Commissariat
à la protection de
la vie privée du Canada

L'ère de l'analyse prédictive :

des tendances aux prédictions

Rapport préparé par le Groupe de recherche du Commissariat à la protection de la vie privée du Canada

Août 2012

Table des matières

Introduction 1

Le concept : l'« analyse prédictive » 2

L'analyse prédictive et son contexte 3

Le catalyseur : les plates-formes et les incitatifs font de nous tous des produits 4

Les ingrédients : les données que nous laissons derrière nous 5

L'impulsion : être au courant d'un événement avant qu'il ne se produise 6

Les applications : qui tente de prédire quoi? 6

La publicité ciblée 7

Les sciences sociales par le biais des médias sociaux 7

L'application de la loi et les services de renseignements 8

Le pistage de l'emplacement 9

La prévention de la fraude 10

Les conséquences sur la protection de la vie privée 10

L'analyse prédictive peut s'avérer « effrayante » 10

Des processus et des résultats opaques 11

Discrimination et atteinte à la réputation 12

La préemption pourrait nuire à l'application régulière de la loi 13

L'analyse predictive et la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE) 13

Connaissance et consentement 14

Ouverture et transparence 14

Responsabilité 15

L'analyse prédictive et la Loi sur la protection des renseignements personnels 16

Conclusion 17

Introduction

Dans le secteur public tout comme dans le secteur privé, on constate une fascination générale à l'égard des prédictions de comportements : qu'achèteront les consommateurs? De quelle façon utiliseront-ils la technologie? À quel moment une personne se comportera-t-elle mal, enfreindra-t-elle la loi ou commettra-t-elle une fraude? Danah Boyd a constaté cette tendance et affirme qu'« il ne s'agit plus de ce que l'on "fait" mais de ce que l'on "pourrait faire", et cela englobe également ce que les autres pourraient faire dans la mesure où ces gestes pourraient vous impliquer ou avoir une influence sur votre comportement »[1]. Ce type d'analyse ne constitue pas un phénomène entièrement nouveau, mais la forme que prend cette analyse passe des prédictions fondées sur l'expérience, l'intuition et la pensée critique à des prédictions fondées sur l'analyse technologique de données brutes — l'analyse prédictive.

Dans un article récent intitulé *How Companies Learn Your Secrets*[2] et dans son livre *The Power of Habit: Why we do what we do in life and business* [3], le reporter du New York Times Charles Duhigg décrit comment certaines entreprises utilisent l'« analyse prédictive » pour comprendre les habitudes personnelles et les habitudes d'achat des clients, de façon à pouvoir leur vendre des produits de façon plus efficace. Charles Duhigg a particulièrement mis en lumière les pratiques du grand magasin américain Target et a révélé que l'entreprise utilisait l'analyse prédictive dans le but d'identifier les femmes susceptibles d'être en début de grossesse, afin de pouvoir leur adresser des publicités ciblées avant ses concurrents. L'algorithme utilisé pour réaliser cette analyse a été surnommé « l'algorithme de prédiction de grossesse ».

La conception d'un algorithme de prédiction de grossesse illustre l'émergence d'une tendance au sein des entreprises, qui accordent désormais une grande importance à leur « service d'analyse des données » ou à leur « équipe de sciences des données » internes. Charles Duhigg révèle que Target dispose de 50 employés dont le travail consiste uniquement à dégager des tendances et des pratiques à partir des données recueillies par l'entreprise sur ses clients.[4] Pour mettre au point l'algorithme de prédiction de grossesse, l'équipe de scientifiques de données a testé des théories et analysé des habitudes à partir de données recueillies sur les clients, de registres de bébés et de données démographiques obtenues auprès de courtiers en données; elle a ainsi découvert qu'en réalisant certains schémas et en couplant des données, on pouvait déceler des habitudes d'achat prévisibles chez les femmes enceintes. Selon Charles Duhigg, la volonté de mettre au point un tel algorithme s'appuie sur une théorie selon laquelle les gens ont plus tendance à modifier leurs habitudes d'achat lorsqu'ils vivent des événements marquants tels qu'une grossesse; ainsi, le fait de cibler ces clients dans les campagnes de publicité pourrait renforcer l'habitude de magasiner chez Target.

Ce rapport de recherche a pour but de permettre une meilleure compréhension du concept d'analyse prédictive, soit le processus sous-jacent décrit dans l'article de Charles Duhigg. L'analyse prédictive est un processus analytique polyvalent pouvant être appliqué à des secteurs aussi variés que la vente au détail, dans le but de faire grimper les ventes, les forces de l'ordre, afin de prédire les

---

[1] Danah Boyd, « Networked Privacy », *Personal Democracy Forum*, New York, New York, le 6 juin 2011. [html]
[2] Charles Duhigg, « How Companies Learn your Secrets », *The New York Times*, février 2012.
[3] C. Duhigg, *The Power of Habit: Why we do what we do in life and business,* Doubleday Canada, Random House Canada Ltd., 2012.
[4] C. Duhigg, *The Power of Habit: Why we do what we do in life and business,* Doubleday Canada, Random House Canada Ltd., 2012, p. 190.

activités criminelles, et les programmes de santé, pour suivre les éclosions de maladies. Par conséquent, il ne s'agit pas d'un processus simple à définir ou à décrire, et ses répercussions sur la protection de la vie privée pourraient être nulles ou importantes, selon l'utilisation qui en est faite. En outre, il importe de souligner que le concept d'analyse prédictive est étroitement lié à des notions de forage des données déjà connues, mais qu'elle permet d'étendre les inférences au-delà de l'analyse des tendances rétrospectives et d'en arriver à un résultat plus prospectif et anticipatoire.

Une grande partie des recherches menées sur l'analyse prédictive et dont il est mention dans ce rapport se réfèrent au contexte américain et aux pratiques adoptées par les entreprises américaines. En raison de l'absence de données réelles, nous ne savons toujours pas de façon précise dans quelle mesure les entreprises canadiennes utilisent l'analyse prédictive et à quelles fins. Néanmoins, en nous penchant sur le contexte des États-Unis et sur les recherches qui y sont effectuées, nous pouvons tirer certaines conclusions sur les pratiques actuelles ou futures en la matière au Canada. En suivant les tendances qui émergent dans le domaine de l'analyse prédictive, nous pouvons mieux comprendre comment cet outil pourrait être utilisé par des entreprises, des organismes et des gouvernements, et quelles conséquences cette pratique pourrait avoir sur les personnes dans un contexte comme celui du Canada.

Ainsi, ce rapport constitue un point de départ en vue de l'étude de certains thèmes sous-jacents et cadres conceptuels permettant de comprendre l'analyse prédictive, et aborde quelques-unes des nombreuses applications possibles de cette technologie dans les secteurs privé et public.

Ce rapport vise à :

- étudier le concept d'analyse prédictive;
- fournir un aperçu du contexte dans lequel l'analyse prédictive est utilisée;
- cerner certaines applications dans les secteurs privé et public;
- décrire les conséquences plus larges de ce concept sur la protection de la vie privée pour les personnes et pour la société dans son ensemble;

examiner l'analyse prédictive en lien avec les principes de protection des renseignements personnels contenus dans la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE) et la *Loi sur la protection des renseignements personnels* (LPRP).

Le concept : l'« analyse prédictive »

La première question que plusieurs personnes se posent consiste à déterminer en quoi l'*analyse prédictive* diffère du *forage des données*, pratique utilisée par les entreprises et les gouvernements depuis déjà quelque temps. Le forage des données est défini comme étant le « *processus* permettant de découvrir des tendances intéressantes et d'acquérir une connaissance à partir de *grandes* quantités de données »[5]. Le forage des données et l'analyse prédictive sont tous deux des processus qui appliquent des analyses mathématiques et statistiques complexes dans le but d'en extraire des connaissances et des schémas. Bien que ces deux concepts soient liés, peut-être même synonymes en ce qui a trait aux processus qu'ils utilisent, l'analyse prédictive nous fournit de nouveaux indices en ce qui concerne l'évolution des pratiques de forage des données, qui deviennent de plus en plus intelligentes.

