d'encourager les systèmes de filtrage ou de retrait du contenu, notamment par des coupures d'Internet, et de permettre le contrôle et le retrait par l'État des voix dissidentes (parfois dans des pays où l'État subventionne également sa propre désinformation). Même les États fortement engagés en faveur des droits de la personne ont été confrontés à des conséquences inattendues. Ce qui est préoccupant, ce sont les lois qui criminalisent la désinformation et qui n'ont pas de définitions suffisamment précises des fausses informations et/ou des préjudices. Des lois formulées en termes généraux ont été utilisées par des gouvernements contre la société civile, les journalistes et les opposants politiques[159]. Les lois civiles peuvent être

---

[153] Khan, *supra* note 30, par. 38.

[154] Voir *Keegstra*, *supra* note 137. Le juge en chef Dickson, représentant la majorité, a fait le raisonnement suivant :

> En fait, l'expression peut être utilisée au détriment de la recherche de la vérité. L'État ne devrait pas être le seul juge de ce qui constitue la vérité; par contre, il ne faut pas accorder une importance exagérée à l'opinion selon laquelle la raison prévaudra toujours contre le mensonge sur le marché non réglementé des idées. Il est en fait très peu probable que des déclarations destinées à fomenter la haine contre un groupe identifiable soient vraies, ou que la vision de la société qu'elles traduisent conduira à un monde meilleur. C'est donc un leurre de les présenter comme cruciales pour la détermination de la vérité et pour l'amélioration du milieu politique et social.

Voir aussi Cynthia Khoo, *Deplatforming Misogyny* (2021) *LEAF*, en ligne : https://www.leaf.ca/wp-content/uploads/2021/04/Full-Report-Deplatforming-Misogyny.pdf.

[155] Thakur et Hankerson, *supra* note 32, p. 10-11, et d'autres exemples.

[156] Crawford, *supra* note 109.

[157] Khan, *supra* note 30, par. 39.

[158] Voir Ruth Levush, *Government Responses to Disinformation on Social Media Platforms* (2019) Library of Congress, en ligne : https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1180&context=scholcom.

[159] Voir la discussion sur la *Déclaration conjointe*, *supra* note 140, p. 53-54. La *Déclaration conjointe* va jusqu'à dire que les lois sur la diffamation criminelle devraient être abolies au paragraphe 2(b), ce qui met le Canada en porte-à-faux avec les droits de la personne internationaux : *Code criminel*, *supra* note 40, art. 297-316; *R. c. Lucas,* [1998] 1 RCS 439.

légitimes mais elles doivent être étroitement adaptées, comme c'est le cas de notre législation sur la diffamation, dans laquelle le défendeur dispose d'un ensemble complet de moyens de défense visant à protéger la liberté d'expression, notamment la vérité, le commentaire loyal et la communication responsable dans l'intérêt public[160].

Lois canadiennes sur la désinformation

Il existe diverses lois qui s'appliquent aux individus qui communiquent de fausses déclarations. Mais il n'existe aucune loi qui cible directement les personnes qui communiquent de la mésinformation ou de la désinformation telles qu'elles sont présentées et examinées dans ce document.

En 1992, dans l'affaire R. c. Zundel[161], la Cour suprême du Canada (CSC) a invalidé la disposition du *Code criminel* qui interdisait la diffusion de fausses nouvelles. L'article 181 du *Code criminel* disposait que :

> 181. Est coupable d'un acte criminel et passible d'un emprisonnement maximal de deux ans quiconque, volontairement, publie une déclaration, une histoire ou une nouvelle qu'il sait fausse et qui cause, ou est de nature à causer, une atteinte ou du tort à quelque intérêt public[162].

Dans un jugement rendu par quatre voix contre trois, la majorité a jugé que l'article 181 portait atteinte au droit à la liberté d'expression garanti par l'alinéa 2b) de la *Charte* et n'était pas justifié en vertu de l'article 1. La majorité a souligné la sévérité des sanctions pénales et l'importance de la « liberté d'expression »[163]. Les juges ont considéré que le critère selon lequel la fausse déclaration cause ou est de nature à causer « une atteinte ou du tort à quelque intérêt public » était vague et trop général et représentait un grand danger pour les groupes minoritaires et leur pleine participation à la société. L'une des principales divergences entre la majorité et la minorité portait sur la manière de définir une fausse expression. Selon la majorité des juges, la disposition exigeait qu'un tribunal décide de la signification qui devait être jugée vraie ou fausse : « Diverses personnes peuvent attribuer à la même déclaration des sens différents à des moments différents »[164]. À leur avis, la vérité est une question de perception, et interdire la diffusion de fausses nouvelles permettrait à des groupes dominants d'imposer leur perception de la vérité à la minorité. Le jugement dissident a considéré les fausses informations différemment, en concluant qu'il existe des faits prouvables et que la disposition du *Code criminel* était étroitement axée sur la tromperie. Selon les juges, l'intention de tromper

---

[160] *Ibid*. Voir *WIC Radio c. Simpson,* [2008] CSC 40, et *Grant c. Torstar*, [2009] CSC 61. Pour des exemples propres à divers pays, voir Daniel Funke et Daniela Flamini, « A guide to anti-misinformation actions around the world », *Poynter*, en ligne : https://www.poynter.org/ifcn/anti-misinformation-actions/.
[161] *R. c. Zundel,* [1992] 2 RCS 731.
[162] *Code criminel*, *supra* note 40.
[163] *Zundel*, *supra* note 161.
[164] *Ibid*.

en diffusant des informations manifestement fausses et préjudiciables sape la valeur de la liberté d'expression[165].

Malgré l'affaire Zundel, la Cour suprême a confirmé la constitutionnalité des lois pénales et civiles en matière de diffamation[166], parce que les fausses informations entravent la recherche de la vérité et ne bénéficient pas du même niveau de protection que les discours politiques, bien qu'elle ait ensuite élargi la défense civile pour les questions d'intérêt public[167]. Ces affaires portent sur l'éventail des expressions de faible valeur qui peuvent être considérées comme éloignées des principes fondamentaux de la protection de l'expression[168]. Il existe également d'autres dispositions du *Code criminel* qui criminalisent un aspect de la fausseté, en particulier la propagande haineuse[169], l'incitation au terrorisme[170] et la fraude[171]. Plusieurs causes au civil comportent également un élément de fausseté, en particulier la diffamation et la représentation sous un faux jour[172]. Dans la mesure où la malinformation peut être visée par une disposition du *Code civil*, l'infliction intentionnelle de souffrance morale et la divulgation publique de faits privés embarrassants peuvent s'appliquer[173].

En 2018, le gouvernement du Canada a modifié le paragraphe 91(1) de la *Loi électorale du Canada*[174] pour supprimer le mot « sciemment » d'une disposition qui interdisait de faire ou de publier de fausses déclarations sur les caractéristiques personnelles ou la conduite d'un candidat en période électorale. La contestation constitutionnelle de la disposition portait sur la suppression du terme « sciemment » et sur la question de savoir si cela signifiait que l'intention n'était plus une condition de l'infraction[175]. La Cour supérieure de l'Ontario a conclu que la modification interdisait la diffusion de fausses informations accidentelles ou inconnues, comme

---

[165] *Ibid*.
[166] *Lucas*, *supra* note 159.
[167] *Ibid*; *Hill c. Église de scientologie de Toronto*, [1995] 2 RCS 130; *Grant*, *supra* note 160.
[168] La Cour suprême du Canada a toujours soutenu que toutes les formes d'expression ne sont pas traitées de manière égale, mais que la justification de l'atteinte au droit de parole repose sur un éventail d'expressions de faible à forte valeur qui contribuent à la recherche de la vérité, à la démocratie et à l'épanouissement personnel. Voir par exemple *Keegstra, supra* note 137; *Saskatchewan (Human Rights Commission) c. Whatcott*, 2013 CSC 11; *Lucas, supra* note 159 (« la valeur négligeable de l'expression diffamatoire »); *Hill, supra* note 167 (« les déclarations diffamatoires ont un lien très ténu avec les valeurs profondes qui sous-tendent l'alinéa 2b). Elles s'opposent à toute recherche de la vérité », par. 106). Mais voir ensuite l'arrêt *Grant, supra* note 160, dans lequel la Cour suprême a déclaré que l'arrêt *Hill* « doit être replacé dans le contexte de cette affaire » (par.57), et en adoptant une nouvelle défense de communication responsable dans l'intérêt public : « Le droit en matière de diffamation n'accorde actuellement aucune protection aux énoncés portant sur des questions d'intérêt public publiés sans destinataire précis s'il est impossible, pour une raison ou pour une autre, d'en prouver la véracité. Or, ce type d'énoncés favorisent les deux raisons d'être de la liberté d'expression dont il a été question précédemment – le débat démocratique et la recherche de la vérité – et il est donc nécessaire que le droit en matière de diffamation leur accorde une certaine protection. », par. 65.
[169] *Code criminel*, *supra* note 40, art. 319.
[170] *Ibid* art. 83.221.
[171] *Ibid* art. 320.
[172] *Yenovkian c. Gulian,* 2019 ONSC 7279.
[173] Voir *Jane Doe 72511 c. NM,* 2018 ONSC 6607.
[174] LC 2000, ch. 9, modification 2018, ch. 31, art. 61.
[175] *CCF c. Canada (AG)*, 2021 ONSC 1224.

dans la mésinformation, et que cela constituait une limite injustifiable au droit à la liberté d'expression. La disposition a été jugée sans effet[176].

Le droit de la concurrence s'applique à un aspect de la désinformation comme la publicité. La *Loi sur la concurrence*[177] interdit les représentations fausses ou susceptibles d'induire en erreur et les pratiques trompeuses pour promouvoir un produit, un service ou un intérêt commercial[178]. Cela comprend, par exemple, le fait de tromper les consommateurs pour obtenir leurs données[179]. Dans la mesure où la désinformation vise l'un de ces objectifs promotionnels, le Bureau de la concurrence peut faire enquête. Par exemple, des allégations fausses et trompeuses ont été faites sur des traitements de la COVID-19, et ont fait l'objet d'une enquête du Bureau de la concurrence[180]. Les influenceurs doivent également indiquer si leurs publications sont parrainées, que ce soit par des paiements, des rabais, des produits et services gratuits, ou d'autres moyens semblables[181].

Comme le montre le bref aperçu ci-dessus, il n'existe actuellement aucune loi au Canada qui vise directement la mésinformation et la désinformation. Dans le droit criminel post-Zundel, la poursuite serait motivée par un autre tort d'ordre supérieur. Par exemple, une fausse déclaration qui constitue un acte de haine, de terrorisme ou de fraude. En droit civil, une dynamique semblable est observable. La conduite peut donner lieu à des poursuites si une fausse déclaration porte atteinte à la réputation (diffamation) ou présente une personne sous un faux jour (en Ontario). Si la malinformation est comprise comme étant essentiellement du *doxing*, alors la divulgation publique, en common law, de faits privés embarrassants peut être applicable dans certaines provinces[182]. Le fait qu'il n'y ait pas de loi qui s'applique directement à la mésinformation ou à la désinformation peut être approprié. En reliant la conduite à une autre faute d'ordre supérieur, on peut engager des poursuites ou une action au civil dans les circonstances les plus flagrantes. Cependant, on peut se demander si le résultat serait différent si l'affaire Zundel était jugée aujourd'hui.

---

[176] Voir Eve Gaumond, « Why a Canadian Law Prohibiting False Statements in the Run-Up to an Election Was Found Unconstitutional » (16 mars 2021), en ligne : https://www.lawfareblog.com/why-canadian-law-prohibiting-false-statements-run-election-was-found-unconstitutional.

[177] L.R.C 1985, ch. C-34; Voir aussi la *Loi sur la modernisation des élections,* L.C. 2018, ch. 31, qui a nécessité la création de registres de publicité politique.

[178] *Loi sur la concurrence*, *supra* note 177, art. 52 et partie 74.01. Il existe un régime criminel et un régime civil. Voir explication : Gouvernement du Canada, « Indications fausses ou trompeuses et pratiques commerciales trompeuses », en ligne : https://www.bureaudelaconcurrence.gc.ca/eic/site/cb-bc.nsf/fra/03133.html.

