Aller au-delà des renseignements personnels : ouvrir la voie à des lois qui protègent la population canadienne contre les préjudices liés au numérique

Christelle Tessono
Center for Information Technology Policy
Université de Princeton

Introduction

Fin février, à la suite d'une évaluation menée par la dirigeante principale de l'information du Canada et dont les résultats ont montré que TikTok présentait un « niveau inacceptable de risque envers la vie privée et la sécurité », le gouvernement canadien a interdit l'utilisation de l'application sur les appareils mobiles gouvernementaux. Les détracteurs et détractrices de cette interdiction, pour leur part, l'ont qualifiée de « diversion », dans la mesure où ces préoccupations ne sont ni nouvelles ni propres à TikTok. Une équipe de recherche du Citizen Lab a publié un rapport analysant la plateforme et a découvert que les types de données recueillis par TikTok pour suivre les utilisateur·rice·s et leur montrer des publicités ciblées sont en réalité semblables à ceux recueillis par les autres plateformes de réseaux sociaux populaires. Cela nous a poussé·e·s à nous interroger sur la question suivante : que savent de nous ces plateformes? Quelles sont les stratégies qu'elles emploient pour recueillir et analyser nos données? Et, plus important encore, quels sont les moyens à notre disposition au Canada pour nous protéger des préjudices liés au numérique? Dans l'article suivant, nous démontrerons que les cadres législatifs existants au Canada ne permettent pas de faire face aux préjudices individuels et collectifs que peuvent causer ces plateformes, car ils se centrent sur la protection des renseignements permettant d'identifier les personnes plutôt que de s'intéresser à toutes les formes de ce que Teresa Scassa appelle des « données d'origine humaine » dans son travail.

Que font les plateformes avec nos données?

Les plateformes de réseaux sociaux recueillent une grande variété de données allant de la géolocalisation de notre téléphone au contenu que nous partageons et aimons sur Instagram et TikTok, en passant par des informations sur notre santé recueillies à partir de technologies portables connectées qui suivent notre état de santé comme les montres Fitbit, sans oublier nos achats et nos habitudes de navigation, pour ne citer que quelques exemples. Comme le montrent les équipes de recherche de Linnet Taylor et al. et de Graef et van der Sloot, une fois ces informations recueillies, bien souvent, elles sont dépersonnalisées, puis soigneusement

sélectionnées afin de créer d'énormes bases de données contenant des informations qui reflètent le comportement et les activités des utilisateur·rice·s. Par la suite, des outils informatiques sont appliqués à ces bases de données agrégées, dont ils tirent des informations exploitables dans le but d'identifier des tendances, des préférences et des comportements chez les groupes de personnes associés à ces données. Comme l'expliquent Barocas et Nissenbaum, le processus assisté par ordinateur consistant à analyser d'immenses bases de données pour générer de nouvelles informations, communément appelé *exploration de données*, « brise notre intuition première selon laquelle c'est l'identité qui constitue la plus grande source de danger potentiel, car il substitue l'inférence au fait d'utiliser des informations permettant d'identifier les personnes et s'en sert comme d'un pont pour parvenir à découvrir des informations supplémentaires ». En d'autres termes, les informations obtenues à partir de ces ensembles de données peuvent fournir d'autres informations sur un individu ou un groupe, sans faire usage des éléments permettant d'identifier les personnes.

Qu'implique l'utilisation de ces technologies?

Quels sont les systèmes informatisés appliqués au niveau individuel? À quel moment le sont-ils? Pourquoi et comment? Les mêmes questions se posent concernant les inférences que font ces systèmes font à notre sujet. Mais, dans les deux cas, il est difficile d'y répondre. À l'heure actuelle, les équipes de recherche, les lanceur·euse·s d'alerte et les journalistes sont les trois acteurs principaux qui permettent de révéler au grand jour les problèmes que cela soulève. Par exemple, l'enquête du *Wall Street Journal* portant sur Meta Platforms inc. a révélé que la plateforme était au courant des répercussions négatives d'Instagram sur les adolescentes.