[5] J. Han, M. Kamber et J. Pei, *Data Mining Concepts and Techniques* (troisième édition), Elsevier Inc., p. 6 et 8.

L'analyse prédictive marque une progression du simple fait de dégager des habitudes à l'établissement de prédictions fondées sur ces habitudes. En 2006, le magazine informatique *Computerworld* définissait l'*analyse prédictive* comme étant « la branche du forage des données chargée d'établir des probabilités »[6]. Cette définition nous permet de constater que l'analyse prédictive est un concept unique tourné vers l'avenir et, lorsque les renseignements personnels sont utilisés en tant que données brutes, l'analyse prédictive est le processus utilisé pour tenter de prédire nos comportements et intentions futurs. Selon SAS, l'une des plus grandes entreprises d'analyse commerciale au monde, l'analyse prédictive consiste à « révéler des schémas, des *sentiments* et des relations qui n'avaient pas été identifiés antérieurement [souligné par l'auteur] »[7]. Alors que le forage des données décrit un processus exploratoire consistant à rechercher des schémas et des connaissances à partir de données, l'analyse prédictive tente d'approfondir les connaissances tirées des données pour en anticiper la signification et ainsi formuler des prédictions quant à l'avenir.

L'analyse prédictive est un processus analytique polyvalent qui permet à différentes entités d'identifier des schémas à partir de données; ceux-ci peuvent ensuite être utilisés pour prédire différents résultats n'ayant pas tous une incidence sur les personnes. Les entreprises ont maintenant plus facilement accès aux produits logiciels leur permettant d'intégrer l'analytique à leur modèle de gestion[8]. Elles peuvent ainsi les utiliser dans le but de réduire les risques qu'elles encourent, de rentabiliser les clients non rentables, de conserver les clients rentables, de réduire leurs dépenses, de détecter les cas de fraude, d'éviter la défaillance de leurs processus ou même d'analyser les effets de certains traitements de santé[9]. Les nouvelles applications comprennent les analyses en temps réel et l'analyse d'information non structurée telle que du texte[10]. Le principal aspect à souligner est que le type de décisions pouvant être prises en fonction des résultats d'analyses s'étend sans cesse à de nouveaux domaines. On constate une tendance à s'éloigner du forage des données, car ce processus présente des données qui ne sont que des regroupements ou des caractérisations de schémas tirés de données;[11] l'analyse prédictive, quant à elle, se démarque par sa capacité de tenter de prévoir, d'anticiper ou d'inférer.

L'analyse prédictive et son contexte

L'innovation technologique et la nature changeante de la consommation sur Internet ont joué un rôle important dans l'émergence de l'analyse prédictive en tant qu'outil pouvant être utilisé par les entreprises et les gouvernements. Les appareils que nous utilisons et la convergence de différentes technologies ont engendré de nouvelles voies de transmission et de nouvelles sources de données, ce qui a favorisé une prolifération de données à un rythme exponentiel. Au cours des dernières années, nous avons vu apparaître le terme « données massives » pour décrire la mine de renseignements recueillis à partir de nos activités quotidiennes, des articles que nous consommons et de nos interactions avec les autres personnes et les objets. Ces données massives sont à leur tour

---

[6] Jan Matlis, « QuickStudy: Predictive Analytics », *Computerworld*, 9 octobre 2006, http://www.computerworld.com/s/article/267042/Predictive_Analytics (consulté en ligne le 15 avril 2012).
[7] http://www.sas.com/technologies/analytics/datamining/
[8] Citons par exemple l'« outil de gestion des décisions analytiques » d'IBM ainsi que d'autres produits d'analyse prédictive fournis par SAS, Oracle, etc.
[9] SAS, *Drive Your Business with Predictive Analytics*, 2012.
[10] Paul M. Schwartz, *Data Protection Law and the Ethical Use of Analytics,* The Centre for Information Policy Leadership, Hunton & Williams LLP, 2010.
[11] J. Millar, « Core Privacy: A Problem for Predictive Data Mining », dans *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society*, Ian Kerr, Val Steeves, Carole Lucock (Éd.), Oxford University Press, 2009, p. 103 à 119.

devenues de plus en plus précieuses pour les entreprises, pour qui le fait de pouvoir formuler des prédictions constitue désormais l'un des ingrédients de la formule gagnante pour réussir. La section suivante traitera de certains éléments essentiels du contexte général d'utilisation des techniques d'analyse prédictive et décrira comment la technologie joue un rôle catalyseur en ce qui a trait à ce type d'analyse, comment les données massives constituent l'ingrédient essentiel, et comment le désir d'atteindre la réussite en prenant des décisions guidées par les données trace la voie de cette tendance.

Le catalyseur *: les plates-formes et les incitatifs font de nous tous des produits*

L'environnement en ligne est au cœur de la prolifération des données et, en fin de compte, de l'émergence des outils permettant de les analyser. Dans le cadre de la plupart de nos activités en ligne, que ce soit lorsque nous utilisons les médias sociaux ou d'autres applications, ou lorsque des entreprises exigent que l'on s'enregistre sur leur site et que l'on crée un profil d'acheteur, nous sommes invités ou incités à révéler des renseignements sur nous-mêmes. Dans certains cas, nous n'exerçons aucun contrôle sur ce qui est exigé, par exemple lorsque nous devons remplir les champs requis pour activer un service ou effectuer un achat en ligne. Dans d'autres cas, les gens divulguent volontairement des renseignements dans un contexte social ou en échange d'un avantage réel ou perçu. En raison de la gratuité des plateformes, des services et des contenus qu'on y trouve, Internet regorge d'incitatifs, de récompenses et d'avantages à participer à cet échange d'information, que ce soit en raison de l'aspect pratique des plateformes, du plaisir que nous procurent les échanges sociaux ou de la perspective intéressante de réaliser de bonnes affaires sur les produits qui nous intéressent. Les plateformes et les incitatifs sont tous conçus expressément de manière à encourager les gens à divulguer des renseignements sur eux-mêmes[12] et, comme l'a judicieusement souligné le professeur Zittrain en 2011, « si ce que vous obtenez en ligne est gratuit, alors vous n'êtes pas le client, mais bien le produit »[13].

Plus particulièrement à titre de consommateurs, nos habitudes de consommation et nos activités en ligne peuvent générer une quantité importante de renseignements. Les médias sociaux et d'autres types d'entreprises modifient les habitudes de consommation sur Internet et créent des liens entre les gens et les objets par le biais d'interactions sociales et de plusieurs types de technologies différentes : « l'utilisation et la convergence d'Internet, des téléphones cellulaires, des systèmes financiers électroniques, des systèmes d'identification biométrique, de l'IRF, des systèmes GPS, de l'intelligence ambiante et autres, contribuent à la production de données automatiques qui sont utilisées par des systèmes de forage des données et de suivi encore plus envahissants et plus puissants »[14].

Ces technologies peuvent permettre à des entreprises d'avoir un aperçu de ce qui se passe à l'heure actuelle ou de ce qui se produira sous peu, le nouvel objectif des produits numériques consistant désormais à fournir l'information « plus vite qu'en temps réel »[15]. Les entreprises se livrent une vive concurrence afin d'avoir accès à des renseignements uniques, ce qui encourage chacune d'entre elles à rechercher de nouveaux moyens de

---

[12] Alexander Furnas, « It's Not All About You: What Privacy Advocates Don't Get about Data Tracking on the Web », *The Atlantic*, 15 mars 2012, http://www.theatlantic.com/technology/archive/2012/03/its-not-all-about-you-what-privacy-advocates-dont-get-about-data-tracking-on-the-web/254533/ (consulté en ligne le 11 avril 2012).
[13] Citation tirée d'un article de Jonathan Zittrain sur http://news.harvard.edu/gazette/story/2011/06/hyper-public-spaces/.
[14] S. Gutwirth et M. Hildebrandt, « Some Caveats on Profiling », dans *Data Protection in a Profiled World*, de S. Gutwirth et al. (éd.), Springer Science & Business Media B.V., 2010.
[15] Andrew Keen, « Should we fear mind-reading future tech? », publié sur CNN Tech le 19 juin 2012.

recueillir des données[16]. Ainsi, les données transactionnelles des clients en temps réel et l'intégration de multiples plateformes et technologies en ligne ne feront que rendre plus exactes et significatives les descriptions de schémas de comportements et les prévisions de comportements futurs[17].

