Un manifeste des données utilisateurs, aujourd'hui ?

Le User Data Manifesto a été initié par Frank Karlitschek un militant du logiciel libre qui a fondé Nextcloud et Owncloud et participé à d'autres projets open source.
La source de cette traduction française figure sur ce dépôt Github, la dernière traduction que je reprends ici avec quelques modifications mineures date de 2015 et semble essentiellement due à Hugo Roy. Le dernier contributeur en date est Philippe Batailler.

> ***[EDIT]** Hugo Roy nous apporte cette précision :*
> *hello – la traduction est bien de moi, mais le texte en anglais aussi la version actuelle du manifeste est une œuvre collaborative avec Frank et @jancborchardt*

À la lecture on est frappé de la pertinence des propositions, cependant malgré quelques avancées du côté des directives de l'Union européenne, certains droits revendiqués ici sont encore à conquérir ! Et après 4 ans il faudrait peut-être ajouter d'autres éléments à ce manifeste : le droit d'échapper au pistage publicitaire, le droit d'anonymiser vraiment sa navigation, le droit de ne pas fournir ses données biométriques etc.

Mais c'est plutôt à vous de dire ce qui manque ou est à modifier dans ce manifeste pour qu'il soit solidement inscrit dans les lois et les usages. Comme toujours, le commentaires sont ouverts et modérés.

Manifeste des données utilisateur

Ce manifeste a pour but de définir les droits fondamentaux des utilisateurs sur leurs données à l'ère d'Internet. Chacun

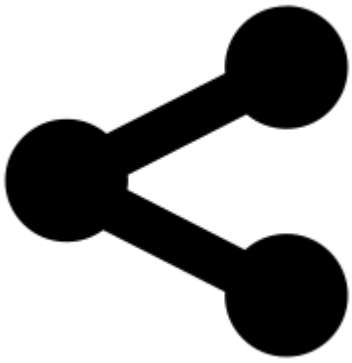
devrait être libre sans avoir à faire allégeance aux fournisseurs de service.

Par **données utilisateur**, on entend les données envoyées par un utilisateur ou une utilisatrice pour son propre usage.

Par exemple, les données utilisateur comprennent :

- les fichiers qu'un utilisateur ou qu'une utilisatrice synchronise entre plusieurs appareils ou qu'il ou elle partage avec un·e proche
- une bibliothèque d'albums photos, de livres ou d'autres fichiers qu'un utilisateur envoie depuis son appareil afin de pouvoir lire, voir, et modifier tout cela en ligne
- les données générées par un appareil de l'utilisateur (comme un thermostat ou une montre connectée) et envoyées vers un serveur
- les requêtes d'un utilisateur à un moteur de recherche, si de telles requêtes sont enregistrées comme telles

Ainsi, les utilisateurs devraient pouvoir…



1. Maîtriser leur accès à leurs données

Les données explicitement et volontairement envoyées par une utilisatrice devraient être sous la pleine maîtrise de l'utilisatrice. Les utilisateurs devraient être capables de décider à qui accorder un accès direct à leurs données et avec quelles permissions et licences cet accès devrait être accordé.

Lorsque les utilisateurs maîtrisent l'accès aux données qu'ils envoient, les données censées restées privées ou partagées à un cercle restreint ne devraient pas être rendues accessibles au fournisseur du service, ni divulguées aux États.

Cela implique que le droit d'utiliser le chiffrement ne devrait jamais être bafoué.

Cela implique également que lorsque des utilisateurs n'ont pas la pleine maîtrise sur l'envoi de leurs données (par exemple s'ils n'utilisent pas le chiffrement avant l'envoi) un fournisseur de service **ne** doit **pas** :

- forcer les utilisateurs à divulguer des données privées (ce qui inclut la correspondance privée) pour eux, ni
- imposer des conditions de licence (ex. : de droit d'auteur ou d'exploitation des données personnelles) qui vont au-delà de ce qui est nécessaire pour l'objectif du service.