[179] *Ibid*.

[180] Bureau de la concurrence Canada, « Le Bureau de la concurrence lutte contre les indications commerciales trompeuses au sujet de la prévention et du traitement de la COVID-19 » (6 mai 2020), en ligne : https://www.canada.ca/fr/bureau-concurrence/nouvelles/2020/05/le-bureau-de-la-concurrence-lutte-contre-les-indications-commerciales-trompeuses-au-sujet-de-la-prevention-et-du-traitement-de-la-covid-19.html.

[181] Gouvernement du Canada, « Le marketing d'influence et la *Loi sur la concurrence* », en ligne : https://www.bureaudelaconcurrence.gc.ca/eic/site/cb-bc.nsf/fra/04512.html.

[182] *Jones c. Tsige*, 2021 ONCA 312. Il peut également s'agir d'une atteinte à la vie privée en vertu de la loi dans certaines provinces.

Partie III Le droit et la gouvernance des médias sociaux

La question qui s'ensuit est la suivante : quels mécanismes de droit et de gouvernance existe-t-il pour tenir les médias sociaux et autres services en ligne responsables des contenus préjudiciables publiés par leur intermédiaire? Il s'agit d'un domaine d'étude très riche, et une analyse détaillée dépasse la portée du présent document, mais le lecteur est invité à consulter les sources citées pour obtenir de plus amples renseignements[183].

En fin de compte, la réglementation dépend de « la mesure dans laquelle le gouvernement délègue réellement une participation aux acteurs privés »[184] [traduit par nos soins]. Trois types de réglementation entrent en jeu pour les services de médias sociaux. Premièrement, il y a les lois qui s'appliquent aux médias sociaux sur le plan de la réglementation du contenu. Ce domaine du droit est connu sous le nom de responsabilité des intermédiaires, en raison du rôle d'intermédiaire de ces sociétés, qui relient les créateurs de contenu aux consommateurs de contenu. Leur rôle, traditionnellement, peut être compris comme un rôle facilitateur et secondaire par rapport à celui des créateurs de contenu, et donc moins moralement responsable[185]. Souvent, le terme « plateforme » est maintenant utilisé pour désigner les intermédiaires qui ont un pouvoir social ou culturel particulier sur le marché[186]. Un autre domaine du droit qui a une incidence sur la sécurité en ligne est le droit de la protection de la vie privée dans le secteur privé. Les données des utilisateurs sont au cœur de la fonctionnalité et de la rentabilité des médias sociaux, et les entreprises concernées ont des obligations envers les utilisateurs en matière de protection de leurs renseignements personnels. Ce domaine du droit n'est pas utile pour traiter de la légalité des publications individuelles, mais plutôt pour déterminer si la conception des médias sociaux et des transactions de données particulières protège suffisamment la vie privée des utilisateurs.

Le deuxième type de réglementation est la coréglementation qui est une forme d'autoréglementation soutenue par le gouvernement, comme les codes de pratique et les

---

[183] Pour un aperçu général de la responsabilité des intermédiaires canadiens, voir Emily B. Laidlaw, « Mapping Current and Emerging Models of Intermediary Liability » (2019), rédigé pour le Groupe d'examen du cadre législatif en matière de radiodiffusion et de télécommunications, en ligne : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3574727; Pour la modération du contenu, voir Evelyn Douek, « Content Moderation as Administration » *À venir* 136 *Harvard Law Review*, ébauche en ligne : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4005326.

[184] Chris T. Marsden, *Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace* (Cambridge University Press, 2011) p. 53.

[185] Selon la définition de l'OCDE, les intermédiaires « rassemblent ou facilitent les transactions entre tiers sur Internet » [traduit par nos soins] : OCDE, *Economic and Social Role of Internet Intermediaries*, (avril 2010), en ligne : https://www.oecd.org/internet/ieconomy/44949023.pdf, p. 9.

[186] L'exploration des différents types de plateformes comporte plusieurs autres facettes qui dépassent le cadre du présent document. Voir, par exemple, Tarleton Gillespie, « Platforms are not Intermediaries », (2018) 2 *Geo L Tech Rev.* 198 209; José van Dijck, Thomas Poell et Martijn De Waal, *The Platform Society* (Oxford University Press, 2019); Robert Gorwa, « What is Platform Governance?» (2019) 22(6) *Information, Communication & Society* 854.

organismes industriels[187]. La coréglementation est fondée sur la collaboration et aide à combler l'écart entre l'obligation juridique et la démarche volontaire; elle est au cœur de la gouvernance d'Internet depuis la commercialisation d'Internet. Ce type de réglementation n'est que brièvement abordé ici. Le troisième type de réglementation est l'autoréglementation, ou dans le cas présent, la modération du contenu par les médias sociaux[188]. L'absence de lois fédérales sur la responsabilité des intermédiaires au Canada signifie que la modération du contenu a été la principale force réglementaire dans le cas du Convoi.

Aperçu des aspects juridiques

Le Canada n'a pas de loi fédérale exhaustive sur la responsabilité des intermédiaires, contrairement à l'Europe[189] et aux États-Unis d'Amérique[190]. Aux États-Unis, l'article 230 du Communications Decency Act (CDA)[191] offre une vaste immunité en matière de responsabilité des intermédiaires pour les contenus publiés par des tiers, sauf en ce qui concerne le droit pénal fédéral, le droit de la propriété intellectuelle ou la confidentialité des communications électroniques, et une modification récente, vivement critiquée, visant à lutter contre la traite des êtres humains[192]. Il en résulte que les intermédiaires disposent d'une sphère de sécurité pour éviter toute responsabilité en cas de mésinformation ou de désinformation que les utilisateurs publieraient et qui pourrait être illégale, parce que, par exemple, elle serait diffamatoire ou révélerait des informations privées. Il est essentiel de faire ressortir quelques éléments de l'article 230. Tout d'abord, l'immunité concerne les décisions de laisser le contenu en place ou de le retirer. Ainsi, les médias sociaux sont protégés pour élaborer des pratiques de modération de contenu plus strictes que la loi et conformes à leurs valeurs. Le problème de l'article 230 est qu'il n'incite pas à la responsabilité et que les entreprises peuvent profiter, et ont profité, de la protection de l'article 230 pour laisser du contenu illégal en ligne sans prendre les mesures nécessaires pour mettre en œuvre des pratiques de modération[193].

---

[187] La réglementation ne se limite pas à ces catégories, même si celles-ci sont les trois principales. Voir Emily B. Laidlaw, « The Challenge Designing Intermediary Liability Laws » dans Catherine Easton et David Mangan, *The Philosophical Foundations of Information Technology Law* (Oxford University Press, parution en 2023).

[188] Comme le dit Chris Marsden, « l'autoréglementation n'existe généralement pas sous sa forme pure, car le gouvernement est souvent dans l'ombre et fait pression sur les entreprises pour qu'elles agissent » [traduit par nos soins] : *supra* note 184, p. 48.

[189] *Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (« directive sur le commerce électronique »)* et *Règlement du Parlement européen et du Conseil relatif à un marché intérieur des services numériques (Législation sur les services numériques) et modifiant la directive 2000/31/CE*, COM(2020) 825 (Directive sur le commerce électronique).

[190] *Communications Decency Act,* 47 USC 230 (CDA). Il existe, bien entendu, d'autres États et territoires à prendre en considération, mais je mentionne les États-Unis et l'Union européenne, car ils ont des systèmes juridiques similaires et ont été deux des premières entités à mettre en œuvre des lois de grande portée sur la responsabilité des intermédiaires.

[191] *Ibid*.

[192] *Ibid*.

[193] En outre, la large protection juridique accordée à la liberté d'expression en vertu du premier amendement signifie qu'un contenu qui serait considéré comme de la propagande haineuse au Canada est un discours protégé

L'UE, en revanche, a adopté un modèle de sphère de sécurité conditionnelle avec la *Directive sur le commerce électronique* (DCE)[194]. Selon ce modèle, un intermédiaire qui héberge le contenu d'un tiers bénéficie d'une exonération conditionnelle de responsabilité pour le contenu illicite publié par les utilisateurs. Toutefois, l'intermédiaire risque de perdre son immunité s'il prend connaissance du contenu illicite et n'agit pas pour en interdire l'accès. Ce modèle fonctionne donc comme un régime de notification et de retrait qui s'articule autour de la *connaissance* des activités illégales et de l'obligation *d'agir* en ce qui concerne le contenu illégal en le retirant.

Par souci de clarté, il convient de préciser que les obligations des intermédiaires conformément à la DCE ne seraient déclenchées que pour des contenus illégaux et qu'une part importante de la manipulation de l'information est légale même si elle est gênante. Ainsi, la Commission européenne a dirigé la rédaction d'un *Code de bonnes pratiques contre la désinformation*[195], une initiative volontaire de l'industrie. Le retrait de contenu illicite est important[196], mais les modèles de sphère de sécurité conditionnelle sont plus difficiles à mettre en œuvre qu'il n'y paraît à première vue. Ils tendent à encourager le retrait de contenus sans protection correspondante des contenus légitimes ou des décisions prises par des entreprises plus sensibles aux droits de la personne[197].

La tendance actuelle en matière de réforme législative est de passer à un modèle de diligence raisonnable pour les intermédiaires, y compris en Europe, qui a complété la DCE avec la *Législation sur les services numériques*[198]. Ces mesures législatives sont désignées de diverses manières : obligation de diligence, évaluation des risques et diligence raisonnable. Fondamentalement, ces modèles s'écartent d'un modèle binaire « laisser/supprimer » pour confier aux intermédiaires la tâche de gérer les risques de préjudice de leurs services. Ils ont l'avantage de ne pas se limiter à la réglementation du contenu et peuvent être davantage axés sur l'acteur, le comportement et la distribution. Il existe des variations importantes dans la manière dont ces modèles pourraient être mis en œuvre, avec les risques associés à une

---

aux États-Unis et ne ferait donc pas l'objet d'un examen constitutionnel ou d'un examen au titre de l'article 230. L'article 230 a eu des répercussions à l'échelle mondiale, avec des controverses lorsqu'il entre en conflit avec les lois d'autres pays. Voir par exemple : *Google Inc c. Equustek Solutions Inc*: [2017] SCC 34; *Google Inc c. Equustek Solutions Inc*, [2017] WL 5000834 (ND Cal Nov 2, 2017). Voir ensuite *Equustek Solutions Inc c. Jack*, [2018] BCSC 610; Michael Geist, « The *Equustek* Effect: A Canadian Perspective on Global Takedown Orders in the Age of the Internet » dans Giancarlo Frosio, ed., *The Oxford Handbook of Online Intermediary Liability* (Oxford University Press, 2020).

[194] La mise en œuvre de la directive a varié considérablement d'un État membre à l'autre, et c'est l'une des raisons pour lesquelles la *Législation sur les services numériques* a été élaborée : *supra* note 189. Voir aussi le *Digital Millennium Copyright Act*, Pub. L. No. 105-304, 112 Stat. 2860 (1998).

[195] Voir le *Code de bonnes pratiques contre la désinformation de 2018*, en ligne : https://digital-strategy.ec.europa.eu/fr/library/2018-code-practice-disinformation et le *Code de bonnes pratiques renforcé contre la désinformation* (juin 2022), en ligne : https://ec.europa.eu/commission/presscorner/detail/fr/ip_22_3664.

[196] Mais voir la discussion sur l'efficacité à la partie III, Modération du contenu.

[197] David Kaye, *Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression,* A/HRC/32/38 (2016), par. 43-44.

[198] *Législation sur les services numériques* (DSA, Digital Services Act), *supra* note 189.

mauvaise exécution. Je suis plutôt en faveur d'une approche de gestion des risques en matière de responsabilité des intermédiaires, même si le diable se cache dans les détails. Les modèles de diligence raisonnable seront examinés plus loin dans la section sur la réforme des lois.

Comme on l'a expliqué, le Canada n'a pas de loi fédérale sur la responsabilité des intermédiaires semblable à celle des États-Unis ou de l'Europe. Le *Code criminel* prévoit un mécanisme pour le retrait de contenu. Un tribunal peut ordonner le retrait de contenus en ligne de propagande terroriste ou haineuse, de pornographie juvénile, de voyeurisme et de divulgation non consensuelle d'images intimes[199]. Le Québec est la seule province ayant une loi générale sur la responsabilité des intermédiaires, qui prévoit une sphère de sécurité à condition que, si un intermédiaire prend connaissance d'une activité illicite dans ses services, il agisse rapidement pour bloquer l'accès au contenu[200].