Les ingrédients *: les données que nous laissons derrière nous*

Notre société contemporaine évolue dans un environnement axé sur les données. Le concept de données massives n'est pas nouveau en soi, mais son échelle évolue rapidement, non seulement en ce qui a trait à la quantité de données recueillies, mais également à la façon dont ces données deviennent de plus en plus interdépendantes.

Le concept des données massives est étroitement relié à l'analyse prédictive, puisque les points de données sont les ingrédients qui alimentent l'application des algorithmes prédictifs. Dans notre société guidée par la technologie, presque tous nos gestes génèrent un flux de renseignements personnels[18]. Entre le contenu généré par l'utilisateur, volontairement divulgué par les personnes, et les renseignements personnels que les entreprises extraient de nos activités de consommation, il devient de plus en plus possible d'obtenir une image assez détaillée de la façon dont nous organisons et menons nos vies. « Nous sommes passés d'un contexte où de petits fragments d'information à notre sujet étaient stockés à plusieurs endroits différents, en ligne et hors ligne, à celui où il est possible d'obtenir une image pleinement détaillée de qui nous sommes, le tout étant enregistré et sauvegardé de façon numérique »[19]. Il devient de plus en plus difficile de maintenir la distinction entre la personne que nous sommes hors ligne et les activités que nous menons en ligne, en particulier lorsque nous utilisons des technologies mobiles.

L'augmentation constante des données massives s'accompagne d'une hausse de leur valeur. Certaines personnes prétendent que « les données représentent le nouveau pétrole »[20], c'est-à-dire qu'elles constituent un nouveau produit de base pouvant être raffiné (analysé) et ensuite exploité. L'explosion des données recueillies sur les consommateurs a créé une industrie à part entière qui se dit en mesure d'extraire une signification de ces données. Les entreprises recueillent ainsi de grandes quantités de données sur les consommateurs dans le but de bénéficier d'un avantage concurrentiel[21]. Plusieurs d'entre elles estiment que la seule façon de survivre au sein d'une économie mondiale est « d'exploiter le pouvoir de l'information et d'en tirer profit »[22]. Le déluge des données qui en résulte, celui-ci comprenant nos renseignements personnels utilisés comme données brutes, et la capacité croissante en ce qui a trait au stockage et aux technologies de l'information, contribuent à l'utilisation accrue de l'analytique[23]. La valeur réelle d'un tel processus vient du fait qu'il permet de mesurer nos comportements et inclinations avec une grande précision et de les utiliser pour formuler des prédictions sur des événements futurs.

---

[16] Paul.M Schwartz, *Data Protection Law and the Ethical Use of Analytics*, The Centre for Information Policy Leadership, Hunton & Williams LLP, 2010.
[17] Pridmore et Zwick, 2011, p. 271.
[18] Misty Harris, « The erosion of anonymity: Today's digital world forces us to share more of our personal information », *The Vancouver Sun*, 12 avril 2012. http://www.vancouversun.com/technology/erosion+anonymity/6396220/story.html.
[19] Terrence Craig et Mary E. Ludloff, *Privacy and Big Data,* O'Reilly Media Inc., 2011, p. 5.
[20] Argumentaire développé dans : http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/.
[21] Jason Pridmore et Detlev Zwick, « Editorial: Marketing and the Rise of Commercial Consumer Surveillance », *Surveillance & Society*, vol. 8, n° 3, 2011, p. 269-277, http://www.surveillance-and-society.org.
[22] SAS, *Drive your Business with Predictive Analytics*, 2012.
[23] Paul M. Schwartz, *Data Protection Law and the Ethical Use of Analytics,* The Centre for Information Policy Leadership, Hunton & Williams LLP, 2010.

L'impulsion *: être au courant d'un événement avant qu'il ne se produise*

Le désir de comprendre le sens des données est le facteur qui motive une grande partie de la cueillette, du stockage et de la diffusion d'information et c'est pour cette raison que l'on souhaite concevoir des techniques d'analyse de l'information toujours plus avancées[24]. On cherche désormais à connaître tout ce qu'il est possible de savoir au sujet d'une personne, de ses attributs et de ses actions passées, dans le but de comprendre ses prédispositions et de prédire ses actions futures[25]. Helen Nissenbaum a étudié les bouleversements technologies clés et décrit cette tendance comme étant une « confiance illimitée accordée aux processus d'information et à l'analyse de l'information et à leur capacité de résoudre des problèmes sociaux profonds et urgents », puis explique que « cette confiance engendre une recherche effrénée d'informations et d'outils d'analyse de plus en plus raffinés »[26]. Helen Nissenbaum nous rappelle que cette tendance ne se limite pas aux intérêts commerciaux, mais que l'attrait que présentent les méthodes analytiques et outils tels que l'analyse prédictive suscitera l'intérêt d'une multitude de décideurs dans différents domaines comme la finance, les assurances, les sociétés émettrices de cartes de crédit, la santé/les hôpitaux, la sécurité nationale et le maintien de l'ordre[27]. Selon le type d'application, l'utilisation des données massives jumelée au désir de pouvoir prédire des événements et aux capacités intelligentes des algorithmes prédictifs pourraient être des éléments susceptibles d'amplifier véritablement les conséquences de ces pratiques sur la protection de la vie privée.

Les applications : qui tente de prédire quoi?

L'intérêt grandissant à l'égard de l'utilisation de données pour guider la prise de décisions et l'accent mis sur celle-ci peuvent être influencés par de nombreux facteurs; dans certains cas, il s'agit d'anticiper des résultats susceptibles d'accroître les profits et dans d'autres, de gérer des risques ou de prévenir certaines répercussions négatives. Dans un chapitre de son livre à paraître, Ian Kerr décrit différentes utilisations possibles des technologies et de l'analyse prédictives[28]. L'une de ces utilisations porte sur ce qu'il appelle des « *prédictions préférentielles* », soit une tentative d'anticiper des préférences ou des inclinations personnelles, souvent dans le but d'adapter une offre de produits et services. Un autre type de prédictions que ces technologies permettent d'effectuer sont des « *prédictions préventives* », soit la tentative d'anticiper et de prévenir certaines actions susceptibles d'engendrer des risques sur le plan social[29]. Ces deux concepts peuvent aider à mieux comprendre certains résultats recherchés par les différentes utilisations de l'analyse prédictive. La section qui suit présentera des exemples issus des secteurs privé et public afin d'illustrer les résultats auxquels on souhaite parvenir grâce à l'utilisation de prédictions *préférentielles* et *préventives*.

---

[24] Helen Nissenbaum, *Privacy In Context: Technology, Policy, and the Integrity of Social Life*, Stanford University Press, 2009.
[25] Helen Nissenbaum, *Privacy In Context: Technology, Policy, and the Integrity of Social Life*, Stanford University Press, 2009, p. 44.
[26] *ibid*, p 42.
[27] *ibid*, p.45.
[28] Ian Kerr, « Prediction, Preemption, Presumption: The Path of Law After the Computational Turn »* (publication préliminaire), à paraître dans *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Mireille Hildebrandt & Ekaterina De Vries (Éd.).
[29] Ian Kerr, « Prediction, Preemption, Presumption: The Path of Law After the Computational Turn »* (publication préliminaire), à paraître dans *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Mireille Hildebrandt & Ekaterina De Vries (Éd.).