Lorsque les utilisateurs rendent des données accessibles à

d’autres, qu’il s’agisse d’un groupe de gens restreint ou d’un groupe plus large, ils devraient pouvoir décider sous quelles permissions l’accès à leurs données est autorisé. Cependant, ce droit n’est pas absolu et ne devrait pas empiéter sur le droit des tierces personnes à utiliser et exploiter ces données une fois qu’elles leur ont été rendues accessibles. Qui plus est, cela ne signifie pas que les utilisateurs devraient avoir le droit d’imposer des restrictions injustes à d’autres personnes. Dans tous les cas, les systèmes techniques ne doivent pas être conçus pour faire appliquer de telles restrictions (par exemple avec des DRM).

Les données reçues, générées ou collectées à partir de l’activité des utilisateurs dans l’utilisation du service (ex. : les métadonnées ou les données du graphe social) devraient leur être rendues accessibles et être également sous leur maîtrise. Si cette maîtrise n’est pas possible, alors ce type de données devrait être anonyme ou bien ne pas être stockée pour une période plus longue que nécessaire.

Certains services permettent aux utilisateurs de soumettre des données avec l’intention de les rendre publiquement accessibles à toutes et à tous. Y compris dans ces cas de figure, quelques données utilisateur restent privées (ex. : les métadonnées ou les données du graphe social). L’utilisatrice et l’utilisateur devraient pouvoir contrôler aussi ces données.

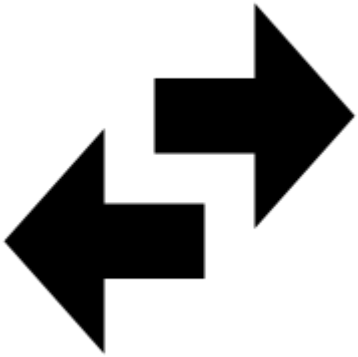
2. Savoir comment les données sont stockées

Quand les données sont envoyées à un fournisseur de service particulier, les utilisateurs et utilisatrices devraient être informé·e·s du lieu de stockage des données du fournisseur de service, de la durée, de la juridiction dans laquelle le fournisseur de service particulier opère et des lois qui s'y appliquent.

Lorsque les utilisateurs utilisent des services centralisés pour envoyer leurs données à un fournisseur de stockage particulier plutôt que de reposer sur des systèmes pair à pair, il est important de savoir où les fournisseurs pourraient stocker ces données car ils pourraient être obligés par les États à divulguer ces données qu'ils ont en leur possession.

Ce point est sans objet si les utilisateurs sont capables de stocker leurs propres données sur leurs appareils (ex. : des serveurs) dans leur environnement personnel et sous leur contrôle direct ou bien s'ils font confiance à des systèmes sans contrôle centralisé (ex. : le pair à pair).

Les utilisateurs ne devraient pas reposer sur des services centralisés. Les systèmes pair à pair et les applications *unhosted* sont un moyen d'y arriver. À long terme, tous les utilisateurs devraient être capables d'avoir leur propre serveur avec des logiciels libres.



3. Être libres de choisir une plateforme

Les utilisatrices devraient toujours être en mesure d'extraire leurs données d'un service à tout moment sans subir l'enfermement propriétaire.

Les utilisateurs ne devraient pas être bloqués par une solution technique particulière. C'est pourquoi ils devraient toujours être capables de quitter une plateforme et de s'installer ailleurs.

Les formats ouverts sont nécessaires pour garantir cela. Évidemment, sans le code source des programmes utilisés pour les données utilisateurs, cela n'est pas pratique. C'est pourquoi des programmes devraient être distribués sous une licence libre.

Si les utilisateurs ont ces droits, ils ont la maîtrise de leurs données plutôt que d'être sous la coupe des fournisseurs de service.

De nombreux services qui gèrent les données utilisateur à ce jour sont gratuits, mais cela ne signifie pas qu'ils soient libres. Plutôt que de payer avec de l'argent, les utilisateurs font allégeance aux fournisseurs de services pour que ceux-ci puissent exploiter les données utilisateurs (par ex. en les vendant, en offrant des licences ou en construisant des profils pour les annonceurs publicitaires).