Les lois canadiennes sur la responsabilité des intermédiaires en matière de manipulation de l'information se sont développées principalement dans le cadre de la législation sur la diffamation[201]. En pratique, ces lois fonctionnent de manière semblable à la DCE comme un système d'avis et de retrait. Si TikTok, par exemple, apprend qu'elle héberge une vidéo au contenu diffamatoire, elle est obligée de désactiver l'accès à la vidéo ou de risquer d'être tenue responsable du tort sous-jacent[202]. Le contenu diffamatoire ne représente qu'une fraction des formes de manipulation de l'information en cause, notamment les fausses informations qui portent atteinte à la réputation d'une personne ou d'une entité[203]. La mésinformation et la désinformation portent souvent sur des sujets généraux, comme la santé. Par ailleurs, le gouvernement du Canada est limité en matière de lois sur la responsabilité des intermédiaires

---

[199] *Code criminel*, *supra* note 40, art. 320.1(5), 83.223, 164.1(5).

[200] *Loi concernant le cadre juridique des technologies de l'information*, RLRQ, ch. C-1.1.
Deux différences entre la *Directive sur le commerce électronique* (DCE) et la loi québécoise sont notables. Premièrement, la loi québécoise parle d'activité illicite, un concept plus large que celui de contenu illégal. Pierre Trudel explique que même si cela englobe un contenu légal, les contraintes constitutionnelles font qu'il serait interprété de manière restrictive. Deuxièmement, l'article 22 prévoit que l'intermédiaire « peut engager sa responsabilité », ce qui signifie que le critère d'analyse est de savoir si l'intermédiaire a agi avec diligence dans les circonstances : Pierre Trudel, « La responsabilité des plateformes : la loi du Québec » (dans un fichier de l'auteur), p. 2-3.

[201] Pour les lois sur la responsabilité des intermédiaires en matière de droit d'auteur, voir la *Loi sur le droit d'auteur*, LRC (1985), ch. C-42, art. 41.25-41.27.

[202] Voir, par exemple, *Weaver c. Corcoran*, [2015] BCSC 165; En ce qui concerne la publication, voir *Crookes c. Newton*, [2011] CSC 47. De plus, en matière de diffamation, il suffit que l'intention soit de diffuser l'information, ce qui fait que des personnes peuvent faire l'objet d'une diffamation involontaire : *Hulton c. Jones*, [1910] AC 20. Par conséquent, la mésinformation peut faire l'objet d'une poursuite en diffamation (diffusion intentionnelle d'une fausse information que l'on croit être vraie), bien qu'il faille noter les moyens de défense des commentaires loyaux et la communication responsable dans l'intérêt public : *WIC Radio, supra* note 160; *Grant*, *supra* note 160.

[203] *Hill*, *supra* note 167. Diverses mesures de défense sont importantes pour protéger de manière générale la liberté d'expression qui pourrait néanmoins nuire à la réputation, mais cette question n'est pas examinée ici. Il convient de noter que la question de savoir si un intermédiaire pourrait être tenu responsable d'un délit d'atteinte à la vie privée n'a pas été testée en droit, bien que l'on puisse supposer qu'un tribunal s'inspirerait des principes de la diffamation.

qu'il peut mettre en place en raison de ses engagements commerciaux en vertu de l'Accord Canada-États-Unis-Mexique (ACEUM)[204].

Les lois sur la protection de la vie privée dans le secteur privé (aux niveaux fédéral et provincial)[205] sont à la base de la protection de la vie privée, et la culpabilité morale des entreprises est plus directe que le domaine de la responsabilité des intermédiaires. Les lois en matière de protection de la vie privée exigent que les organisations soient responsables des renseignements personnels concernant une personne identifiable qu'elles recueillent, utilisent ou communiquent dans le cadre d'activités commerciales[206]. Ce qui est complexe, étant donné l'opacité de nombreuses pratiques commerciales des médias sociaux, c'est d'identifier les flux d'informations pour déterminer précisément ce que les médias sociaux recueillent, utilisent et communiquent, ainsi que les différentes tierces parties avec lesquelles ils effectuent des transactions. La meilleure illustration de cette complexité est l'*Enquête conjointe du Commissariat à la protection de la vie privée du Canada et du Bureau du Commissaire à l'information et à la protection de la vie privée de la Colombie-Britannique au sujet de Facebook, Inc.* (« l'enquête conjointe »)[207] concernant le scandale Cambridge Analytica. La société Cambridge Analytica a utilisé les données d'une application appelée *This is Your Digital Life*, qui recueillait des informations sur les utilisateurs de Facebook pour établir leur profil psychologique, qui était ensuite utilisé pour envoyer des publicités ciblées afin d'influencer les électeurs lors de diverses élections, notamment pendant la course à l'investiture républicaine pour la présidentielle américaine et l'élection qui a suivi. L'enquête conjointe a conclu que

---

[204] Voir l'article 19.17 de l'Accord Canada-États-Unis-Mexique (ACEUM), en ligne : https://www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/cusma-aceum/text-texte/toc-tdm.aspx?lang=fra.
L'article 19.17 de l'ACEUM introduit une large exonération de responsabilité pour les intermédiaires, inspirée de l'article 230 du CDA, bien que la question de savoir si elle va aussi loin que l'article 230 fasse l'objet d'un débat. L'article 19.17 interdit de traiter un intermédiaire « comme un fournisseur de contenu informatif pour déterminer la responsabilité » [traduit par nos soins], ce qui laisse une marge de manœuvre pour les recours équitables et les modèles d'obligation de diligence/gestion des risques : voir Vivek Krishnamurthy et al., « CDA 230 Goes North American? Examining the Impacts of the USMCA's Intermediary Liability Provisions in Canada and the United States » (7 juillet 2020) CIPPIC, en ligne : https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3645462. Le gouvernement du Canada, dans « l'Énoncé des mesures de mise en œuvre du Canada » , en ligne : https://www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/cusma-aceum/implementation-mise_en_oeuvre.aspx?lang=fra indique que l'article19.17 signifie que les intermédiaires ne sont pas tenus civilement responsables des publications des utilisateurs ni des mesures prises pour les modérer, ce qui est conforme à l'article 230 du CDA. L'ACEUM est entré en vigueur le 1er juillet 2020.

[205] Notre loi fédérale sur la protection de la vie privée dans le secteur privé est la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE), L.C. 2000, ch. 5. Plusieurs lois provinciales ont été jugées essentiellement similaires, comme le *Personal Information Protection Act* de l'Alberta, SA 2003, ch. P-6.5.

[206] LPRPDE, *supra* note 205, art. 4(1). La personne doit être identifiable, mais le Commissariat à la protection de la vie privée interprète cela de manière large.

[207] Commissariat à la protection de la vie privée du Canada, *Enquête conjointe du Commissariat à la protection de la vie privée du Canada et du Bureau du Commissaire à l'information et à la protection de la vie privée de la Colombie-Britannique au sujet de Facebook Inc.* (25 avril 2019), en ligne : https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/enquetes/enquetes-visant-les-entreprises/2019/lprpde-2019-002/.

Facebook (aujourd'hui Meta) a violé les lois en matière de protection de la vie privée, car elle n'a pas assumé la responsabilité de protéger efficacement la vie privée, n'a pas obtenu un consentement valable des utilisateurs et n'a pas mis en place des mesures de sécurité adéquates[208].

Au moment de la rédaction du présent document, le gouvernement du Canada a présenté le projet de loi C-27 visant à réformer la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE) et à étendre la protection de la vie privée[209]. L'analyse de ce projet de loi dépasse la portée du présent document, mais le lecteur est invité à l'examiner sous l'angle de la responsabilité des médias sociaux et de la protection contre les préjudices en ligne. Les données constituent la base de la manipulation de l'information. Ainsi, la législation sur la protection de la vie privée est un cadre naturel pour traiter de la protection des utilisateurs contre les systèmes généraux de surveillance[210], pour assurer la surveillance des modèles commerciaux qui rendent les utilisateurs vulnérables[211], pour introduire des mesures de responsabilité algorithmique et une réglementation de l'intelligence artificielle, pour compléter les lois relatives aux pratiques commerciales trompeuses par des obligations précises en matière de protection des données, et pour renforcer les droits et les obligations concernant les types de pratiques en matière de données qui sont acceptables dans un environnement d'information dont nous ne pouvons pas prétendre de façon réaliste que nous pouvons nous retirer.

Où cela nous mène-t-il? Les médias sociaux sont certainement réglementés au Canada, mais nos lois comportent d'importantes lacunes en ce qui concerne la responsabilité des intermédiaires pour les préjudices en ligne, et plus particulièrement pour la manipulation de l'information. En général, la voie la plus viable pour la responsabilité des intermédiaires est une réclamation en vertu de la législation en matière de diffamation, mais seulement certaines formes de manipulation de l'information sont diffamatoires. En dehors de la responsabilité des intermédiaires, la protection des données est un outil juridique important pour traiter les aspects de la responsabilité des plateformes liés à la protection de la vie privée, mais ce n'est qu'une partie de la solution et elle ne règle pas directement le problème de la manipulation de l'information ou des préjudices en ligne de manière plus générale.

---

[208] Le Bureau de la concurrence a également enquêté sur Facebook à propos de cette affaire pour représentations fausses ou trompeuses. Facebook a réglé pour 9,2 millions de dollars : Gouvernement du Canada, « Facebook payera une sanction de 9 millions de dollars pour régler les préoccupations du Bureau de la concurrence à propos d'indications trompeuses quant à la confidentialité » (18 mai 2020), en ligne : https://www.canada.ca/fr/bureau-concurrence/nouvelles/2020/05/facebook-payera-une-sanction-de-9millions-de-dollars-pour-regler-les-preoccupations-du-bureau-de-la-concurrence-a-propos-dindications-trompeuses-qu.html.

[209] Projet de loi C-27 : *Loi édictant la Loi sur la protection de la vie privée des consommateurs, la Loi sur le Tribunal de la protection des renseignements personnels et des données et la Loi sur l'intelligence artificielle et les données et apportant des modifications corrélatives et connexes à d'autres lois*, 1re session, 44e législature, 2022.

[210] Les lois sur les intermédiaires, telles que la *Directive sur le commerce électronique*, disposent qu'il n'y a pas d'obligation générale de surveillance : voir *Directive sur le commerce électronique*, *supra* note 189, art. 15.

[211] C'est là que la législation sur la concurrence pourrait jouer un rôle clé, spécialement en ce qui concerne le microciblage et son rôle dans la manipulation de l'information.

Cela nous amène à examiner les cadres de gouvernance qui existent en plus du droit traditionnel, à savoir les pratiques de modération du contenu des entreprises. Il convient de noter que même dans les pays dotés d'une législation complète en matière de responsabilité des intermédiaires, les politiques de modération du contenu jouent un rôle important. Plusieurs raisons expliquent ce phénomène. Les entreprises sont incitées à modérer le contenu afin de traiter même le contenu légal mais gênant, bien que l'absence de modération du contenu soit le modèle commercial de certaines plateformes. Ces entreprises sont mondiales et les politiques permettent de créer des normes d'application générale. Une autre raison est que peu de plaintes sont portées devant les tribunaux. En droit civil, les procès sont trop coûteux et trop lents pour valoir la peine pour la plupart des plaignants. En droit criminel, on sait que les préjudices en ligne sont rarement signalés et font l'objet de peu d'enquêtes. Les groupes vulnérables hésitent souvent à se plaindre à la police, et les forces de l'ordre n'ont pas toutes les ressources ou les connaissances spéciales nécessaires pour enquêter sur certaines plaintes. De plus, il semblerait que ces plaintes ne soient pas prises au sérieux[212].