La publicité ciblée

Comme nous le démontre l'exemple des magasins Target, l'utilisation de l'analyse prédictive peut aider les entreprises à cibler leurs publicités de façon plus efficace dans le but d'accroître leurs profits. Les entreprises veulent être en mesure d'inférer les préférences de leurs clients, d'identifier des clients potentiels ou rentables et d'offrir les produits et services appropriés précisément au moment opportun[30]. Terry O'Reilly qualifie cette pratique d'*hyperciblage* et affirme que celle-ci permettra aux entreprises d'adresser directement aux consommateurs des publicités tout à fait personnalisées, issues d'une connaissance approfondie de leur vie personnelle, au moment précis où ils s'apprêtent à effectuer un achat[31]. Le contexte publicitaire démontre bien comment la possibilité de réaliser d'importants profits peut façonner la manière dont les entreprises valorisent et utilisent les renseignements personnels, de même que les facteurs qui les motivent à utiliser l'analyse prédictive pour mieux comprendre qui nous sommes, ce que nous voulons et quand nous le voulons, en temps réel. L'analyse prédictive permettra aux entreprises de livrer efficacement ce type de publicité ciblée, et elle illustre parfaitement comment le fait de pouvoir formuler des prédictions préférentielles pourrait s'avérer être une raison lucrative d'utiliser l'analyse prédictive dans le secteur privé.

Les sciences sociales par le biais des médias sociaux

L'analyse prédictive est devenue un outil permettant de tirer des perspectives utiles à partir de données collectives non structurées et de données générées par les utilisateurs; celles-ci révèlent certaines caractéristiques au sujet de différentes habitudes, notamment en ce qui concerne le comportement humain, la communication, l'analyse de sentiments et les modèles d'influence sociale. « Les recherches effectuées sur Google et les messages affichés sur Facebook et sur Twitter, par exemple, font qu'il est possible d'évaluer les comportements et les sentiments avec une grande précision et au moment où ils se produisent »[32].

Facebook possède l'ensemble de données le plus complet jamais recueilli sur le comportement social humain, et ses équipes de scientifiques de données recherchent de nouvelles façons d'exploiter la mine de données qu'elle détient afin de mieux comprendre la communication humaine et les comportements sociaux[33]. Puisque Facebook recueille des données sur des utilisateurs qui interagissent en temps réel, son équipe de scientifiques de données jouit d'une position unique lui permettant d'effectuer certaines expériences et analyses sur les habitudes et les motivations qui guident les gens dans leurs comportements sociaux, leurs préférences et leurs interactions. Par exemple, les données qu'elle recueille peuvent lui donner un aperçu des raisons pour lesquelles des idées ou des modes se répandent parmi les gens ou dans quelle mesure les actions futures d'une personne sont influencées par ses interactions avec ses amis. Facebook pourrait ainsi suivre une tendance sociale ou calculer le « bonheur national brut » d'un pays en temps réel en analysant des mots et des phrases clés qui révèlent des émotions positives ou négatives à l'égard de différentes choses[34]. Il pourrait s'avérer extrêmement lucratif pour Facebook de vendre les « perspectives » extraites de ses données à des entreprises qui souhaiteraient inciter les gens à partager des

---

[30] Paul M. Schwartz, *Data Protection Law and the Ethical Use of Analytics,* The Centre for Information Policy Leadership, Hunton & Williams LLP, 2010.
[31] CBC Radio, *Under the Influence with Terry O'Reilly*, émission traitant de l'hyperciblage, diffusée le 28 avril 2012, http://www.cbc.ca/undertheinfluence/season-1/2012/04/28/hyper-targeting-1/.
[32] Steve Lohr, « Big Data's Impact in the World », *New York Times*, 11 février 2012, http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=2&pagewanted=all.
[33] « What Facebook Knows : The Planetary Experiment », *Technology Review*, mot de l'éditeur, août 2012, p. 12.
[34] *Ibid*.

contenus ou à cliquer sur des publicités, ou encore à des économistes et à d'autres chercheurs qui étudient le comportement social humain.

Toutefois, le profit n'est pas le but recherché par toutes les entités intéressées par ce type d'application. Une nouvelle initiative des Nations Unies, baptisée « Global Pulse », cherche à effectuer des analyses de sentiments à partir de messages affichés sur les réseaux sociaux et de messages texte afin de pouvoir éventuellement prévoir les pertes d'emplois, les réductions de dépenses ou les éclosions de maladies dans une région donnée. Le but des Nations Unies est d'utiliser les premiers signes d'avertissement pour diriger ses programmes d'aide avant que les problèmes ne surviennent, ce qui lui permettrait par exemple d'éviter qu'une région sombre à nouveau dans la pauvreté[35].

L'application de la loi et les services de renseignements

Les organismes chargés de l'application de la loi et les services de renseignements utilisent depuis longtemps le forage des données et les techniques de profilage dans le but de prédire ou d'identifier des menaces ou des activités criminelles potentielles. Dans une société de plus en plus préoccupée par les risques et les menaces qui la guettent,[36] on constate une inquiétude constante liée au fait que n'importe qui peut s'avérer être « quelqu'un de malveillant » et un enthousiasme à l'égard des technologies prédictives qui devancent ou préviennent des comportements perçus comme pouvant présenter un risque sur le plan social.[37] Les organismes chargés de l'application de la loi sont de plus en plus intéressés par les produits issus de l'analyse prédictive et, aux États-Unis, les forces de l'ordre les utilisent déjà pour prédire quels seront les « points chauds » en se basant sur les heures et les endroits auxquels ont été précédemment commis d'autres crimes et en combinant ces données à celles des registres d'incidents et aux données historiques et sociologiques sur les comportements et tendances en matière de criminalité[38]. Ces technologies de détection utilisées avant qu'un crime ne soit commis continuent de faire l'objet de tests et de mises au point, et on prétend que certaines permettent déjà de prédire à quel moment les crimes seront commis et quels en seront les auteurs, avant même qu'ils n'aient lieu[39]. Grâce à son outil d'analyse, IBM affirme que l'analyse prédictive aidera les policiers à passer d'une approche de type « détection et réponse » à une approche consistant à « prédire et agir »[40]. D'autres programmes cherchent à analyser des comportements et des exemples d'attributs associés

[35] Steve Lohr, « Big Data's Impact in the World », *New York Times,* 11 février 2012, http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=2&pagewanted=all.

[36] Terme inventé par Ulrich Beck en1999.

[37] Ian Kerr, « Prediction, Preemption, Presumption: The Path of Law After the Computational Turn »* (publication préliminaire), à paraître dans *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Mireille Hildebrandt & Ekaterina De Vries (Éd.).

[38] Comme les programmes Blue CRUSH (Memphis) et Watson (Washington) d'IBM, http://www.cnn.com/2012/07/09/tech/innovation/police-tech/index.html.

[39] Ce logiciel, conçu par Richard Berk, professeur à l'Université de Pennsylvanie, est déjà utilisé à Baltimore et à Philadelphie afin de prédire quelles personnes parmi celles qui sont en période de probation ou en liberté conditionnelle seront les plus susceptibles de commettre un meurtre ou d'être victimes d'un meurtre : http://abcnews.go.com/Technology/software-predicts-criminal-behavior/story?id=11448231.

aux activités criminelles ou terroristes[41]. Les responsables de la sécurité publique et de la sécurité nationale cherchent avant tout à être en mesure de prédire quelles sont les personnes les plus susceptibles de commettre des actes terroristes ou de perpétrer des crimes. Un tel objectif suscite un intérêt sans cesse croissant à l'égard des prédictions de type préventives.