Abandonner ainsi la maîtrise de sa vie privée et d'autres

droits semble être un acte trivial pour de nombreuses personnes, un faible prix à payer en échange du confort que ces services Internet apportent.

Les fournisseurs de service ont ainsi été obligés de transformer leurs précieux services Internet en systèmes massifs et centralisés de surveillance. Il est crucial que chacun réalise et comprenne cela, puisqu'il s'agit d'une menace importante pour les libertés de l'humanité et le respect de la vie privée de chacun.

Enfin, pour assurer que les données utilisateurs soient sous la maîtrise des utilisateurs, les meilleurs conceptions techniques incluent les systèmes distribués ou pair-à-pair, ainsi que les applications *unhosted*. Juridiquement, cela signifie que les conditions générales d'utilisation devraient respecter les droits des utilisateurs et leur donner la possibilité d'exercer leurs droits aux données définis dans ce manifeste.

Illustration réalisée avec https://framalab.org/gknd-creator/

---

21 degrés de liberté – 15

La communication, sur Internet et ses médias sociaux comme dans la vie, doit rester dans le cadre des lois. Mais est-il normal qu'un puissant média comme Facebook s'arroge le droit de décider, le plus souvent sans explication ni recours, de quels sujets on parle de façon privée ?

Voici déjà le 15e article de la série écrite par Rick Falkvinge. Le fondateur du Parti Pirate suédois prend ici l'exemple de la censure exercée par Facebook, qui en quelque

sorte se substitue aux lois en imposant la sienne et bride la liberté d'expression.

Le fil directeur de la série de ces 21 articles, comme on peut le voir clairement dans les épisodes précédents que nous vous avons déjà livrés, c'est la ***perte de certaines libertés*** *dont nous disposions encore assez récemment, avant que le passage au tout-numérique ne nous en prive.*

Les médias numériques interdisent à nos enfants d'aborder certains sujets.

Source : Rick Falkvinge sur privateinternetaccess.com

Traduction Framalang : wyatt, draenog, goofy, et un·e anonyme

Dans le pire des cas il pouvait être interdit à nos parents de se rencontrer. Mais aujourd'hui, on empêche nos enfants de parler de certains sujets, une fois la conversation en cours. Cette évolution est effrayante.

Lorsqu'un lien vers The Pirate Bay est publié sur Facebook par nos enfants, une petite fenêtre fait son apparition à l'écran avec pour message « Le lien que vous venez de publier n'est pas approprié. Veuillez ne plus publier de tels liens ».

Oui, même dans les conversations privées. Particulièrement dans les conversations privées.

Cela peut paraître anodin, c'est véritablement inouï. Les discussions de nos enfants ne sont pas restreintes en soi, mais elles sont en revanche contrôlées si elles abordent les sujets sensibles dont le régime ne souhaite pas qu'on discute et on les empêche d'en discuter. C'est bien pire que de simplement interdire à certaines personnes de se rencontrer.

L'équivalent analogique de cette pratique serait une conversation téléphonique classique de nos parents dans laquelle une troisième voix menaçante interviendrait, parlant lentement sur un ton assez doux pour être perçu comme une menace : « Vous avez fait mention d'un sujet interdit. Veuillez ne plus discuter de sujets interdits à l'avenir. »

Nos parents auraient été effrayés si cela s'était produit – et à juste raison !

Mais dans le monde numérique de nos enfants, au lieu d'être conspuée, cette pratique est acclamée par les mêmes personnes qui la réprouveraient si elles venaient à en être les victimes.

Dans le cas de notre exemple bien sûr, n'importe lequel des liens vers The Pirate Bay est considéré comme sujet interdit, selon le postulat – le postulat ! – qu'ils mènent à la production de copies qui seraient décrétées en violation du droit d'auteur par un tribunal.