Modération du contenu par les médias sociaux

On prétend souvent que les utilisateurs ont des « droits » par rapport aux entreprises de médias sociaux ou qu'une plateforme a violé leur droit à la liberté d'expression garanti par la *Charte* en raison d'une décision de modération du contenu. La *Charte* ne s'applique pas aux activités des entreprises privées, à moins que celles-ci n'entreprennent des activités gouvernementales[213]. Cela veut dire qu'en général, les utilisateurs n'ont pas le droit de s'exprimer librement sur les médias sociaux, car ceux-ci relèvent du secteur privé. Cela ne signifie pas que le droit à la liberté d'expression n'a aucune signification juridique en ce qui concerne la modération du contenu par les plateformes. Par exemple, les gouvernements doivent se conformer à la *Charte* dans la mise en œuvre de toute loi. Ainsi, quel que soit le projet de loi sur les préjudices en ligne présenté par le gouvernement du Canada, il devra être conforme à la *Charte*[214].

Si le fait que les droits semblent privatisés dans les espaces numériques crée un certain malaise, cette préoccupation est légitime. La manière dont une plateforme interprète la liberté d'expression, par exemple, qu'elle soit fondée sur les valeurs de l'entreprise, le droit national

---

[212] Une partie de ce problème sera examinée dans un rapport à paraître du Conseil des académies canadiennes : *La sécurité publique à l'ère du numérique*, en ligne : https://www.rapports-cac.ca/reports/la-securite-publique-a-lere-du-numerique/.

[213] Voir articles 1 et 32 de la *Charte*, *supra* note 135. La *Charte* s'applique aux pouvoirs législatif, judiciaire et exécutif et dans les cas où le gouvernement a délégué son autorité à une partie privée ou que cette partie agit en tant qu'agent de l'État. Cet aspect a été étudié par des chercheurs et spécialistes, mais pas dans la même mesure que dans l'analyse de l'application horizontale indirecte de la *Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales* de 1950. La législation sur les préjudices en ligne devra être examinée de près si le gouvernement délègue délibérément ses pouvoirs, car cela pourrait entraîner un autre type d'examen de la *Charte* pour les activités des plateformes.

[214] Les tribunaux doivent également adapter la common law aux valeurs de la *Charte*. Par exemple, le droit à la vie privée et à la libre expression a influencé l'évolution des délits de diffamation et d'atteinte à la vie privée : *Hill*, *supra* note 167; *Jones*, *supra* note 182.

(souvent le premier amendement) ou les droits de la personne internationaux, est un système de gouvernance privée de sa conception sans aucune des caractéristiques normales de responsabilité auxquelles nous nous attendons des systèmes gérés par l'État[215]. La décision de plusieurs médias sociaux de bannir l'ancien président Trump de leur plateforme, par exemple, est importante du point de vue de la gouvernance privée, invite à examiner qui fixe les conditions de modération, et met en évidence le pouvoir énorme de ces plateformes. Facebook dispose d'un mécanisme d'appel officiel, le Oversight Board, qui a examiné la décision de bannir le président Trump et a conclu que la décision de restreindre l'accès était appropriée, mais que la sanction de suspension indéfinie ne l'était pas[216].

Un phénomène similaire est observable dans ce qu'Elena Chachko appelle la « sécurité nationale par plateforme »[217]. Comme elle l'explique, les plateformes sont maintenant au cœur de la géopolitique et de la sécurité[218]. De nombreux médias sociaux collaborent avec les gouvernements, emploient des directeurs de politique étrangère, des équipes d'intervention en cas d'incident, des mécanismes officiels de modération du contenu et d'appel, des équipes chargées de la confiance et de la sécurité et des politiques axées sur les questions de sécurité nationale, telles que la mésinformation et la désinformation, l'intégrité des élections, le terrorisme et l'extrémisme violent. La plupart des médias sociaux assument ce rôle par nécessité, en raison de la manière dont leurs plateformes sont utilisées et exploitées. Cependant, certaines plateformes épousent des idéologies politiques qui influencent leur conception et la modération de leur contenu. Un risque important est le fait que, puisque cet arrangement est souvent indirect et informel, un acteur privé puisse « choisir les fonctions qu'il souhaite remplir »[219] [traduit par nos soins]. En effet, les plateformes peuvent choisir de ne pas s'engager du tout[220]. Cela déstabilise la protection de la sécurité nationale, car le gouvernement n'exerce qu'une surveillance minimale sur ce que font les plateformes, les problèmes sont nouveaux et les plateformes disposent d'une énorme marge de manœuvre pour décider de la marche à suivre, s'il y a lieu[221]. On a pu observer cela avec la réaction de Facebook face à la diffusion de contenus violents et extrémistes, ainsi qu'à la mésinformation et à la désinformation, au Myanmar, et sa contribution à la violence contre les Rohingyas. Le rapport

---

[215] Par privatisation, je veux dire qu'une partie privée remplit une fonction normalement réservée au gouvernement. Ce phénomène a été décrit comme un basculement : les droits de la personne ont été initialement structurés dans le cadre des relations entre les citoyens et l'État, et aujourd'hui, avec les technologies numériques, nous exerçons et vivons nos droits dans le cadre de relations entre utilisateurs et entreprises technologiques : voir Emily B. Laidlaw, *Regulating Speech in Cyberspace: Gatekeepers, Human Rights and Corporate Responsibility* (Cambridge University Press, 2015), ch. 6.

[216] Oversight Board, Décision 2021-001-FB-FBR, en ligne : https://www.oversightboard.com/decision/FB-691QAMHJ.

[217] Elena Chachko, « National Security by Platform » (2021) 25 *Stanford Technology Law Review*, p. 55.

[218] *Ibid*.

[219] *Ibid* p. 125.

[220] *Ibid* p. 127. Ce problème de manque d'orientation et d'informalité est observable dans la réglementation de la technologie en général, un problème signalé par des spécialistes de la réglementation de la technologie depuis des années.

[221] *Ibid.*

du Conseil des droits de l'homme, *Rapport de la mission d'enquête internationale indépendante sur le Myanmar*, est allé jusqu'à qualifier la réponse de Facebook de « lente et inefficace »[222].

La modération du contenu n'est pas entièrement volontaire. Il s'agit plutôt d'une étape importante dans l'accomplissement de la responsabilité des entreprises auxquelles il incombe de respecter les droits de la personne, conformément aux *Principes directeurs des Nations Unies relatifs aux entreprises et aux droits de l'homme*[223] (les « Principes directeurs »). Ces principes imposent aux entreprises des obligations de diligence raisonnable concernant leur incidence sur les droits de la personne. En d'autres termes, les entreprises doivent éviter de nuire aux droits de la personne, surveiller leur conformité et s'efforcer de prévenir et d'atténuer les dommages, et fournir un accès à un mécanisme de réparation. Les Principes directeurs constituent le plan directeur des entreprises pour l'intégration des droits de la personne dans leurs activités et pour leur reddition de comptes, mais ils reposent toujours sur un engagement de bonne foi, et bon nombre des politiques de modération du contenu des médias sociaux utilisés lors du Convoi ne font aucune allusion aux Principes directeurs et ne les reflètent pas[224].

Par ailleurs, les efforts des entreprises sont souvent le fruit d'une collaboration ou sont encouragés par le gouvernement, sous la forme d'une coréglementation, comme le *Code de bonnes pratiques contre la désinformation* de l'UE mentionné plus haut[225]. Un autre exemple est le Global Internet Forum to Counter Terrorism (GIFCT) (Forum mondial de l'Internet contre le terrorisme), qui est une collaboration entre divers services en ligne visant à lutter contre le terrorisme et l'extrémisme violent. Il a été créé grâce à la collaboration de diverses parties intéressées, et comprend, outre ses fondateurs du secteur, des représentants du monde universitaire, de la société civile et d'organismes tels que la Direction exécutive du Comité contre le terrorisme des Nations Unies et l'Union européenne[226]. Le GIFCT travaille à l'élaboration de mécanismes de prévention et d'intervention en cas d'incident.

---

[222] Conseil des droits de l'homme, *Rapport de la mission d'enquête internationale indépendante sur le Myanmar* (2018), A/HRC39/64, en ligne (en anglais) : https://www.ohchr.org/sites/default/files/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf at para 74.

[223] Voir Haut-Commissariat des droits de l'homme, *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* (2011), HR/PUB/11/04, en ligne : https://www.ohchr.org/sites/default/files/Documents/Publications/GuidingPrinciplesBusinessHR_FR.pdf.

[224] Les Principes directeurs sont ancrés dans la licence sociale d'exploitation d'une entreprise. Ils ont été approuvés par le Conseil des droits de l'homme, qui les a fait passer du statut de guide à celui de système de gouvernance. Voir John Ruggie, *Just Business: Multinational Corporations and Human Rights* (Norton, 2013); David Kaye, *Speech Police: The Global Struggle to Govern the Internet* (New York: Columbia Global Reports, 2019). Voir Meta, « Politique relative aux droits humains au sein de l'entreprise », en ligne : https://humanrights.fb.com/fr/policy/; Google, « Droits de l'homme », en ligne : https://about.google/human-rights/.

[225] *Code de bonnes pratiques contre la désinformation*, *supra* note 195.

[226] Voir en ligne : https://gifct.org/; Chachko, *supra* note 217, p. 89.

Il convient d'examiner deux aspects de la modération du contenu. Premièrement, les entreprises de médias sociaux utilisent généralement des technologies, sous une forme ou une autre, pour réglementer les contenus préjudiciables. Bien qu'elles soient souvent présentées comme des solutions techniques, il s'agit de systèmes de gouvernance mis en œuvre techniquement. Deuxièmement, les médias sociaux réglementent le comportement des utilisateurs au moyen de leurs politiques de modération du contenu. Cet aspect sera examiné sous l'angle des médias sociaux utilisés dans le cadre du Convoi.

La technologie de la modération du contenu

La technologie est essentielle pour combattre la manipulation de l'information[227]. Cependant, ce n'est pas une panacée pour les contenus préjudiciables. Elle peut être imprécise, manquer de la finesse nécessaire pour évaluer un contenu ambigu avec précision et en fonction du contexte, et elle est façonnée par l'état d'esprit (et le biais potentiel) du créateur des données, avec une surveillance minimale externe à l'organisation[228].

Il existe de nombreux outils automatisés de modération de contenu utilisés pour filtrer, classer, traiter et organiser le contenu. Beaucoup sont guidés par l'intelligence artificielle[229]. Par exemple, la technologie peut aider à détecter des comptes non authentiques[230]. Le hachage perceptif, tel que celui du logiciel PhotoDNA de Microsoft, est une empreinte numérique utilisée pour détecter les images et les vidéos préjudiciables, comme celles d'abus sexuels sur des enfants, des contenus terroristes et extrémistes violents, ou des contenus contrevenant aux droits d'auteur[231]. Le GIFCT est à l'origine d'une base de données commune pour ses membres[232]. Le projet Arachnid utilise le hachage pour repérer les contenus d'abus pédosexuels[233]. D'autres outils comprennent la reconnaissance d'images pour donner la priorité aux contenus à examiner par un humain, et les techniques de traitement du langage naturel pour détecter les discours haineux et les contenus extrémistes[234].

---

[227] Voir Kreps, *supra* note 62, p. 6-7.

[228] Voir Spandana Singh, « Everything in Moderation » (22 juillet 2019), *New American Foundation*, en ligne : https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation.

[229] Voir limites de l'IA dans Alex Feerst, *The Use of AI in Online Content Moderation or, the tech sector invested in automation and all I got was this questionable adjudication* (septembre 2022) Digital Governance Working Group, en ligne : https://platforms.aei.org/wp-content/uploads/2022/09/The-Use-of-AI-in-Online-Content-Moderation.pdf.

[230] C'est le générateur de textes d'IA pour les fausses informations qui peut également être utilisé pour les identifier. Kreps parle de Grover, un modèle qui, à la fois, génère et identifie les « fausses nouvelles », *supra* note 62, p. 6-7.

[231] Voir Content ID de YouTube (adapté de PhotoDNA) et Google Drive.

[232] Elle se limite aux entités terroristes figurant sur les listes des groupes terroristes désignés par les Nations Unies. Le GIFCT est une ONG fondée en 2017 par Facebook (désormais Meta), Microsoft, Twitter et YouTube, et a élargi le nombre de ses membres depuis lors : *supra* note 226.