Le pistage de l'emplacement

Les téléphones intelligents et les tablettes mobiles font qu'il est possible et courant pour une personne de « géomarquer » son emplacement en temps réel. Les applications mobiles et les services en ligne non seulement encouragent et facilitent cette tendance, mais une étude menée récemment démontre que l'on peut prédire l'emplacement futur d'une personne en effectuant un suivi de l'utilisation de son téléphone cellulaire, et que ces prédictions sont tout à fait exactes et efficaces[42]. Dans le cadre de cette étude, un algorithme a permis de prédire les coordonnées GPS futures d'un utilisateur de téléphone cellulaire dans un rayon d'environ 1 000 mètres carrés. Lorsque la prévision tenait compte de renseignements supplémentaires recueillis auprès d'un seul de ses amis, l'emplacement futur de l'utilisateur pouvait être prédit dans un rayon de 20 mètres[43]. Même sans l'utilisation du géomarquage, l'étude a également démontré que l'on pouvait prédire l'emplacement de l'utilisateur avec un degré d'exactitude semblable en se fondant sur l'emplacement géographique des tours de téléphonie cellulaire[44]. La capacité de suivre et de prévoir les déplacements d'une personne pourrait s'avérer très attrayante pour des entreprises qui souhaiteraient livrer des publicités sur mesure en fonction de prédictions préférentielles ou pour des organismes chargés de l'application de la loi cherchant à prédire et à devancer certaines activités criminelles. Les entreprises pourraient prédire les déplacements d'une personne de façon à lui proposer une offre sur mesure au moment le plus opportun. Par exemple, une entreprise qui utiliserait cette technologie pourrait prédire à quel moment et à quel endroit vous prenez habituellement votre pause-café, de façon à vous offrir un bon de réduction au moment où vous franchissez la porte[45]. Les forces de l'ordre pourraient quant à elles utiliser une telle technologie pour suivre et prédire l'emplacement de certaines personnes soupçonnées d'avoir commis des crimes. Dans la mesure où les forces de l'ordre pourraient obtenir un mandat leur permettant d'utiliser les données de localisation GPS, les

---

[41] Quelques projets sont actuellement en cours au sein du département de la Sécurité intérieure, notamment 1) le projet « *Future Attribute Screening Technology* » (FAST), dont l'objectif est de concevoir des technologies pouvant filtrer les gens en fonction d'attributs comportementaux associés à la perpétration d'actes criminels violents ou autres; et 2) le projet « *Predictive Screening* » ayant pour but de déceler des comportements observables précédant un attentat-suicide à la bombe et de concevoir des algorithmes d'extraction permettant de les identifier et d'alerter le personnel lorsque des comportements liés aux attentats-suicides à la bombe sont observés.

[42] « On fait référence aux travaux de recherche de Musolesi sur ce qu'il appelle les modèles de mobilité, récemment publiés dans le cadre de ses travaux menés à l'université de Birmingham, au Royaume-Uni. Il a récemment reçu le prix Mobile Data Challenge de Nokia en prédisant les déplacements de 25 personnes travaillant dans une ville en Suisse et s'étant portées volontaires pour le projet. Pour ce faire, il a utilisé des données GPS, les numéros de téléphone des participants, les messages textes qu'ils avaient envoyés et la liste des appels qu'ils avaient effectués, et a créé un algorithme capable de prédire les déplacements de ces personnes dans un rayon de 20 mètres. Essentiellement, l'algorithme n'atteignait ce niveau de précision que lorsqu'il suivait également les déplacements et les données de chacun des amis de ces personnes. ». Tiré de http://www.forbes.com/sites/parmyolson/2012/08/06/algorithm-aims-to-predict-crime-by-tracking-mobile-phones/.

[43] http://www.forbes.com/sites/parmyolson/2012/08/06/algorithm-aims-to-predict-crime-by-tracking-mobile-phones/.

[44] http://www.forbes.com/sites/parmyolson/2012/08/06/algorithm-aims-to-predict-crime-by-tracking-mobile-phones/.

[45] http://www.forbes.com/sites/parmyolson/2012/08/06/algorithm-aims-to-predict-crime-by-tracking-mobile-phones/.

algorithmes leur permettraient de suivre les habitudes d'une personne en ce qui a trait à ses déplacements, et d'intervenir lorsque l'algorithme suggère qu'elle serait susceptible de se rendre à un endroit inhabituel[46].

La prévention de la fraude

Il existe également d'autres domaines où le gouvernement et le secteur privé pourraient utiliser l'analyse prédictive aux fins de prévention de la fraude. Les responsables des programmes gouvernementaux confrontés à des transactions ou à des demandes frauduleuses en vue de l'obtention de prestations gouvernementales, d'une indemnisation d'assurances ou d'un rapport de crédit pourraient utiliser l'analyse des schémas et des tendances visant à détecter et à prévenir les demandes fausses ou frauduleuses. Par exemple, la Direction générale des services d'intégrité de Service Canada utilise un logiciel statistique dans le cadre d'un projet pilote d'analyse prédictive des risques conçu pour détecter la fraude et les abus dans le cadre de l'assurance-emploi (AE)[47]. Le concept est que l'outil d'analyse prédictive des risques analyse plusieurs bases de données et améliore ainsi considérablement l'identification des demandeurs d'AE qui ont touché un trop-payé. Chaque dossier balisé pour examen par le système fait ensuite l'objet d'une enquête. Le programme témoigne d'un changement vers une détection automatisée de la fraude et une gestion générale du risque, facilitées par l'utilisation des outils d'analyse.

Les conséquences sur la protection de la vie privée

Il est trop simpliste de supposer que l'analyse (complète ou partielle) des données est entièrement problématique du point de vue de la protection de la vie privée[48]. L'analyse prédictive peut prendre de nombreuses formes, et les conséquences pour la protection de la vie privée varieront selon le contexte dans lequel l'analyse est utilisée, mais aussi selon la portée et la mise en œuvre de cette analyse. D'un côté, les données massives et une analyse prédictive intelligente pourraient contribuer à faire avancer la recherche, à stimuler l'innovation et à générer de nouvelles approches en vue d'une meilleure compréhension du monde et de la prise de décisions importantes et socialement bénéfiques dans des domaines comme la santé publique, le développement et la prévision économiques[49]. D'un autre côté, une analyse poussée entraîne une augmentation de la collecte, du partage et de couplage des données et peut également s'avérer incroyablement invasive, intrusive et discriminatoire, en plus de constituer un autre pilier d'une société de surveillance. La section qui suit examinera certaines des conséquences plus vastes sur la protection de la vie privée pour la personne et la société qui découlent de l'application de l'analyse prédictive.

L'analyse prédictive peut s'avérer « effrayante »

S'il ne s'agit pas d'une réaction universelle, l'analyse prédictive peut, dans certains contextes, susciter un sentiment de frayeur ou de malaise à l'idée d'être sous le regard d'un observateur omniprésent qui connaît

---

[46] *Ibid.*
[47] Selon un résumé sur l'EFVP publié par Ressources humaines et Développement des compétences Canada (RHDCC) sur http://www.hrsdc.gc.ca/fra/access_information/privacy/PhaseII.shtml
[48] Helen Nissenbaum, *Privacy In Context: Technology, Policy, and the Integrity of Social Life*, Stanford University Press, 2009, p. 50.
[49] Steve Lohr, « Big Data's Impact in the World », *New York Times*, 11 février 2012, http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=2&pagewanted=all

des choses sur nous et sur notre comportement[50]. Danah Boyd explique que la collecte des données en soi ne constitue pas une violation de la vie privée, mais précise que « le fait d'assembler ces données et de les utiliser pour "observer" constitue par contre une grave violation des normes en matière de protection de la vie privée »[51].