Copie d'écran de Marc Rees en illustration d'un article de NextInpact : [Facebook censure (toujours) les messages privés qui l'ennuient](#).

La première fois que j'ai vu la fenêtre Facebook m'enjoignant à ne pas discuter de sujets interdits, je tentais de partager via The Pirate Bay du contenu à caractère politique que j'avais créé. Cette façon de faire s'est avérée très efficace pour partager des gros fichiers, c'est exactement la raison pour laquelle le site est très utilisé (qui aurait pensé à ça, hein ?), notamment par des personnes qui comme moi veulent partager de vastes séries de documents politiques.

Il existe des canaux de communications privés, mais très peu de personnes les utilisent, et les politiciens (oui, nos parents analogiques inclus) s'en réjouissent, à cause du « terrorisme » et autres épouvantails.

La vie privée demeure de votre responsabilité.

Sécurité de nos données : sur qui compter ?

Un des meilleurs experts indépendants en sécurité informatique résume ici parfaitement ce qui selon lui constitue un véritable problème : notre dépendance aux *commodités* que nous offrent les entreprises hégémoniques de l'Internet. Nous bradons bien facilement nos données personnelles en échange d'un confort d'utilisation dont on ne peut nier sans hypocrisie qu'il nous rend la vie quotidienne plus facile.

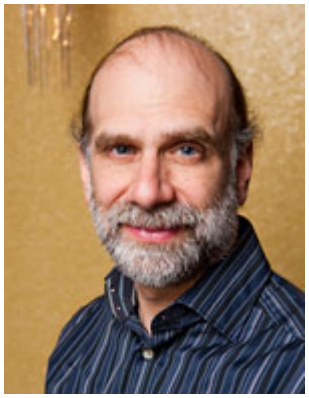
Dès lors que nous ne pouvons renoncer aux facilités que nous procurent Google, Facebook et tous les autres, pouvons-nous espérer que les technologies de sécurité nous épargnent un pillage de nos données personnelles ? Rien n'est moins sûr, selon Bruce Schneier, qui en appelle plutôt à la loi qu'à la technique.

Goofy.

Traduction Framalang : Simon, Docendo, KoS, goofy, audionuma, seb, panini, lamessen, Obny, r0u

Article original : Everyone Wants You To Have Security, But Not from Them

Ils veulent tous notre sécurité, mais pas grâce à d'autres



par Bruce Schneier

En décembre dernier, le PDG de Google Eric Schmidt a été interviewé lors d'une conférence sur la surveillance de l'Institut CATO. Voici une des choses qu'il a dites, après avoir parlé de certaines des mesures de sécurité que son entreprise a mises en place après les révélations de Snowden : « si vous avez des informations importantes, l'endroit le plus sûr pour les garder, c'est chez Google. Et je peux vous assurer que l'endroit le plus sûr pour ne pas les conserver en sécurité, c'est partout ailleurs ».

J'ai été surpris, parce que Google collecte toutes vos informations pour vous présenter la publicité la plus ciblée possible. La surveillance est le modèle économique d'Internet, et Google est l'une des entreprises les plus performantes en la matière. Prétendre que Google protège vos données mieux que quiconque, c'est méconnaître profondément ce pourquoi Google conserve vos données gratuitement.

Je m'en suis souvenu la semaine dernière lorsque je participais à l'émission de Glenn Back avec le pionnier de la cryptographie Whitfield Diffie. Diffie a déclaré :

> *Vous ne pouvez pas avoir de vie privée sans sécurité, et je pense que nous avons des défaillances flagrantes en sécurité informatique, pour des problèmes sur lesquels nous travaillons depuis 40 ans. Vous ne devriez pas vivre avec la peur d'ouvrir une pièce jointe dans un message. Elle devrait être confinée ; votre ordinateur devrait être en mesure de la*