[233] Voir en ligne : https://www.projectarachnid.ca/fr/.

[234] Singh, *supra* note 228.

La technologie est également utilisée pour inciter les utilisateurs à modifier leur comportement[235]. Actuellement, ces stratégies sont nécessairement expérimentales, parce que la recherche sur leur efficacité n'en est qu'à ses débuts[236]. Par exemple, un outil couramment utilisé par les services en ligne pour lutter contre la manipulation de l'information est la correction de l'information. Cet outil est intéressant, parce que l'interférence avec le droit à la liberté d'expression est relativement mineure[237]. Les utilisateurs ont toujours accès à l'information, mais le contenu est signalé comme étant une fausse information ou provenant d'un faux compte. La correction de l'information est-elle efficace? Les résultats de la recherche sont mitigés. L'argument le plus solide contre la correction de l'information a été exposé dans un article rédigé en 2010 par Brendan Nyhan et Jason Reifler. Il s'agit de ce qu'ils ont appelé « l'effet de retour de flamme » [traduit par nos soins], c'est-à-dire que le fait de démystifier une fausse information est inefficace et peut avoir l'effet inverse et renforcer les perceptions erronées des lecteurs[238]. Toutefois, il a été démontré depuis que l'effet de retour de flamme est exagéré, que le problème est plus subtil et qu'il faut poursuivre les recherches pour mesurer l'efficacité de la correction de l'information[239]. Par exemple, celle-ci peut avoir l'effet d'un « écho des croyances »[240] [traduit par nos soins]. Et comme toute information diffusée donne l'illusion d'être crédible, il y a un risque que la correction de l'information donne à l'histoire une illusion de vérité[241]. La réponse à la correction de l'information pourrait être la douceur avec laquelle elle est effectuée; c'est ainsi que Facebook est passée de la correction de l'information à la présentation d'un éventail diversifié de nouvelles sur le sujet concerné[242]. Une autre solution pourrait consister à prévoir la correction pour le moment où le public prend conscience de la fausseté de l'information[243].

---

[235] Le « *nudging* » est la théorie de Richard Thaler et Cass Sunstein selon laquelle une architecture de choix indirecte et subtile est efficace pour inciter à des changements de comportement, comme le fait d'imposer le choix du don d'organes avec le renouvellement du permis de conduire : *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Penguin Book, 2008).

[236] On peut voir des plateformes effectuer des changements à la suite de nouvelles recherches : Tessa Lyons, « Replacing Disputed Flags With Related Articles » (20 décembre 2017) *Meta,* en ligne : https://about.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/.

[237] Helm, *supra* note 101, p. 315-318.

[238] Brendan Nyhan et Jason Reifler, « When Corrections Fail: The Persistence of Political Misperceptions » (2010) 32 *Political Behaviour* 303; Helm, *supra* note 101, p. 315-318; Wardle et Derakhshan, *supra* note 27, p. 45; Timothy Caulfield, « Does Debunking Work? Correcting COVID-19 Misinformation on Social Media » dans Colleen Flood et al., *Vulnerable: The Law, Policy and Ethics of COVID-19* (Presses de l'Université d'Ottawa, 2020), p. 188-193.

[239] Voir l'étude de la documentation de Caulfield, *supra* note 238. Il conclut en disant : « même si l'effet de retour de flamme peut se produire dans certaines circonstances, c'est un domaine dans lequel il serait utile d'approfondir les recherches; il ne s'agit certainement pas d'un phénomène robuste et mesurable au point qu'il devrait nous empêcher de déployer des efforts pour contrer la mésinformation sur les médias sociaux » [traduit par nos soins] : p. 190.

[240] Thorson, *supra* note 95.

[241] Caulfield, *supra* note 238, p. 190-191.

[242] Voir Lyons, *supra* note 236.

[243] Caulfield, *supra* note 238, p. 191-192. Caulfield énumère plusieurs principes qui peuvent maximiser les effets de la correction de l'information : utiliser les faits; communiquer clairement et simplement; utiliser des sources fiables, bien que la confiance soit un défi; établir qu'il y a un consensus scientifique le cas échéant; être aimable et authentique; écrire dans un style narratif; utiliser des arguments rationnels en soulignant les lacunes de la logique,

De même, la suppression et le blocage du contenu sont parfois utilisés. Je les inclus dans la discussion sur les solutions techniques, bien qu'il s'agisse souvent d'un mélange d'actions automatisées et humaines. Une question qui se pose est celle de l'efficacité de la suppression du contenu. Il n'est pas évident qu'elle soit efficace pour changer les croyances, parce qu'il n'y a pas de nouvelle information pour remplacer ce qui a été supprimé[244]. De plus, le contenu problématique est retransmis presque instantanément après sa publication, et il est donc rarement retiré de la circulation. La vidéo en direct de la fusillade de Buffalo a été supprimée par Twitch en deux minutes, mais elle avait déjà été copiée et rediffusée sur diverses plateformes[245]. Le retrait peut également servir de balise en attirant plus d'attention sur le message ou en renforçant les croyances[246]. La suppression de comptes peut réussir à perturber l'élan d'un groupe. Il perd des adeptes et peine à en gagner de nouveaux, ce qui perturbe la monétisation. Les membres peuvent passer à d'autres plateformes, mais le groupe ne retrouve pas son niveau précédent[247]. Le retrait de contenu pourrait avoir une fonction différente de celle de changer les croyances, en jouant un rôle expressif renforçant ce qui est un comportement acceptable, bien qu'il faille veiller à équilibrer les différents droits[248].

Sur le plan de la législation et de la gouvernance, nous sommes dans une période d'expérimentation technique et réglementaire. Il n'existe pas de consensus sur les stratégies qui serviront à lutter contre la manipulation de l'information. Par conséquent, les solutions changent et se retournent parfois contre nous. En raison du pouvoir social de certaines plateformes, le retour de flamme peut être énorme. Le résultat est un mélange d'interventions. Par exemple, l'application de messagerie WhatsApp a récemment mis à jour ses caractéristiques techniques afin d'améliorer la protection de la vie privée des utilisateurs, de leur permettre de quitter des groupes sans en avertir les canaux, de cacher qu'ils sont en ligne et de bloquer la capture d'écran des messages destinés à être consultés une seule fois[249]. Ce sont là des solutions positives propres à cette application. Twitter intègre la théorie du « *nudge* » à l'aide d'outils techniques[250]. Le « *nudge* » incite les utilisateurs à repenser à la

---

etc.; encadrer la correction de l'information pour mettre l'accent sur les faits et non sur la mésinformation; le public doit être le grand public et non les personnes qui croient en la mésinformation : p. 193-198.

[244] Voir discussion, Helm, *supra* note 101, p. 321.

[245] Mia Sato, « How the Buffalo shooting livestream went viral » (17 mai 2022) *Verge*, en ligne : https://www.theverge.com/2022/5/17/23100579/buffalo-shooting-twitch-livestream-viral-content-moderation.

[246] C'est ce qu'on appelle l'effet Streisand. On en discute également dans Helm, *supra* note 101, p. 321-322.

[247] Voir la discussion sur l'analyse documentaire d'Amarnath Amarasingam : « Does Deplatforming Work? A quick survey of literature in the wake of the Capitol Hill Attack » 12 janvier 2021) *Intrepid*, en ligne : https://www.intrepidpodcast.com/blog/2021/1/12/does-deplatforming-work-a-quick-survey-of-literature-in-the-wake-of-the-capitol-hill-attack. Voir la monétisation de YouTube pour la comparer à celle de Rumble, BitChute et Odysee. Les stratégies de monétisation plus généreuses de certaines nouvelles plateformes de partage de vidéos méritent d'être étudiées pour leurs répercussions sur les plateformes de partage de vidéos traditionnelles.

[248] En s'appuyant sur l'argument selon lequel l'un des objectifs de la loi est de renforcer ou de modifier les normes, le retrait du contenu devrait être fondé sur les principes des droits de la personne.

[249] Michelle Toh, « WhatsApp is going to stop letting everyone see when you're online » (9 août 2022) *CNN,* en ligne : https://www.cnn.com/2022/08/09/tech/whatsapp-privacy-changes-meta-intl-hnk/index.html.

[250] Thaler et Sunstein, *supra* note 235.

publication des gazouillis contenant des propos injurieux et à lire leurs messages avant de les diffuser[251].

En revanche, la modification apportée par Facebook à son algorithme visant à améliorer le bien-être des utilisateurs s'est retournée contre elle. Vers 2017-2018, Facebook a modifié son algorithme de classement de l'engagement[252] pour favoriser les interactions sociales significatives. Les publications populaires et celles des amis et des proches ont été amplifiées et les nouvelles professionnelles ont été désamplifiées. Les Facebook Files divulguées par Frances Haugen ont également révélé qu'une partie des ajustements algorithmiques impliquait de donner un coup de pouce aux publications qui génèrent de fortes réactions émotionnelles sous forme d'émoticônes. Les émoticônes « amour », « rire », « tristesse » et « colère » ont cinq fois plus de valeur que l'émoticône « j'aime »[253]. Des recherches internes ont montré que les publications qui suscitaient des réactions de colère en émoticônes étaient « disproportionnellement susceptibles de contenir des informations erronées, toxiques et de mauvaise qualité »[254] [traduit par nos soins]. Par conséquent, l'algorithme de Facebook a favorisé la diffusion de mésinformation et de désinformation[255].

Politiques de modération du contenu

En ce qui concerne le Convoi, les discussions qui ont alimenté sa création ont commencé bien avant janvier 2022 sur divers médias sociaux grand public et alternatifs, dans des groupes discutant de l'obligation de vaccination, des restrictions liées à la COVID-19 et des théories du complot, et elles ont été amplifiées par des influenceurs et des organismes d'information alternatifs. Le public et les participants étaient acquis. Les risques systémiques associés aux services des médias sociaux sont un élément clé de leur responsabilité. Comment sont-ils conçus? Comment l'algorithme fonctionne-t-il? Comment surveillent-ils les répercussions de leurs services, et donnent-ils suite à leurs constatations? À l'exception de la protection de la vie privée, ils n'ont aucune obligation légale de gérer les risques systémiques, d'autant plus qu'une grande partie du contenu est légale. L'opacité des médias sociaux, au-delà des rapports de

---

[251] Twitter Safety, « How Twitter is nudging users to have healthier conversations » (1er juin 2022), en ligne : https://blog.twitter.com/common-thread/en/topics/stories/2022/how-twitter-is-nudging-users-healthier-conversations.

[252] Le classement de l'engagement est controversé et a fait l'objet d'une partie du témoignage de Frances Haugen. Une partie de la controverse vient du fait que l'augmentation de l'engagement est considérée comme intéressée dans la mesure où elle permet de garder les utilisateurs sur Facebook : Jeremy B. Merrill et Will Oremus, « Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation » (26 octobre 2021) *Washington Post*, en ligne : https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/.

[253] Voir « The Facebook Files », *Wall Street Journal*, en ligne : https://www.wsj.com/articles/the-facebook-files-11631713039.

[254] Merrill et Oremus, *supra* note 252.

[255] Je mentionne à la fois la mésinformation et la désinformation, parce que l'algorithme de Facebook a été exploité pour des opérations d'information, comme les publicités d'origine russe sur Meta, « An Update On Information Operations On Facebook » (6 septembre 2017), en ligne : https://about.fb.com/news/2017/09/information-operations-update/.

transparence, qui ne sont pas encore mûrs et normalisés pour cette industrie, signifie que cet aspect de ce qui a alimenté la mise en place du Convoi est une question d'autoréglementation des médias sociaux. Par ailleurs, bien que la structure de monétisation des différents médias sociaux ne soit pas explorée en profondeur dans le présent document, la Commission pourrait envisager de pousser plus loin ses recherches sur la motivation financière de certains influenceurs et sur les pratiques de monétisation des médias sociaux utilisés dans le Convoi. Par exemple, dans le cadre du Programme Partenaire de YouTube, un créateur gagnerait de l'argent grâce aux publicités qui entourent ses vidéos sur YouTube. Le contenu produit par ces influenceurs pourrait alors être soumis à un examen en vertu des lignes directrices de modération et d'autres politiques de monétisation[256].