Au niveau collectif, les méthodes automatisés d'analyse de données ont une incidence sur la façon dont les groupes de personnes sont identifiés. Comme l'ont montré les chercheur·euse·s Lanah Kammourieh et al., ces systèmes sont capables d'identifier des groupes de quatre façons différentes. Tout d'abord, ils sont capables d'identifier des groupes et d'inférer des informations à leur sujet sans hypothèse prédéfinie. Deuxièmement, ils sont capables d'identifier les groupes au sein d'une population dont les membres n'avaient pas de rapport les un·e·s avec les autres avant l'analyse. Troisièmement, ils sont capables d'identifier des groupes grâce à de nouvelles approches analytiques et ainsi de créer des groupes basés sur des caractéristiques jusque-là

inconnues. Enfin, ces systèmes pourraient être capables d'identifier des groupes à l'insu des analystes, risquant ainsi de porter préjudice à certaines personnes.

Ce qui est particulièrement délicat avec ce genre d'inférences, c'est que ces outils d'analyse informatique peuvent discriminer les gens en les répartissant dans des groupes qui n'appartiennent pas à des catégories protégées par la loi (par exemple, la race, le genre ou le handicap) et ce, sans que les renseignements personnels de ces personnes ne soient exposés pour autant. Il est donc difficile de déterminer si une personne a fait l'objet d'un profilage ou de discrimination. Par conséquent, les stratégies législatives en matière de protection de la vie privée et des données qui se concentrent uniquement sur la protection des renseignements personnels permettant d'identifier les personnes « détournent l'attention des problèmes mettant en jeu des groupes de personnes anonymes ayant fait l'objet d'un profilage à partir d'énormes ensembles de données numériques, et peuvent même donner lieu à ce type de phénomène ».

Quels sont les types de préjudices collectifs qu'on peut observer en raison des inférences tirées de ces bases de données? Comme le souligne le Citizen Lab dans un rapport sur la collecte de données sur la mobilité, bien que les bases de données puissent contenir des données dépersonnalisées ou agrégées, le risque de réidentification demeure, car il est possible d'établir « des inférences ou des corrélations à partir des données ou en les superposant avec des renseignements personnels connus ». Une étude réalisée en 2009 par Latanya Sweeney, professeure à Harvard, l'a prouvé : dans ce cadre, plus de 40 % des personnes ayant participé de façon anonyme à une étude ADN ont pu être réidentifiées. Outre le risque d'identification, il existe aussi un risque de surveillance de groupes historiquement marginalisés, voire de ciblage politique, comme nous l'a appris le scandale de Cambridge Analytica. Plus important encore, les systèmes de prise de décision automatisés déployés pour analyser ces données ont tendance à mal identifier les gens, à mal les catégoriser et même à prédire des résultats de façon inexacte.

Comment le Canada s'en sort-il face à ces défis?

En matière de législation sur la protection des données et de la vie privée, le gouvernement canadien dispose de deux ensembles de lois. Premièrement, la *Loi sur la protection des renseignements personnels*, qui régit la collecte, l'utilisation, la divulgation, la conservation et la suppression des renseignements personnels par le gouvernement fédéral.

Deuxièmement, la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE), qui décrit comment le secteur privé doit traiter les renseignements personnels lors d'une activité commerciale. De plus, les provinces et les territoires disposent également de leurs propres lois régissant l'utilisation des renseignements personnels par les secteurs privé et public. Cela étant dit, le présent article se concentrera uniquement sur les lois fédérales. La *Loi sur la protection des renseignements personnels* et la LPRPDE se concentrent toutes deux principalement sur la protection des renseignements personnels. Ces lois définissent les renseignements personnels comme étant « les données qui concernent un "individu identifiable" », une définition qui n'aborde absolument pas la question de la protection des données ne permettant pas d'identifier les individus.