> *traiter. Et si nous avons continué depuis des dizaines d'années sans résoudre ces problèmes, c'est en partie parce que c'est très difficile, mais aussi parce que beaucoup de gens veulent que vous soyez protégés contre tout le monde… sauf eux-mêmes. Et cela inclut tous les principaux fabricants d'ordinateurs qui, grosso modo, veulent contrôler votre ordinateur pour vous. Le problème, c'est que je ne suis pas sûr qu'il existe une alternative viable.*

Cela résume parfaitement Google. Eric Schmidt veut que vos données soient sécurisées. Il veut que Google soit le lieu le plus sûr pour vos données tant que vous ne vous préoccupez pas du fait que Google accède à vos données. Facebook veut la même chose : protéger vos données de tout le monde sauf de Facebook. Les fabricants de matériels ne sont pas différents. La semaine dernière, on a appris que Lenovo avait vendu des ordinateurs avec un logiciel publicitaire préinstallé, appelé Superfish, qui casse la sécurité des utilisateurs pour les espionner à des fins publicitaires.

C'est la même chose pour les gouvernements. Le FBI veut que les gens utilisent un chiffrement fort, mais veut des portes dérobées pour pouvoir accéder à vos données. Le Premier ministre britannique David Cameron veut que vous ayez une sécurité efficace, tant qu'elle n'est pas trop forte pour vous protéger de son gouvernement. Et bien sûr, la NSA dépense beaucoup d'argent pour s'assurer qu'il n'y a pas de sécurité qu'elle ne puisse casser.

Les grandes entreprises veulent avoir accès à vos données pour leurs profits ; les gouvernements les veulent pour des raisons de sécurité, que ces raisons soient bonnes ou moins bonnes. Mais Diffie a soulevé un point encore plus important : nous laissons beaucoup d'entreprises accéder à nos informations parce que cela nous facilite la vie.

J'ai abordé ce point dans mon dernier livre, *Data and Goliath*

:

> *Le confort est l'autre raison pour laquelle nous cédons volontairement des données hautement personnelles à des intérêts privés, en acceptant de devenir l'objet de leur surveillance. Comme je ne cesse de le dire, les services basés sur la surveillance sont utiles et précieux. Nous aimons pouvoir accéder à notre carnet d'adresses, notre agenda, nos photos, nos documents et tout le reste sur n'importe quel appareil que nous avons à portée de la main. Nous aimons des services comme Siri et Google Now, qui fonctionnent d'autant mieux quand ils savent des tonnes de choses sur nous. Les applications de réseaux sociaux facilitent les sorties entre amis. Les applications mobiles comme Google Maps, Yelp, Weather et Uber marchent bien mieux et plus rapidement lorsqu'elles connaissent notre localisation. Permettre à des applications comme Pocket ou Instapaper de connaître nos lectures semble un prix modique à payer pour obtenir tout ce que l'on veut lire à l'endroit qui nous convient. Nous aimons même quand la publicité cible précisément ce qui nous intéresse. Les bénéfices de la surveillance dans ces applications, et d'autres, sont réels et non négligeables.*

Comme Diffie, je doute qu'il existe une alternative viable. Si Internet est un exemple de marché de masse à l'échelle de la planète, c'est parce que toute l'infrastructure technique en

est invisible. Quelqu'un d'autre s'en occupe pour vous. On veut une sécurité forte, mais on veut aussi que les entreprises aient accès à nos ordinateurs, appareils intelligents et données. On veut que quelqu'un d'autre gère nos ordinateurs et smartphones, organise nos courriels et photos, et nous aide à déplacer nos données entre nos divers appareils.
Tous ces « quelqu'un d'autre » vont nécessairement avoir la capacité de violer notre vie privée, soit en jetant carrément un coup d'œil à nos données soit en affaiblissant leur sécurité de façon à ce qu'elles soient accessibles aux agences nationales de renseignements, aux cybercriminels, voire les deux. La semaine dernière, on apprenait que la NSA s'était introduite dans l'infrastructure de la société néerlandaise Gemalto pour voler les clés de chiffrement de milliards, oui, des milliards de téléphones portables à travers le monde. Cela a été possible parce que nous, consommateurs, ne voulons pas faire l'effort de générer ces clés et configurer notre propre sécurité lorsque nous allumons pour la première fois nos téléphones ; nous voulons que ce soit fait automatiquement par les fabricants. Nous voulons que nos données soient sécurisées, mais nous voulons que quelqu'un puisse les récupérer intégralement lorsque nous oublions notre mot de passe.