Les caractérisations erronées et les inexactitudes sont certainement un résultat problématique de l'analyse; toutefois, des prévisions exactes pourraient constituer une invasion encore plus importante de la vie privée dans certains contextes. J. Millar fait remarquer que les résultats de l'analyse prédictive peuvent révéler des choses qui font partie de ce qu'il appelle nos « renseignements privés de base ». Il explique que l'usage de l'analyse prédictive et de l'exploration des données peut violer les fondements de base de la protection de la vie privée lorsqu'elle révèle les désirs, les croyances et les intentions *inexprimés* d'un individu auxquels seul ce dernier aurait accès[52]. Il poursuit en expliquant que les conclusions tirées à notre sujet sur la base de l'utilisation d'algorithmes prédictifs, ou d'une analyse approfondie par une équipe de spécialistes formés ayant accès à de vastes quantités de données, pourraient nous amener à avoir le sentiment que nos renseignements personnels de base ont été violés parce qu'il s'agit souvent de conclusions qui dépassent ce qui pourrait autrement être facilement observé ou connu par d'autres à notre sujet[53].

Autrement dit, les hypothèses qui sont établies au sujet des personnes sur la base des activités qui sont facilement observables ne suscitent en général aucun malaise. Par contre, les hypothèses qui sont fondées sur une inspection plus approfondie et détaillée de nos activités, ou qui sont établies à l'aide d'une assistance technique peuvent être inattendues ou dépasser nos attentes raisonnables.

Des processus et des résultats opaques

L'analyse prédictive est habituellement un processus opaque. Même lorsqu'une entreprise indique, par exemple dans l'énoncé de la politique de confidentialité, qu'elle utilisera les renseignements personnels dans des analyses, nous savons que la plupart des gens ne lisent pas ou ne comprennent pas le langage juridique complexe et ardu utilisé habituellement pour rédiger ces textes. Un récent sondage commandé par le Commissariat à la protection de la vie privée du Canada a révélé que seulement 50 % des Canadiens qui ont répondu au sondage ont indiqué consulter « rarement » ou « jamais » les politiques de confidentialité, tandis qu'une majorité (62 %) de Canadiens estiment que les politiques de confidentialité des sites Web qu'ils visitent sont quelque peu vagues ou très vagues lorsqu'il s'agit de leur donner de l'information sur ce que l'entreprise fera avec leurs renseignements personnels[54]. En outre, même lorsque les gens lisent les petits caractères d'une politique de confidentialité ou des modalités d'utilisation, l'être humain en général n'est pas très doué pour évaluer l'ampleur des conséquences à long terme, et les risques augmentent en fonction de la quantité

---

[50] Alexander Furnas, « It's Not All About You: What Privacy Advocates Don't Get about Data Tracking on the Web », *The Atlantic*, 15 mars 2012.
[51] Danah Boyd, « Networked Privacy ».
[52] J. Millar, « Core Privacy: A Problem for Predictive Data Mining », dans *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society*, Ian Kerr, Val Steeves, Carole Lucock (Éd.), Oxford University Press, 2009, p. 103 à 119.
[53] J. Millar, « Core Privacy: A Problem for Predictive Data Mining », dans *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society*, Ian Kerr, Val Steeves, Carole Lucock (Éd.), Oxford University Press, 2009, p. 103 à 119.
[54] Commissariat à la protection de la vie privée du Canada, *Sondage auprès des Canadiens sur les enjeux liés à la protection de la vie privée*. Mené par Phoenix Strategic Perspectives Inc., janvier 2013.
http://www.priv.gc.ca/information/por-rop/2013/por_2013_01_f.pdf

de renseignements divulgués. Or, la conception des médias sociaux conditionne justement les personnes à divulguer de grandes quantités de renseignements[55].

La plupart du temps, l'extraction des renseignements personnels s'effectue sans que les consommateurs ne sachent exactement dans quelle mesure ils ont effectivement communiqué des données. Habituellement, les gens n'ont aucun choix quant aux renseignements personnels qu'ils doivent fournir, ne savent pas qui peut accéder à leurs données, ni comment celles-ci sont ensuite utilisées[56]. Si les personnes disposent d'un certain contrôle au début du processus à l'égard de ce qu'elles publient ou communiquent sur Internet et des transactions qu'elles effectuent, leur choix de participer ne se fonde pas généralement sur une compréhension globale de la manière dont leurs renseignements sont manipulés en arrière-plan et après la transaction.

Certains chercheurs qualifient ce phénomène de « flux asymétrique de l'information »[57]. Les entreprises, les organisations et les gouvernements tentent tous de découvrir des détails intimes sur les consommateurs et les citoyens, leur comportement, leurs habitudes, leurs intentions, etc., mais proportionnellement, les individus en savent très peu sur les organisations avec lesquelles ils interagissent[58]. En général, de nombreuses personnes, si ce n'est la majorité, ne comprennent pas les processus complexes sous-jacents à l'analyse prédictive ou aux techniques de forage des données. Elles sont donc perplexes et inconscientes des raisons pour lesquelles on « leur refuse un prêt, les cible pour le message d'une campagne politique particulière ou les inonde de publicités à un moment ou un endroit où elles ont révélé être particulièrement vulnérables aux mesures de marketing »[59]. Même si les entreprises donnaient des renseignements détaillés sur leurs activités, par exemple dans leur politique de confidentialité, il semble peu probable que les personnes ou les consommateurs auraient la motivation ou la capacité d'en apprendre autant sur ces entreprises que ces dernières en savent sur eux.

En outre, lorsque le gouvernement utilise des techniques d'analyse prédictive, le processus analytique et ses résultats sont bien souvent dissimulés au public pour des raisons de sécurité nationale ou de sécurité publique. Peu importe leur application particulière, de nombreux algorithmes et applications logicielles de prédiction sont opaques, parce qu'ils sont « protégés par les lois sur le droit d'auteur et les secrets commerciaux, de sorte que le public ne peut pas savoir qui les a élaborés, comment ils fonctionnent, ni si les hypothèses sur lesquelles ils sont fondés sont valables »[60].

Discrimination et atteinte à la réputation

Lorsque l'analyse prédictive est utilisée pour tirer des conclusions relatives aux futurs comportements et intentions, il existe un risque que les personnes soient profilées et classées dans des catégories, et qu'elles puissent même faire l'objet d'une discrimination fondée sur ces prédictions. Si l'on considère cette

---

[55] Cory Doctorow, « What Facebook Knows : The Curious Case of Internet Privacy. », *Technology Review*, 6 juin 2012.
[56] Vincent Manzerolle et Sandra Smeltzer, « Consumer Databases and the Commercial Mediation of Identity: A Medium Theory Analysis », 2011, p. 326 et 331.
[57] Vincent Manzerolle et Sandra Smeltzer, « Consumer Databases and the Commercial Mediation of Identity: A Medium Theory Analysis », *Surveillance & Society* 8(3), p. 323 à 337, http://www.surveillance-and-society.org.
[58] Citation tirée de Turow (2006), extrait de Manzerolle et Smeltzer, 2011, p. 327
[59] Mark Andrejevic, « Surveillance and Alienation in the Online Economy », 2011, p. 287.
[60] Ian Kerr, « Prediction, Preemption, Presumption: The Path of Law After the Computational Turn »* (publication préliminaire), à paraître dans *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Mireille Hildebrandt & Ekaterina De Vries (Éd.).

problématique dans le contexte de la publicité ciblée, on pourrait craindre que ces profils de consommateurs entraînent « l'exclusion à l'accès des biens et services » ou une « discrimination de prix non fondée sur les biens et les services, mais sur l'identité ou le profil des consommateurs »[61]. Dans ce scénario, la tarification dynamique se fonde sur des caractéristiques personnelles telles que la richesse (capacité de payer), l'urgence du besoin, et la vulnérabilité à certaines incitations ou tactiques de marketing »[62].

Le problème ne se limite pas à la collecte des renseignements, mais touche également les conclusions qui sont tirées de ces données. La réputation peut constituer un contrôle d'accès aux services et il est facile de bâtir une réputation fondée sur des données inexactes ou des renseignements sortis de leur contexte[63]. C'est un problème d'être exclu de la liste d'envoi d'un certain type de publicités, mais il est potentiellement beaucoup plus dommageable d'être exclu d'un programme gouvernemental ou de faire l'objet d'une surveillance disproportionnelle sur la base de renseignements erronés.