Les politiques de modération du contenu des médias sociaux utilisés dans le Convoi varient considérablement. Comme nous l'avons observé, les partisans et les organisateurs du Convoi utilisaient généralement Facebook, Twitter, TikTok, YouTube, Rumble, Telegram, Zello, BitChute, Odyssey, GoFundMe et GiveSendGo. Les plateformes grand public ont des politiques de modération de contenu relativement élaborées, bien que la présente analyse ne tienne pas compte de l'efficacité ou de la légitimité de leurs méthodes, ni de leur application[257]. En utilisant le mot « élaborées », je veux simplement dire que ces plateformes ont une ou plusieurs politiques qui traitent de manière substantielle des risques de préjudice de leurs services, ainsi qu'un système permettant d'agir sur les contenus qui enfreignent ces politiques, comprenant notamment un mécanisme qui permet aux utilisateurs de signaler les contenus et un mécanisme d'appel. Certains médias sociaux correspondent théoriquement à ce profil, mais ne font qu'une modération minimale. Et une différence importante entre les divers médias sociaux utilisés dans le Convoi est la mesure dans laquelle ils modèrent de manière proactive les contenus préjudiciables[258].

Facebook et Twitter, par exemple, ont diverses politiques en matière de discours haineux, de terrorisme et d'extrémisme violent, de violence, de manipulation des médias, de faux comptes, etc.[259] Les sujets couverts sont similaires, mais les politiques ne le sont pas, ce qui illustre les différences entre les plateformes, mais aussi les différences d'éthique en ce qui concerne leur

---

[256] YouTube, « Politiques de monétisation », en ligne : https://www.youtube.com/howyoutubeworks/policies/monetization-policies/.

[257] Pour une analyse complète de la modération du contenu et de la législation sur les droits de la personne, voir Mackenzie Common, *Rule of law and human rights issues in social media content moderation* (2020) thèse de doctorat, London School of Economics and Political Science; Douek, *supra* note 183.

[258] La nécessité d'une modération proactive du contenu est un problème important, mais il est difficile de concilier cette nécessité avec celle de ne pas imposer de systèmes généraux de surveillance ou de ne pas compromettre le cryptage, qui peuvent constituer une atteinte à la vie privée. La ligne de démarcation se trouve généralement entre la surveillance générale et la surveillance ciblée, comme la recherche proactive de contenu haché qui est précisément du contenu pédopornographique, du contenu terroriste ou extrémiste violent, mais il y a beaucoup de contenu dans la zone grise qui manipule les utilisateurs.

[259] Voir « Les règles de Twitter », en ligne : https://help.twitter.com/fr/rules-and-policies/twitter-rules; voir les « Standards de la communauté Facebook » de Meta, en ligne : https://transparency.fb.com/fr-fr/policies/community-standards/.

position sur certaines questions[260]. Les politiques servent notamment à souligner et à confirmer ce qui est illégal (p. ex. proférer des menaces)[261], mais elles fixent également des règles concernant ce qui, au-delà de la réglementation, est acceptable. La modération du contenu est donc essentielle pour traiter les propos licites mais offensants. En matière de discours haineux, les politiques fixent la barre de l'expression acceptable plus haut que le droit criminel ou les droits de la personne[262]. Par exemple, Facebook définit un discours haineux comme « un discours violent ou déshumanisant, des stéréotypes offensants, une affirmation d'infériorité, une expression de mépris, de dégoût ou de renvoi, une insulte ou un appel à l'exclusion ou à la ségrégation »[263]. Facebook traite la mésinformation comme une question d'intégrité et d'authenticité. Ce réseau dispose d'une politique en matière de désinformation, qui cible les types de désinformation : atteinte à l'intégrité physique et violence, santé, élections, ingérence dans le recensement, et manipulation des médias[264]. Twitter a instauré une politique sur la mésinformation en cas de crise en mai 2022[265]. La désinformation est traitée séparément dans des politiques portant, par exemple, sur le pourriel, les comptes coordonnés ou les comportements non authentiques.

Certains des mécanismes plus souples utilisés par les médias sociaux sont importants parce qu'ils vont au-delà du modèle binaire « laisser/supprimer ». Comme nous l'avons vu plus haut, la correction ou la diversité de l'information, la rétrogradation ou la limitation de la visibilité, les avertissements, les étiquettes et les invitations à repenser ou à lire avant de diffuser sont autant de formes de friction stratégique. Pour une modération plus officielle du contenu, Facebook a recours à une approche hybride utilisant une combinaison d'examen algorithmique et humain, ainsi qu'un examen proactif et basé sur les plaintes. Lorsqu'un utilisateur envoie un contenu à publier, celui-ci est soumis à un filtrage visant à déterminer s'il correspond aux bases de données de hachage de matériel pédopornographique et de contenu terroriste. En cas de correspondance, la publication du contenu est bloquée. Une fois le contenu en ligne, Facebook le surveille à l'aide d'algorithmes permettant d'identifier les contenus répréhensibles en fonction de divers paramètres, comme des mots, des images ou des comportements considérés comme généralement associés aux différents types de contenus répréhensibles, l'identité de l'auteur, le contexte de la publication et les commentaires. Si l'algorithme détermine que le contenu viole clairement les normes de la communauté, il est supprimé, mais si l'algorithme n'est pas clair, un modérateur humain examine le contenu. Les utilisateurs peuvent également signaler qu'un contenu enfreint les règles de la communauté. Les sanctions en cas d'infraction

---

[260] Pendant longtemps, Twitter a résisté à l'idée d'imposer une modération plus stricte du contenu par fidélité à une approche fondée sur le premier amendement, mais ces dernières années, ce média social a commencé à élaborer des pratiques de modération plus complètes.

[261] *Code criminel*, *supra* note 40, art. 264.1.

[262] *Ibid*, art. 319, qui interdit l'incitation publique à la haine ou la fomentation volontaire de la haine ou de l'antisémitisme; *Keegstra*, *supra* note 137; *Whatcott*, *supra* note 168.

[263] Meta, « Discours haineux », en ligne : https://transparency.fb.com/fr-fr/policies/community-standards/hate-speech/.

[264] Meta, « Fausses informations », en ligne : https://transparency.fb.com/fr-fr/policies/community-standards/misinformation/.

[265] Yoel Roth, « Introducing our crisis misinformation policy » (19 mai 2022), en ligne : https://blog.twitter.com/en_us/topics/company/2022/introducing-our-crisis-misinformation-policy.

vont de l'avertissement à la désactivation permanente ou temporaire du compte, en passant par la restriction de l'accès à certaines fonctionnalités de Facebook, comme la diffusion en direct. Les utilisateurs peuvent faire appel de la décision, notamment auprès du Oversight Board de Facebook[266]. Facebook a supprimé certains groupes, pages et comptes Facebook liés au Convoi, comme ceux des polluposteurs, qui profitaient du Convoi pour attirer les utilisateurs vers des sites Web hors plateforme présentant des publicités payantes, des groupes haineux et des groupes conspirationnistes, comme QAnon[267].

Twitter utilise diverses techniques de friction, mais, comme Facebook, elle utilise également une approche hybride reposant sur l'automatisation et l'examen humain. Les utilisateurs peuvent signaler les contenus qui violent les règles de Twitter. Les sanctions sont graduelles et vont de l'action précise sur les tweets (étiquettes, visibilité, suppression) à la restriction des messages, ou à la restriction au niveau du compte (mode lecture seule, vérification, suspension). Les utilisateurs peuvent faire appel d'un compte bloqué ou suspendu[268]. Twitter a suspendu définitivement un compte du Convoi et celui d'un influenceur[269].

TikTok utilise également des techniques de friction, ainsi qu'un système hybride d'examen automatisé et humain, et, comme on l'a vu ci-dessus, utilise une approche graduelle d'avertissements, de suspensions temporaires puis permanentes de comptes, avec la possibilité de faire appel. Pour certains contenus, comme le contenu pédopornographique, elle applique une politique de tolérance zéro[270]. TikTok interdit la désinformation nuisible qui cause un préjudice important[271]. Plusieurs influenceurs utilisent régulièrement TikTok, mais au moment de la rédaction de ce document, je ne suis pas au courant de mesures prises à l'encontre de comptes liés au Convoi.

La modération du contenu est nettement différente pour ce que l'on a décrit comme des médias sociaux alternatifs. Cela ne fait qu'étouffer les problèmes, car les utilisateurs passent à

---

[266] Oversight Board, « Faites appel pour modeler le futur de Facebook et d'Instagram », en ligne : https://www.oversightboard.com/appeals-process/; Plus généralement voir, en ligne : https://transparency.fb.com/fr-fr/.

[267] Culliford, *supra* note 5.

[268] Twitter, « Notre gamme d'options pour l'application de nos politiques », en ligne : https://help.twitter.com/fr/rules-and-policies/enforcement-options.

[269] Kevin Jiang, « Ontario MPP Randy Hillier 'permanently suspended' from Twitter » (8 mars 2022) *Toronto Star,* en ligne : https://www.thestar.com/politics/provincial/2022/03/08/ontario-mpp-randy-hillier-suspended-from-twitter.html.

[270] TikTok, « Violations du contenu et bannissements », en ligne : https://support.tiktok.com/fr/safety-hc/account-and-user-safety/content-violations-and-bans.

[271] Conformément à la définition de TikTok dans ses « Règles communautaires », la désinformation portant préjudice désigne les publications « susceptibles de nuire gravement à autrui, à notre communauté ou au grand public, et ce quelle que soit l'intention de départ. Parmi les préjudices graves figurent les blessures physiques sérieuses, les maladies ou les décès, les traumatismes psychologiques aigus, les dégâts matériels à grande échelle et l'atteinte à la confiance du public dans les institutions et les processus civiques tels que les gouvernements, les élections et les organismes scientifiques. Nous n'incluons pas ici des informations purement inexactes, des légendes ou des atteintes à la réputation des personnes ou des entreprises. », en ligne : https://www.tiktok.com/community-guidelines?lang=fr#37.

des plateformes moins modérées, que ce soit pour éviter les règles de modération ou parce que leur compte a été suspendu. C'est ce que on a pu observer avec la suspension par GoFundMe de la campagne du Convoi, qui s'est ensuite déplacée sur GiveSendGo[272].

Comme on l'a mentionné, le Convoi était un mouvement qui se nourrissait de vidéos. YouTube était utilisée par les organisateurs et les partisans du Convoi, mais les utilisateurs se sont également tournés vers d'autres plateformes de diffusion de vidéos, principalement BitChute, Rumble et Odysee. Comme les autres médias sociaux grand public, YouTube a des directives qui interdisent de manière générale les pratiques trompeuses, le harcèlement, le discours haineux et les contenus nuisibles ou dangereux ou d'autres contenus similaires[273]. L'application de ces directives est assurée par une combinaison d'examens automatisés et humains, ainsi que par le signalement des utilisateurs[274]. En revanche, BitChute a des directives pour sa communauté, et une méthode pour signaler du contenu et faire appel des décisions[275]. Toutefois, la modération est principalement basée sur les signalements des utilisateurs et ne semble pas comprendre de mesures proactives. De plus, et c'est là un point crucial, les directives à l'intention des utilisateurs n'interdisent principalement que le contenu illégal, ce qui fait que BitChute est devenue un refuge pour l'expression d'opinions extrêmes. La plateforme ne limite le contenu haineux que s'il est illégal, comme l'incitation à la haine. Elle interdit à peine les contenus terroristes et extrémistes des entités désignées dans la législation antiterroriste, qui ne traitent pas des groupes nationalistes blancs et des préjudices plus larges qui touchent à la modération des médias sociaux[276]. Elle n'a aucune politique concernant la mésinformation et la désinformation ou d'autres formes de manipulation de l'information, notamment la manipulation des médias ou les faux comptes, bien qu'elle interdise le pourriel et le harcèlement en bande organisée.

---

[272] Amanda Connolly, « GoFundMe, GiveSendGo defend handling of convoy blockade fundraising campaigns » (3 mars 2022) *Global News*, en ligne : https://globalnews.ca/news/8656947/gofundme-givesendgo-convoy-blockade-campaigns/.