Nous ne résoudrons jamais ces problèmes de sécurité tant que nous serons notre pire ennemi. C'est pourquoi je crois que toute solution de sécurité à long terme ne sera pas seulement technologique, mais aussi politique. Nous avons besoin de lois pour protéger notre vie privée de ceux qui respectent les lois, et pour punir ceux qui les transgressent. Nous avons besoin de lois qui exigent de ceux à qui nous confions nos données qu'ils protègent nos données. Certes, nous avons besoin de meilleures technologies de sécurité, mais nous avons également besoin de lois qui imposent l'usage de ces technologies.

Crédit photo : Nicubunu (CC BY-SA 2.0)

Les algos peuvent vous pourrir la vie

Les algorithmes[1] ne sont guère qu'une série d'instructions pas-à-pas généralement exécutées par un programme sur une machine. Cependant leur complexité et leur opacité pour le commun des mortels sont redoutables, et bien plus encore leur omniprésence dans tous les compartiments de notre vie, y compris la plus intime. Si le code fait la loi, c'est justement parce que les algorithmes sont à la fois puissants, invasifs et sont devenus aujourd'hui indispensables.
L'article ci-dessous ne met pas l'accent sur les nombreux domaines où nous utilisons des algorithmes sans en avoir conscience, il pointe davantage les risques et menaces qu'ils représentent lorsque ce sont les algorithmes qui déterminent notre existence, à travers quelques exemples parmi bien d'autres. Il pose également l'intéressante question de la responsabilité de ceux qui élaborent les algorithmes. Suffira-t-il de réclamer des concepteurs d'algorithmes un sympathique engagement solennel à la manière de celui des acteurs du Web ?

Les codeurs dont les algos contrôlent nos vies, qui les contrôle ? Pouvons-nous avoir un droit de regard sur les algorithmes qui désormais menacent de régir nos vies ?

Les algorithmes sont formidables mais peuvent aussi ruiner des vies

Extrait de l'essai (en anglais) The Formula: How Algorithms Solve All Our Problems—and Create More par **Luke Dormehl.**

Source : article du magazine **Wired** Algorithms are great and all, but can also ruin our lives

Traduction Framalang : Wan, r0u, goofy, Sphinx, sinma, Omegax, ylluss, audionuma

Le 5 avril 2011, John Gass, 41 ans, a reçu un courrier du service d'enregistrement des véhicules motorisés (Registry of Motor Vehicles ou RMV) de l'État du Massachusetts. La lettre informait M. Gass que son permis de conduire avait été annulé, qu'il lui était désormais interdit de conduire et que cela prenait effet immédiatement. Le seul problème, c'est qu'en bon conducteur n'ayant pas commis d'infraction grave au code de la route depuis des années, M. Gass n'avait aucune idée du motif de ce courrier.

Après plusieurs appels téléphoniques frénétiques, suivis par une entrevue avec les fonctionnaires du service, il en a appris la raison : son image avait été automatiquement signalée par un algorithme de reconnaissance faciale conçu

pour parcourir une base de données de millions de permis de conduire de l'État, à la recherche de possibles fausses identités criminelles. L'algorithme avait déterminé que Gass ressemblait suffisamment à un autre conducteur du Massachusetts pour présumer d'une usurpation d'identité, d'où le courrier automatisé du RMV.

Les employés du RMV se sont montrés peu compréhensifs, affirmant qu'il revenait à l'individu accusé de prouver son identité en cas d'erreur quelconque et faisant valoir que les avantages de la protection du public l'emportaient largement sur les désagréments subis par les quelques victimes d'une accusation infondée.