La préemption pourrait nuire à l'application régulière de la loi

Le but de l'analyse prédictive et préventive est de formuler des hypothèses concernant ce qui se produira avant que cela ne survienne. D'un point de vue éthique, l'adoption imprudente et excessive de technologies qui anticipent un acte répréhensible avant même que celui-ci se produise pourrait avoir des conséquences importantes sur nos modèles traditionnels de justice, d'application régulière de la loi et de libertés individuelles[64]. Le concept d'application régulière de la loi prévoit que les personnes ont la capacité d'observer ce qui se passe, de comprendre les décisions ou les mesures importantes les concernant, de participer à la prise de ces décisions ou de ces mesures, et d'y réagir[65]. La nature opaque et complexe de l'analyse pourrait nuire à l'atteinte du niveau de transparence et d'équité nécessaire pour l'application régulière de la loi.

L'analyse prédictive et la Loi sur la protection des renseignements personnels et les documents électroniques (LPRPDE)

Au Canada, les entreprises du secteur privé qui effectuent des analyses prédictives sur la base des renseignements personnels doivent veiller à ce que leurs pratiques respectent les principes en matière de protection de la vie privée énoncés dans la LPRPDE. Le consentement, le but et les limites de l'utilisation, l'ouverture et la transparence, de même que la responsabilité sont autant d'éléments à prendre en considération lorsque nous étudions la manière dont l'analyse prédictive est utilisée au pays dans divers contextes. La section qui suit cernera certains éléments déclencheurs du recours à l'analyse prédictive qui pourraient être source de préoccupations en vertu de la LPRPDE.

---

[61] Helen Nissenbaum, *Privacy In Context: Technology, Policy, and the Integrity of Social Life*, 2009.
[62] Helen Nissenbaum, *Privacy In Context: Technology, Policy, and the Integrity of Social Life*, 2009.
[63] Extrait de la présentation de Jennifer Barrigar sur les applications dans le cadre des conversations sur la protection de la vie privée au CPVP.
[64] Ian Kerr, « Prediction, Preemption, Presumption: The Path of Law After the Computational Turn »* (publication préliminaire), à paraître dans *Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*, Mireille Hildebrandt & Ekaterina De Vries (Éd.).
[65] *Ibid.*

Connaissance et consentement

Le consentement est un principe directeur essentiel en vertu de la LPRPDE et stipule que la personne doit avoir une connaissance élémentaire de la manière dont ses renseignements seront utilisés afin de pouvoir donner son consentement éclairé pour la collecte et l'utilisation de ses renseignements. Le problème est que le modèle traditionnel de consentement est un peu difficile à appliquer au scénario complexe et dynamique de l'analyse des données. Bien que les gens accordent une grande importance à leur droit à la vie privée, ils semblent largement disposés à échanger leurs renseignements personnels contre des biens en ligne ou des services gratuits. Les entreprises ont tendance à obtenir le consentement pour ces pratiques en incluant un « avis » ambigu aux utilisateurs, généralement bien caché dans le libellé d'une politique de confidentialité rédigée dans un langage juridique complexe. Certains avancent que cette forme de consentement n'a aucune signification à moins que les gens soient pleinement conscients des profils compilés et aient une compréhension totale de la manière dont leurs données sont manipulées (« pour savoir quelles données vous voulez cacher, vous devez savoir quel profil les entreprises dressent; pour savoir si vous voulez que les programmes et les profils s'adaptent automatiquement à votre comportement, vous devez savoir quand et comment les données pertinentes sont traitées… [et] ces initiatives ne doivent pas être simplement laissées à d'éventuels incitatifs du marché »[66]. C'est beaucoup demander aux gens de comprendre ce processus et de donner leur consentement éclairé en lisant le texte au langage alambiqué des politiques de confidentialité. En outre, le fait est que les fusions et la sous-traitance, les accords de partage de données entre les entreprises et les organisations, la portée de la collecte de données et les couplages effectués, ainsi que les capacités des algorithmes prédictifs que seuls les spécialistes des données sont en mesure de comprendre, sont autant d'éléments qui rendent cet environnement technique et commercial complexe et variable au fil du temps.

Ouverture et transparence

La complexité et la variabilité de l'environnement commercial en ligne posent également des problèmes sur le plan de la transparence. Garantir la transparence devrait signifier que les pratiques de traitement de l'information sont communiquées aux utilisateurs d'une manière qui soit pertinente et significative pour les choix qu'ils doivent faire[67]. Lorsqu'on prend en considération la dynamique du pouvoir et les asymétries de l'information, le but des organisations contre celui des personnes (socialiser ou utiliser une technologie novatrice), ainsi que la complexité et le caractère souvent secret des outils d'analyse prédictive, il n'est pas étonnant que la transparence soit un principe de protection de la vie privée difficile à respecter. Helen Nissenbaum juge qu'il est impossible d'expliquer l'écosystème de la publicité en ligne actuel d'une manière utile sans devoir fournir plein de détails. Habituellement, une organisation précise ses pratiques personnelles de traitement de l'information dans de longues modalités d'utilisation ou politiques de confidentialité rédigées dans un langage juridique et contractuel complexe, que la majorité des gens ne lisent pas ou ne comprennent pas. Le concept de « paradoxe de la transparence » utilisé par H. Nissenbaum[68] capture l'essence de cette problématique. Elle explique que si une politique de confidentialité détaille clairement chaque flux, condition, qualification, et exception, il est peu probable que les utilisateurs la comprennent, ou

---

[66] S. Gutwirth et M. Hildebrandt, « Some Caveats on Profiling », dans *Data Protection in a Profiled World*, de S. Gutwirth et al. (éd.), Springer Science & Business Media B.V., 2010, p. 38.

[67] Helen Nissenbaum, « A Contextual Approach to Privacy Online », 2011, http://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf.

[68] Alexis Madrigal, « The Philosopher Whose Fingerprints Are All Over the FTC's New Approach to Privacy », *The Atlantic*, 29 mars 2012 http://www.theatlantic.com/technology/archive/2012/03/the-philosopher-whose-fingerprints-are-all-over-the-ftcs-new-approach-to-privacy/254365/ (consulté en ligne le 23 mai 2012).

même la lisent[69]; toutefois, il n'est pas plus utile de résumer ces pratiques de traitement de l'information de façon plus simple, car ce résumé omet d'importants détails qui vont probablement changer la donne au chapitre de la protection de la vie privée[70]. Décrire les méthodes et les résultats attendus de l'analyse prédictive est déjà un défi en soi en raison des subtilités et des complexités de celle-ci, et des préoccupations devraient être soulevées lorsque les organisations cachent les détails de leurs activités dans le texte ambigu de leur politique de confidentialité[71]. Il faut être attentif à la « transparence simulée »[72], par laquelle l'approche en matière de transparence finit par être incompréhensible pour les utilisateurs et exagérément permissive pour l'entreprise.

Responsabilité

La responsabilité est un principe directeur clé pour les organisations qui utilisent l'analyse prédictive. Être une organisation responsable implique plus que de simplement avoir une politique de confidentialité ou de désigner un chef de la protection des renseignements personnels. Il faut pour cela disposer d'un modèle de gestion qui interprète et applique correctement tous les principes de protection de la vie privée.