[273] YouTube, « Règlement de la communauté », en ligne : https://www.youtube.com/intl/fr_ca/howyoutubeworks/policies/community-guidelines/#:~:text=YouTube%20prend%20des%20mesures%20%C3%A0,%2C%20documentaire%2C%20scientifique%20ou%20artistique, et « Aperçu des politiques », en ligne : https://www.youtube.com/howyoutubeworks/policies/overview/.

[274] YouTube, « Comment YouTube applique-t-il son règlement de la communauté? », en ligne : https://www.youtube.com/intl/fr_ca/howyoutubeworks/policies/community-guidelines/#enforcing-community-guidelines.

[275] BitChute, « Content Moderation Policy », en ligne : https://support.bitchute.com/policy/content-moderation#flagging--reporting, et « Community Guidelines », en ligne : https://support.bitchute.com/policy/guidelines/.

[276] La désignation terroriste ne comprend pas les groupes d'extrême droite, bien que le Canada y ait ajouté, par exemple, les Proud Boys et Blood & Honour récemment : Sécurité publique Canada, « Entités inscrites actuellement », en ligne : https://www.securitepublique.gc.ca/cnt/ntnl-scrt/cntr-trrrsm/lstd-ntts/crrnt-lstd-ntts-fr.aspx. Bitchute a également sa propre liste d'entités interdites, mais il n'y en a que deux sur la liste : « Prohibited Entities List », en ligne : https://support.bitchute.com/policy/prohibited-entities-list.

Rumble, qui a hébergé l'une des vidéos qui ont contribué à susciter l'élan entourant le Convoi[277], adopte une approche de modération du contenu semblable à celle de BitChute et, par conséquent, ce réseau est également devenu populaire pour la diffusion de contenus extrémistes. Le PDG de Rumble décrit la plateforme comme « différente de YouTube et de Facebook, car elle utilise beaucoup moins d'algorithmes pour recommander et examiner le contenu »[278] [traduit par nos soins]. Les vidéos sont affichées par ordre chronologique pour les utilisateurs en fonction des personnes qu'ils suivent sur la plateforme. Rumble n'utilise pas d'algorithmes pour filtrer de manière proactive les contenus à haut risque. Bien que les conditions générales interdisent plus que le contenu illégal, le niveau d'exigence n'est pas beaucoup plus élevé[279]. Ce média n'a pas de politique en matière de mésinformation ou de désinformation.

Les lignes directrices pour l'utilisation d'Odysee interdisent de manière générale l'incitation à la haine ou à la violence, la promotion du terrorisme ou d'activités criminelles, et la violence qui ne mérite pas d'être signalée. La mésinformation et la désinformation, le contenu haineux et l'extrémisme sont permis. Les utilisateurs peuvent signaler des contenus et les sanctions comprennent la suppression de contenus, le blocage de commentaires ou le filtrage d'un canal utilisateur[280]. Cependant, la structure d'Odysee est particulière. Les vidéos ne sont pas stockées sur un serveur centralisé, mais plutôt décentralisées sur un réseau utilisant la technologie de la blockchain[281]. Cela signifie que les vidéos ne peuvent pas être supprimées définitivement, même par l'utilisateur qui les a téléchargées, bien qu'Odysee puisse en bloquer l'accès par l'intermédiaire de l'application[282].

Telegram a été activement utilisée pour organiser le Convoi et obtenir le soutien des gens. Telegram modère très peu son service. Comme application de messagerie, ce média est

---

[277] Broderick, *supra* note 7.

[278] Fizza Kulvi, « Meet Rumble, Canada's new 'free speech' platform – and its impact on the fight against online misinformation » (8 juillet 2021) *The Conversation*, en ligne : https://theconversation.com/meet-rumble-canadas-new-free-speech-platform-and-its-impact-on-the-fight-against-online-misinformation-163343.

[279] Rumble, « Website Terms and Conditions of Use and Agency Agreement », en ligne : https://rumble.com/s/terms; voir la discussion de Kevin Newman, « Investigating Canadian YouTube rival Rumble and its growing popularity among the world's far right » (19 février 2022) *CTV News*, en ligne : https://www.ctvnews.ca/w5/investigating-canadian-youtube-rival-rumble-and-its-growing-popularity-among-the-world-s-far-right-1.5787533.

[280] Odysee, « Community Guidelines », en ligne : https://odysee.com/@OdyseeHelp:b/Community-Guidelines:c, et « Signaler le contenu », en ligne : https://odysee.com/$/report_content?claimId=166ec880e443d4e1bca31dbd142bdf2a4a8aa61f&sunset=lbrytv.

[281] Eviane Leidig, « Odysee: The New YouTube for the Far-Right » (17 février 2021) *Global Network on Extremism and Technology*, en ligne : https://gnet-research.org/2021/02/17/odysee-the-new-youtube-for-the-far-right/#:~:text=Odysee's%20community%20guidelines%20state%20that,not%20allowed%20on%20the%20platform

[282] Odysee explique : « Nous ne pouvons pas supprimer du contenu publié de la blockchain elle-même, bien que nous puissions bloquer du contenu accessible depuis notre application ou d'autres services reposant sur la blockchain. » [traduit par nos soins], mandat, en ligne : https://odysee.com/$/tos. Étant donné que la blockchain est un grand livre immuable, les données qu'elle contient ne peuvent pas être modifiées : Eileen Brown, « Blockchain-based Odysee keeps your social media content online » (8 avril, 2021) *Zdnet*, en ligne : https://www.zdnet.com/finance/blockchain/blockchain-based-odysee-keeps-your-social-media-content-online/.

différent de tous les autres médias sociaux mentionnés ci-dessus. De nombreuses formes de modération utilisées par les plateformes grand public créent des risques importants pour la vie privée si elles sont utilisées pour modérer la messagerie privée[283]. Or, comme on l'a vu à la partie I, les groupes privés de Telegram peuvent compter jusqu'à 200 000 membres (alors qu'Instagram et iMessage limitent les discussions de groupe à 32 personnes) et leurs canaux permettent la diffusion à un nombre illimité d'abonnés[284]. Il est difficile de qualifier ces groupes de privés[285]. Telegram ne modère pas ses groupes ou canaux privés, sauf pour les signalements de pourriels. Par conséquent, les discours haineux, les contenus terroristes et extrémistes violents, la mésinformation et la désinformation, les contenus graphiques, les contenus pédophiles et autres contenus similaires ne sont pas modérés. Dans les groupes et canaux publics, les conditions de service interdisent uniquement la promotion de la violence et les contenus pornographiques illégaux[286].

L'application de walkie-talkie Zello a été utilisée pour organiser les barrages. Ses conditions d'utilisation et ses directives communautaires interdisent de manière générale tout comportement préjudiciable[287], y compris tout ce qu'un représentant de Zello juge inacceptable[288]. Elle interdit la promotion de l'extrémisme violent, mais donne une définition étroite du terrorisme comme étant lié aux organisations figurant sur les listes de sanctions[289]. La mésinformation et la désinformation ne sont pas explicitement mentionnées. Les utilisateurs peuvent signaler les violations, mais il n'y a pas d'autres informations sur le processus d'évaluation ou sur la modération proactive du contenu par Zello. Les sanctions pour violations comprennent la suspension ou la fermeture des comptes.

Les entreprises de médias sociaux n'ont aucune obligation légale d'aller plus loin que la loi en établissant les conditions d'utilisation de leur espace, et les médias sociaux qui vont trop loin sont raisonnablement critiqués[290]. Les Principes directeurs fournissent un plan pour l'élaboration de politiques de modération du contenu, mais il n'y a pas de mécanisme

---

[283] Le présent document n'examine pas non plus les risques pour la vie privée et la sécurité liés à l'affaiblissement du cryptage, mais il s'agit d'une question importante lorsqu'on examine le type d'obligation qu'une application de messagerie devrait avoir en matière de réglementation du contenu.

[284] Telegram, « FAQ », en ligne : https://telegram.org/faq#q-quelle-est-la-difference-entre-les-groupes-et-les-canaux; Sam Andrey, Alexander Rand et Karim Bardeesy, *Rebuilding Canada's Public Square* (Septembre 2021), en ligne : https://static1.squarespace.com/static/5e9ce713321491043ea045ef/t/615478c6a74009181c27d15e/1632925924146/RebuildingCanada%27sPublicSquare.pdf.

[285] Ce problème est difficile. Les groupes Facebook sont également privés et leur taille n'est pas limitée.

[286] Telegram, « Terms of Service », en ligne : https://telegram.org/tos/terms-of-service-for-telegram-premium.

[287] Zello, « Terms of Service », en ligne : https://zello.com/legal/terms/ : contenu « illicite, nuisible, menaçant, abusif, harcelant, délictuel, excessivement violent, diffamatoire, vulgaire, obscène, nu, partiellement nu ou sexuellement suggestif, pornographique, diffamatoire, portant atteinte à la vie privée d'autrui, haineux, racial, ethnique ou autrement répréhensible » [traduit par nos soins].

[288] *Ibid* , « Community Guidelines », en ligne : https://zello.com/community/user-guidelines/.

[289] Zello, *supra* note 287.

[290] Le présent document n'a pas pour but d'explorer la manière dont le droit à la liberté d'expression des utilisateurs fonctionne et devrait fonctionner au Canada en ce qui concerne l'accès aux médias sociaux. C'est un domaine dans lequel je fais actuellement des recherches.

d'application et l'on s'en remet donc à la bonne foi des entreprises. Si les médias sociaux agissent, c'est sous l'effet des incitatifs du marché, de la responsabilité sociale et de la pression publique. Comme nous l'avons vu, une grande partie des comportements à l'origine de mouvements comme le Convoi fonctionnent sur le mode de la combustion lente jusqu'à ce que quelque chose les pousse à l'action, et une grande partie du contenu qui alimente cette combustion lente est légale ou dans une zone grise.

Réforme des lois

Nous sommes dans une période de changement rapide en matière de réforme des lois visant à traiter les préjudices en ligne et la responsabilité des intermédiaires[291]. Au cours des dernières années, le rôle central des médias sociaux et d'autres intermédiaires a été de plus en plus reconnu, et des expériences techniques et réglementaires allant au-delà des modèles de suppression ont été menées. Le régime restrictif de notification et d'action de l'Allemagne, avec une loi d'application du droit aux réseaux sociaux[292], semble être la référence en matière de responsabilité des intermédiaires dans les États occidentaux. Plus récemment, on a assisté à un changement de paradigme dans l'approche des propositions de réforme des lois, mettant en évidence des solutions créatives et prometteuses[293]. Bien qu'un examen détaillé dépasse le cadre du présent document, je vais esquisser quatre thèmes principaux et signaler certains des aspects les plus problématiques.

Premièrement, l'évolution la plus importante en matière de réforme législative est le passage d'une orientation purement axée sur la réglementation du contenu à la gestion du risque systémique. La Commission canadienne sur l'expression démocratique a proposé un devoir d'agir de manière responsable[294]. Patrimoine canadien étudie actuellement une approche de gestion des risques concernant les préjudices en ligne, qui a été au cœur des ateliers du groupe consultatif d'experts. La façon de traiter la mésinformation et la désinformation a fait l'objet de débats[295]. À mon avis, si les médias sociaux et autres services en ligne sont chargés de gérer leurs risques systémiques, cela devrait naturellement comprendre la mésinformation et la désinformation. Toutefois, il ne devrait pas y avoir d'obligation de prendre des mesures à l'égard d'un contenu qui relèverait de la catégorie des informations légales mais gênantes. En fin de compte, la gestion des risques ne consiste pas à prendre des mesures à l'égard

---

[291] La réforme des lois est évidente dans de nombreux domaines de réglementation de la technologie. Je me concentre ici sur des préjudices en ligne précis et sur les cadres de responsabilité des intermédiaires, même si la réforme des lois sur la protection de la vie privée, la réglementation de l'IA et la concurrence sont également pertinentes pour aborder la manipulation de l'information.

[292] Network Enforcement Act (Netzwerkdurchsetzungsgesetz, NetzDG) (2017).

[293] Il ne fait aucun doute que certaines propositions peu judicieuses ont été faites, n'offrant pas l'équilibre délicat que le présent document tente de montrer comme étant nécessaire dans le domaine des préjudices en ligne.