John Gass est loin d'être la seule victime de ces erreurs d'algorithmes. En 2007, un bogue dans le nouveau système informatique du Département des services de santé de Californie a automatiquement mis fin aux allocations de milliers de personnes handicapées et de personnes âgées à bas revenus. Leurs frais d'assurance maladie n'étant plus payés, ces citoyens se sont alors retrouvés sans couverture médicale.

Là où le système précédent aurait notifié les personnes concernées qu'elles n'étaient plus considérées comme éligibles aux allocations en leur envoyant un courrier, le logiciel maintenant opérationnel, CalWIN, a été conçu pour les interrompre sans avertissement, à moins de se connecter soi-même et d'empêcher que cela n'arrive. Résultat : un grand nombre de ceux dont les frais n'étaient plus pris en charge ne s'en sont pas rendu compte avant de recevoir des factures médicales salées. Encore beaucoup n'avaient-ils pas les compétences nécessaires en anglais pour naviguer dans le système de santé en ligne et trouver ce qui allait de travers.

Des failles similaires sont à l'origine de la radiation de votants des listes électorales sans notification, de petites entreprises considérées à tort comme inéligibles aux contrats gouvernementaux, et d'individus identifiés par erreur comme

« parents mauvais payeurs ». Comme exemple notable de ce dernier cas, Walter Vollmer, mécanicien de 56 ans, a été ciblé à tort par le Service fédéral de localisation des parents, et s'est vu envoyer une facture de pension alimentaire à hauteur de 206 000 $. L'épouse de M. Vollmer, 32 ans, a par la suite montré des tendances suicidaires, persuadée que son mari avait eu une vie cachée pendant la majeure partie de leur mariage.

Une possibilité tout aussi alarmante : qu'un algorithme puisse ficher par erreur un individu comme terroriste. Un sort qui attend chaque semaine environ 1500 voyageurs malchanceux qui prennent l'avion. Parmi les victimes passées de ces erreurs de corrélation de données, on retrouve d'anciens généraux de l'armée, un garçon de quatre ans, ainsi qu'un pilote d'*American Airlines*, qui a été détenu 80 fois au cours d'une même année.

Beaucoup de ces problèmes sont dus aux nouveaux rôles joués par les algorithmes dans l'application de la loi. Les budgets réduits menant à des réductions de personnel, les systèmes automatisés, auparavant de simples instruments administratifs, sont maintenant des décideurs à part entière.

Dans nombre de cas, le problème est plus vaste que la simple recherche d'un bon algorithme pour une tâche donnée. Il touche à la croyance problématique selon laquelle toutes les tâches possibles et imaginables peuvent être automatisées. Prenez par exemple l'extraction de données, utilisée pour découvrir les complots terroristes : de telles attaques sont statistiquement rares et ne se conforment pas à un profil bien défini comme, par exemple, les achats sur Amazon. Les voyageurs finissent par abandonner une grande partie de leur vie privée au profit des algorithmes d'extraction de données, avec peu de résultats, si ce n'est des faux-positifs. Comme le note Bruce Schneier, le célèbre expert en sécurité informatique :

> *Chercher des complots terroristes… c'est comme chercher une aiguille dans une botte de foin, ce n'est pas en accumulant*

> *davantage de foin sur le tas qu'on va rendre le problème plus facile à résoudre. Nous ferions bien mieux de laisser les personnes chargées d'enquêtes sur de possibles complots prendre la main sur les ordinateurs, plutôt que de laisser les ordinateurs faire le travail et les laisser décider sur qui l'on doit enquêter.*

Bien qu'il soit clair qu'un sujet aussi brûlant que le terrorisme est un candidat parfait pour ce type de solutions, le problème central se résume encore une fois à cette promesse fantomatique de *l'objectivité* des algorithmes. « Nous sommes tous absolument effrayés par la subjectivité et l'inconstance du comportement humain », explique Danielle Citron, professeur de droit à l'Université du Maryland. « Et à l'inverse, nous manifestons une confiance excessive pour tout ce que peuvent accomplir les ordinateurs ».