En juin 2022, le gouvernement canadien a déposé le projet de loi C-27 : *Loi sur la mise en œuvre de la Charte du numérique*, qui vise à édicter trois lois distinctes. Premièrement, la *Loi sur la protection de la vie privée des consommateurs* (LPVPC) cherche à moderniser la LPRPDE afin que le pays puisse s'adapter aux défis émergents dans le contexte des technologies numériques. Deuxièmement, la *Loi sur le Tribunal de la protection des renseignements personnels et des données* cherche à créer un tribunal pour imposer des sanctions lorsque des infractions à la LPVPC sont commises. Enfin, la *Loi sur l'intelligence artificielle et les données* (LIAD) vise à créer un cadre législatif pour « réglementer les échanges et le commerce internationaux et interprovinciaux en matière de systèmes d'intelligence artificielle par l'établissement d'exigences communes [...] pour la conception, le développement et l'utilisation de ces systèmes ». En ce qui concerne la protection des données, la LPVPC diffère de la LPRPDE, car elle introduit des dispositions sur la dépersonnalisation des données, leur suppression et la protection des enfants. Plus précisément, elle définit la dépersonnalisation des données comme le fait de « modifier des renseignements personnels afin de réduire le risque, sans pour autant l'éliminer, qu'un individu puisse être identifié directement ». De plus, la LPVPC cherche à fournir des garanties aux mineur·e·s en considérant leurs renseignements personnels comme sensibles. Or, malgré cela, ces lois laissent encore beaucoup à désirer.

Comment protéger davantage la population canadienne à l'avenir?

Pour lutter contre les préjudices liés au numérique qui émergent, le gouvernement canadien devrait moderniser sa législation sur la protection de la vie privée et étendre ses

mesures de protection aux renseignements ne permettant pas d'identifier un individu. Cela impliquerait d'apporter les modifications suivantes au projet de loi C-27 :

1. Mettre en place des mesures de protection pour toutes les données d'origine humaine, comme le propose Teresa Scassa. Cela inclurait les renseignements personnels, ainsi que les données dépersonnalisées et anonymisées.

2. Instaurer une interdiction de réidentification des données dépersonnalisées, comme le recommande l'étude parlementaire sur l'utilisation de données sur la mobilité pendant la pandémie de COVID-19.

3. Donner les moyens au Commissariat à la protection de la vie privée du Canada de faire appliquer les lois sur la protection de la vie privée dans les secteurs public et privé, d'enquêter sur les violations, de rédiger des réglementations et de contrôler les sociétés.

4. Définir dans la LPVPC ce qui constitue un « intérêt commercial légitime » et le « bien public » en lien avec la collecte, le stockage, l'utilisation, le transfert et la vente de données privées, comme le recommande l'étude parlementaire sur l'utilisation des données sur la mobilité pendant la pandémie de COVID-19.

En outre, étant donné que les technologies émergentes reposant sur des systèmes d'intelligence artificielle (IA) portent gravement atteinte à la vie privée et à de nombreux autres droits de la personne, la LIAD proposée doit être considérablement améliorée. Le gouvernement devrait chercher à établir un cadre réglementaire solide et indépendant en dotant le Commissariat à la protection de la vie privée du Canada de pouvoirs adéquats pour faire appliquer la loi et élaborer une réglementation propre à chaque secteur. De plus, nous avons besoin d'un cadre législatif qui prenne en compte les principaux risques que posent les systèmes algorithmiques pour les droits de la personne. Cela comprendrait, mais sans s'y limiter, le fait d'établir des limites et des lignes directrices claires sur la conception et le développement de systèmes algorithmiques qui :

1. ont des répercussions sur la santé et les résultats financiers des individus et des communautés;

2. sont utilisés pour accéder à des services sociaux ou à une aide humanitaire;

3. sont utilisés pour profiler les gens et influencer le comportement de ces derniers;
4. utilisent des informations corporelles biométriques ou liées à la santé pour identifier et catégoriser les personnes de manière unique.

Dans un contexte où le projet de loi C-27 est en cours de débat à la Chambre des communes, le gouvernement dispose d'une occasion unique de promulguer un cadre législatif qui non seulement protège la population canadienne contre les préjudices liés au numérique, mais garantit également que les technologies numériques seront développées de façon sécuritaire et équitable.

Christelle Tessono est chercheuse en politiques technologiques et chercheuse émergente au Center for Information Technology Policy de l'Université de Princeton. *Elle remercie les participant·e·s de l'atelier de février 2023 sur la gouvernance des plateformes pour les généreux retours qu'ils et elles lui ont faits sur une version antérieure de cet article.*