L'éthique constitue le fondement de l'information équitable et des principes de protection de la vie privée, ainsi que d'autres notions comme les flux d'information pertinents, les attentes raisonnables en matière de protection de la vie privée et l'intégrité contextuelle. Il est fondamentalement question d'agir en prenant en considération les effets sur les autres et ainsi de toujours jouer un rôle clé dans l'évaluation des conséquences sur la vie privée des diverses applications de l'analyse prédictive. Paul Schwartz met l'accent sur l'importance de l'analyse éthique et d'une approche conceptuelle de la compréhension de ses conséquences. Il souligne que les composants essentiels d'une utilisation responsable et éthique de l'analyse sont les principes de protection de la vie privée, comme les concepts de responsabilité et de proportionnalité[73]. Il présente aussi des considérations éthiques d'ordre général utiles que les entreprises et les organisations doivent respecter avant d'avoir recours à l'analyse prédictive[74] :

- L'utilisation éthique de l'analyse doit être dictée par l'obligation d'une entreprise d'être un acteur socialement responsable.
- Le traitement des données, leur analyse et la prise de décision subséquente de l'organisation doivent respecter les normes culturelles et sociales relatives au comportement acceptable et à l'utilisation des renseignements « délicats ».
- Une entreprise doit être responsable et reconnaître que l'analyse pourrait avoir des conséquences négatives et positives sur les personnes.
- Une entreprise doit évaluer les conséquences de son utilisation de l'analyse sur la confiance que lui accordent un vaste éventail d'intervenants.

---

[69] Pour de plus amples renseignements à ce sujet, consulter l'étude de 2008 par Aleecia M. McDonals & Lorrie Faith Cranor intitulée « The Cost of Reading Privacy Policies », *Privacy Year in Review* 4:3, 2008, http://moritzlaw.osu.edu/students/groups/is/files/2012/02/Cranor_Formatted_Final.pdf.

[70] Helen Nissenbaum, « A Contextual Approach to Privacy Online », 2011, http://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf.

[71] Ce point a été soulevé de manière générale dans Pridmore et Zwick, 2011.

[72] Le concept « transparence simulée » est élaboré dans J.M. Balkin, «How Mass Media Simulate Political Transparency », Université de Yale. http://www.yale.edu/lawweb/jbalkin/articles/media01.htm.

[73] Paul M. Schwartz, *Data Protection Law and the Ethical Use of Analytics,* The Centre for Information Policy Leadership, Hunton & Williams LLP, 2010.

[74] Paul M. Schwartz, *Data Protection Law and the Ethical Use of Analytics,* The Centre for Information Policy Leadership, Hunton & Williams LLP, 2010.

Une évaluation des conséquences possibles est le point de départ pour les organisations qui envisagent d'avoir recours à l'analyse prédictive. Ce concept cadre bien avec le document du CPVP relatif à la responsabilité[75], qui stipule qu'« [u]ne organisation responsable peut prouver à ses clients, ses employés, ses actionnaires, les organismes de réglementation et ses compétiteurs qu'elle tient à la protection de la vie privée, pas seulement à des fins de conformité, mais aussi parce que la protection de la vie privée favorise une saine gestion des affaires ».

L'analyse prédictive et la Loi sur la protection des renseignements personnels

Au Canada, les ministères et organismes gouvernementaux qui effectuent une analyse prédictive en utilisant des renseignements personnels doivent veiller à respecter les dispositions de la *Loi sur la protection des renseignements personnels* et les politiques du Secrétariat du Conseil du Trésor (SCT), comme la Directive sur les pratiques relatives à la protection de la vie privée et la Directive sur les évaluations des facteurs relatifs à la vie privée. La prolifération des données a été un catalyseur et un élément clé dans l'émergence de l'analyse prédictive dans le secteur privé, et la tendance dans le secteur public est semblable en ce sens qu'elle est marquée par un partage accru des renseignements entre divers programmes, par une augmentation de l'externalisation ou des contrats avec des entreprises du secteur privé et par une augmentation de la collecte d'information en particulier en cas de recherche de « renseignement ».

La *Loi sur la protection des renseignements personnels* impose des limites concernant la collecte des renseignements personnels et prévoit notamment que les renseignements personnels ne seront recueillis que lorsqu'ils sont directement liés à un programme ou à une activité de l'institution. Elle prévoit également que les programmes gouvernementaux doivent, autant que possible, recueillir les renseignements personnels directement auprès de la personne concernée. Les ministères et organismes gouvernementaux sont tenus d'informer les personnes du but de la collecte de renseignements personnels, ainsi que de son autorisation, et de définir les nouvelles utilisations compatibles de ces renseignements. Il se peut que ce soit dans ces nouvelles « utilisations compatibles » que nous trouvions l'émergence de l'analyse prédictive comme outil visant à remplacer l'examen manuel et l'analyse de grandes quantités de données. Toutefois, en ce qui concerne les programmes de sécurité publique et de prévention de la fraude, la décision d'avoir recours à l'analyse prédictive doit être précédée d'un examen minutieux des conséquences sur la protection de la vie privée et d'une évaluation approfondie de la nécessité de recourir à l'analyse prédictive, afin de s'assurer que son utilisation est raisonnable et proportionnée au résultat désiré et de cerner si le programme est efficace tout en étant le moins envahissant possible[76].

[75] « Un programme de gestion de la protection de la vie privée : la clé de la responsabilité » http://www.priv.gc.ca/information/guide/2012/gl_acc_201204_f.asp.

[76] Document d'orientation du CPVP, « Une question de confiance : Intégrer le droit à la vie privée aux mesures de sécurité publique au 21e siècle », 2011.

Conclusion

Le cas de Target, à l'origine du présent rapport, a été un exemple révélateur de l'importance accordée à l'analyse prédictive par les entreprises, ainsi qu'au rôle joué par les « experts en matière de données ». L'algorithme de prédiction de grossesse était préoccupant, car il montrait à quel point l'analyse peut mener à la formulation de conclusions très personnelles sur les gens, tout en démontrant que le processus est généralement très opaque et peut engendrer le sentiment d'être manipulé ou profilé.

Il s'agit d'un domaine en rapide évolution, et l'ampleur du regroupement des données et de l'analyse augmente dans une mesure qui dépasse les préoccupations exprimées jusqu'à maintenant sur cette pratique. L'analyse prédictive est un outil qui peut être appliqué de diverses manières, et les considérations éthiques et les principes relatifs à l'équité dans le traitement des renseignements et à la protection de la vie privée peuvent aider à définir une approche contextuelle de l'évaluation des risques pour la protection de la vie privée associés à une application particulière de l'analyse prédictive, ainsi qu'à cerner les utilisations les plus préoccupantes. L'application des considérations éthiques doit toujours commencer par la compréhension que les décisions prises sur la base des résultats de l'analyse peuvent avoir des conséquences négatives sur les gens, que certains renseignements ne doivent pas être recueillis aux fins d'analyse et qu'il doit y avoir des limites raisonnables quant au type d'hypothèses qui peuvent être tirées concernant les futures intentions et les futurs comportements des personnes[77].

Charles Duhigg affirme que les experts de l'analyse prédictive soutiennent que « bientôt, les entreprises pourront connaître nos goûts et prédire nos habitudes mieux que nous ne les connaîtrons nous-mêmes »[78]. La vaste portée potentielle et en pleine expansion de l'analyse prédictive place cette problématique dans la ligne de mire des personnes préoccupées par la protection de la vie privée. Toutefois, il est difficile de saisir l'ampleur de l'utilisation de cette analyse au pays étant donné sa complexité et sa nature opaque, de cerner les résultats désirés, de déterminer la portée de son efficacité ou de délimiter à l'avance quelles formes particulières d'analyse seront source de préoccupations ou non dans le cadre de la protection de la vie privée. Ainsi, ce rapport de recherche ne constitue que la première étape de l'examen des tendances au chapitre de l'analyse prédictive et d'une meilleure compréhension des défis à venir.

[77] Paul M. Schwartz, 2010; Ian Kerr, « Prediction, Preemption, Presumption: The Path of Law After the Computational Turn »* (publication préliminaire) – version provisoire de l'article, l'auteur souhaite qu'on lui adresse une demande d'autorisation avant de le citer.
[78] C. Duhigg, *The Power of Habit*, 2012, p. 212.