Le professeur Citron suggère que l'erreur vient de ce que nous « faisons confiance aux algorithmes, parce que nous les percevons comme objectifs, alors qu'en réalité ce sont des humains qui les conçoivent, et peuvent ainsi leur inculquer toutes sortes de préjugés et d'opinions ». Autrement dit, un algorithme informatique a beau être impartial dans son exécution, cela ne veut pas dire qu'il n'a pas de préjugés codés à l'intérieur.

Ces erreurs de jugement, implicites ou explicites, peuvent être causées par un ou deux programmeurs, mais aussi par des difficultés d'ordre technique. Par exemple, les algorithmes utilisés dans la reconnaissance faciale avaient par le passé de meilleurs taux de réussite pour les hommes que pour les femmes, et meilleurs pour les personnes de couleur que pour les Blancs.

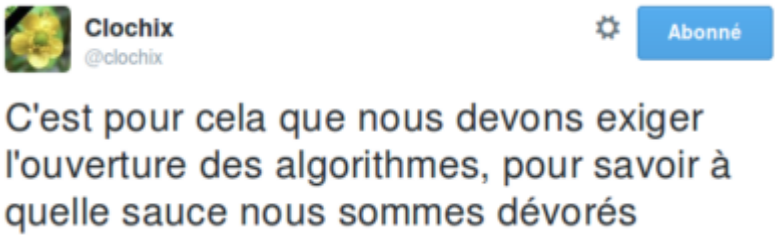
Ce n'est pas par préjugé délibéré qu'un algorithme ciblera plus d'hommes afro-américains que de femmes blanches, mais cela ne change rien au résultat. De tels biais peuvent aussi

venir de combinaisons plus abstraites, enfouies dans le chaos des corrélations de jeux de données.

Prenez par exemple l'histoire de l'afro-américaine Latanya Sweeney, docteure de l'Université d'Harvard. En effectuant des recherches sur Google, elle fut choquée de découvrir que les résultats de ses recherches étaient accompagnés de publicités demandant : « Avez-vous déjà été arrêté(e) ? ». Ces annonces n'apparaissaient pas pour ses collègues blancs. Sweeney se lança alors dans une étude, démontrant que les outils d'apprentissage automatique utilisés par Google étaient incidemment racistes, en associant plus souvent des noms donnés à des personnes noires avec des publicités ayant trait aux rapports d'arrestation.

Le système de recommandation de Google Play révèle un problème similaire : il suggère aux utilisateurs qui téléchargent *Grindr*, un outil de réseautage social basé sur la localisation pour les gays, de télécharger également une application qui assure le suivi géolocalisé des délinquants sexuels. Au vu de ces deux cas, devons-nous conclure que les algorithmes ont fait une erreur, ou plutôt qu'ils sont révélateurs des préjugés inhérents à leurs concepteurs ? Ou, ce qui semble plus probable, ne seraient-ils pas révélateurs d'associations inappropriées et à grande échelle entre – dans le premier cas – les personnes noires et le comportement criminel, et – dans le deuxième cas – l'homosexualité et les agressions sexuelles ?

Peu importe la raison, peu importe la façon répréhensible dont ces corrélations codifiées peuvent exister, elles révèlent une autre face de la culture algorithmique. Quand un seul individu fait explicitement une erreur de jugement, il ne peut jamais affecter qu'un nombre fini de personnes. Un algorithme, quant à lui, a le potentiel d'influer sur un nombre de vies exponentiellement plus grand.



Pour aller plus loin, 4 articles en français sur le même sujet :

- Surveiller les algorithmes
- Ces algorithmes qui vous nous gouvernent
- Ouvrir les modèles, pas seulement les données
- Le jaguar et le bus scolaire

Note

[1] Pour une définition plus élaborée voir Qu’est-ce qu’un algorithme