[294] Commission canadienne sur l'expression démocratique, *Comment rendre les plateformes en ligne plus transparentes et plus responsables envers les utilisateurs/trices canadiens* (mai 2022) Forum des politiques publiques, en ligne : https://ppforum.ca/wp-content/uploads/2022/05/DemX-2-French-May-4-1.pdf. Le devoir d'agir de manière responsable ressemble au devoir de diligence proposé au Royaume-Uni, mais la Commission a cherché à séparer le concept de la jurisprudence en matière de droit de la négligence.

[295] Voir feuilles de travail, *supra* note 121.

d'éléments de contenu, mais il faudrait que cela soit clairement établi dans la législation, en particulier pour le contenu licite. La *Législation sur les services numériques* (DSA) de l'UE constitue un modèle pour une telle approche[296].

L'UE a adopté la DSA en 2022. Celle-ci impose des obligations de gestion des risques aux « très grandes plateformes »[297]. Ces plateformes doivent déterminer les risques systémiques liés à la diffusion de contenus illicites, à tout effet négatif sur les droits de la personne et à la manipulation intentionnelle de leurs services. En particulier, elles doivent prendre en compte les répercussions de leurs systèmes de modération et de recommandation et de leur système de sélection et d'affichage des publicités[298]. Les plateformes doivent ensuite atténuer les risques et mener des audits indépendants pour vérifier la conformité[299]. Parmi les autres dispositions clés figurent le contrôle des systèmes de recommandation par les utilisateurs, la transparence de la publicité et l'accès des chercheurs aux données pour le contrôle de la conformité[300]. La menace de mésinformation et de désinformation est abordée dans les considérants, mais elle n'est pas explicitement mentionnée dans le corps de la DSA. Par contre, peu après l'adoption de la DSA, un nouveau *Code de bonnes pratiques contre la désinformation* a été adopté[301]. L'un des défauts de la DSA est l'accent mis sur la gestion des risques sur les très grandes plateformes. Bien que des obligations spéciales pour ces plateformes puissent être appropriées, la gestion des risques est tout aussi importante pour les autres services en ligne, mais la capacité des petites et moyennes entreprises doit être prise en compte. Le Convoi illustre le fait que les médias sociaux traditionnels et alternatifs, ainsi que les messages dupliqués, ont été utilisés de manière intensive, et que la gestion des risques par quelques-unes des plateformes grand public n'aurait pas permis de lutter efficacement contre les préjudices en ligne.

L'accent mis sur la gestion des risques reflète le fondement de la diligence raisonnable des Principes directeurs[302]. Une variante pourrait être un modèle d'obligation de diligence, qui a été proposé dans le *Online Safety Bill* (OSB)[303] du Royaume-Uni. On ne sait pas ce qu'il adviendra de ce projet de loi sur la sécurité en ligne, qui est actuellement en suspens, mais il est un exemple de législation qui s'est perdue dans la complexité de la réglementation des préjudices en ligne[304]. Le contenu réglementé, les services et la nature des obligations varient tellement que, si la loi est adoptée, il sera difficile pour la plupart des services en ligne de s'y conformer. La complexité est un risque pour toute législation visant à remédier aux préjudices en ligne, si l'objectif est d'imposer des obligations sensibles aux droits de la personne qui font une

[296] DSA, *supra* note 189.
[297] *Ibid* art. 25-33.
[298] *Ibid* art. 26.
[299] *Ibid* art. 27-28.
[300] *Ibid* art. 29-31.
[301] *Supra* note 195.
[302] *Supra* note 223.
[303] *Online Safety Bill*, 2022-2023, HC Bill 121 (*tel qu'amendé au Public Bill Committee*), en ligne : https://publications.parliament.uk/pa/bills/cbill/58-03/0121/220121.pdf.
[304] Voir les graphiques de Graham Smith, « Mapping the Online Safety Bill » (27 mars 2022) *Cyberleagle*, en ligne : https://www.cyberleagle.com/2022/03/mapping-online-safety-bill.html.

distinction entre les différents types de contenus et de services en ligne. Très controversé, l'OSB a cherché à réglementer directement les propos licites mais offensants et a créé des infractions propres à la désinformation[305].

Deuxièmement, les rapports de transparence sont un élément essentiel de la surveillance de la conformité des intermédiaires. La DSA et l'OSB imposent tous deux des obligations en matière de rapports de transparence[306]. Comme je l'ai mentionné, il est difficile de bien faire des rapports de transparence et il s'agit d'un élément nouveau pour les services en ligne, en particulier pour les services de médias sociaux. Nous n'avons peut-être pas une idée précise de la nature de ces rapports, ni des paramètres de réussite, ni même de ce que nous voulons que les médias sociaux fassent en matière de transparence, mais il semble clair que la transparence est essentielle à l'avenir de la réglementation concernant les préjudices en ligne[307].

Troisièmement, la réforme des lois est systématiquement axée sur la création d'organismes de réglementation indépendants pour la sécurité en ligne. Ces organismes sont essentiels pour améliorer l'accès à la justice et pour faire progresser la nécessaire approche coréglementaire en matière de préjudices en ligne. L'Australie est la première juridiction à avoir créé un commissaire de la sécurité en ligne doté d'un mandat de recherche, d'éducation, d'enquête et d'application[308]. Le mandat a commencé par la protection des enfants contre l'intimidation et la diffusion non consensuelle d'images intimes et de matériel violent odieux, et s'est étendu à la protection des adultes et a été assorti d'un pouvoir réglementaire plus large[309]. Comme organisme de réglementation, le bureau du commissaire est conçu pour agir sur le contenu, mais il travaille en étroite collaboration avec l'industrie et a intégré la sécurité par la conception dans son travail[310]. Les questions de mésinformation et de désinformation ne relèvent pas de la compétence du commissaire. Le Royaume-Uni a choisi son organisme de réglementation des télécommunications et de la radiodiffusion, l'OFCOM, pour être l'organisme de réglementation de l'OSB[311]. La DSA prévoit que les États membres désignent un organisme de réglementation pour en assurer l'application[312]. Le rôle et la fonction d'un organisme de réglementation font l'objet de débats. Un modèle utile est celui des commissaires à la vie privée, qui ont à la fois un rôle d'éducation, de recherche et d'enquête[313]. Les pouvoirs de l'organisme de réglementation

---

[305] Caitlin Chin, « The United Kingdom's Online Safety Bill Exposes a Disinformation Divide » (11 août 2022) *Center for Strategic & International Studies,* en ligne : https://www.csis.org/analysis/united-kingdoms-online-safety-bill-exposes-disinformation-divide.

[306] DSA, *supra* note 189, art. 13, 23, 33; OSB, *supra* note 303, art. 64-65.

[307] Daphne Keller, « Some Humility About Transparency » (19 mars 2021) *The Center for Internet and Society*, en ligne : http://cyberlaw.stanford.edu/blog/2021/03/some-humility-about-transparency.

[308] Voir en ligne : https://www.esafety.gov.au/.

[309] *Online Safety Act 2021*, No. 76, 2021 (OSB). Voir la discussion sur le mandat législatif, en ligne : https://www.esafety.gov.au/about-us/who-we-are/our-legislative-functions.

[310] Voir « Safety by Design », en ligne : https://www.esafety.gov.au/industry/safety-by-design.

[311] Office of Communications. Dans une précédente recherche, j'ai constaté que l'OFCOM était mal adapté à la réglementation des droits numériques : Laidlaw, *supra* note 183, chapitre 6.

[312] Voir DSA, *supra* note 189, chapitre IV.

[313] Le commissaire à la sécurité électronique joue un rôle crucial dans l'éducation et le soutien du public et dans la collaboration avec l'industrie.

sont importants, notamment celui de prononcer des ordonnances et d'imposer des amendes[314]. La question de savoir si les utilisateurs doivent avoir accès à un mécanisme de recours en dehors des tribunaux ou des intermédiaires suscite encore plus de débats. Il semble y avoir un large consensus sur l'importance de mettre en place un ombudsman pour soutenir les utilisateurs, en particulier les groupes marginalisés et racialisés qui sont souvent ciblés en ligne. Cependant, la mesure dans laquelle il devrait y avoir un tribunal, une cour des services électroniques ou un conseil des médias sociaux est discutée[315].

Enfin, les droits de la personne sont essentiels à la protection contre les préjudices en ligne. La force de la DSA est de souligner les droits fondamentaux non seulement dans les considérants, mais aussi dans les façons précises dont les obligations sont formulées dans la substance même de la législation. C'est ce que l'on peut observer de la même manière dans l'OSB. Par exemple, le projet de loi établit qu'au moment de décider des mesures de sécurité, l'entité réglementée doit procéder à une évaluation des effets sur la liberté d'expression et la vie privée[316]. Le comité d'experts sur les préjudices en ligne a indiqué que le devoir d'agir de manière responsable devrait comporter deux obligations distinctes pour les services en ligne : la protection contre les préjudices en ligne et la protection des droits de la personne. Ainsi, par exemple, une décision concernant l'utilisation de l'automatisation pour gérer les risques de préjudices en ligne nécessiterait également une évaluation de l'incidence de cette approche sur les droits de la personne, tels que le droit à la vie privée et la liberté d'expression[317].

Conclusion

Les médias sociaux ont permis au Convoi de mobiliser et de réseauter. À bien des égards, c'est précisément ce que les médias sociaux ont été conçus pour faire, et ont fait, pour divers mouvements. Le problème est que les vecteurs d'attaque utilisés pour les fausses informations, les propos haineux, la violence, l'extrémisme, le harcèlement et d'autres formes d'abus sont les mêmes que les vecteurs utilisés pour publier des photos de famille, promouvoir votre entreprise, partager des vidéos d'animaux mignons et apprendre à réparer l'écran de votre iPhone. Pour lutter contre la mésinformation, la désinformation et la malinformation, il faut donc décortiquer le contenu des informations manipulatrices et approfondir la question des acteurs qui diffusent des fausses informations ainsi que les techniques qu'ils utilisent, les effets sur les utilisateurs qui consomment ces informations et la conception des espaces de médias sociaux exploités. Il devrait également y avoir une certaine convergence de vues sur ce dont

[314] Le renforcement du pouvoir des commissaires à la protection de la vie privée a fait l'objet d'une réforme législative et des leçons peuvent en être tirées pour la création d'un organisme de réglementation de la sécurité en ligne.
[315] J'ai préconisé la création d'un tribunal des services électroniques pour les litiges en matière de diffamation dans « Re-Imagining Resolution of Online Defamation Disputes » 56(1) OHLJ 162. Heidi Tworek recommande la création d'un conseil des médias sociaux : « Social Media Councils » (28 octobre 2019) *CIGI*, en ligne : https://www.cigionline.org/articles/social-media-councils/. La Commission canadienne sur l'expression démocratique recommande un tribunal des services électroniques pour la désinformation, *supra* note 294.
[316] OSB, *supra* note 309, art. 19.
[317] Voir *supra* note 121.

nous parlons lorsque nous évoquons la manipulation de l'information. Je suggère de simplifier l'analyse en la limitant à trois catégories : la désinformation (que l'on diffuse intentionnellement en sachant qu'elle est fausse), la mésinformation (que l'on diffuse intentionnellement en croyant qu'elle est vraie) et « tout le reste » des contenus préjudiciables, comme les contenus terroristes et extrémistes violents et les discours haineux, qui recoupent la désinformation et la mésinformation.

Le Convoi a mis en évidence les lacunes des lois et des politiques canadiennes en matière de réglementation des médias sociaux. Toutes les décisions relatives à la façon de traiter le contenu du Convoi publié sur les médias sociaux ont été prises par les entreprises de médias sociaux, en fonction de leurs directives d'utilisation et au moyen de diverses solutions techniques. Bien que chaque plateforme soit différente et que ces entreprises puissent concevoir des solutions créatives et sensibles aux droits de la personne, une discussion importante doit être engagée sur la façon d'encourager ces solutions, de créer des normes de l'industrie et de tenir les entreprises responsables. Dans plusieurs pays, dont le Canada, une réforme des lois est en cours pour s'attaquer aux préjudices en ligne. Les lecteurs sont encouragés à suivre l'évolution de la situation et à réclamer un régime fondé sur les droits de la personne au Canada.