LIVRE BLANC

Les données personnelles à l'heure du big data :

De l'intelligence artificielle au pouvoir des algorithmes

CHAPITRE 1 – Historique et cadre contemporain

Éléments d'une histoire du fichage et des libertés privées en France

L'histoire du fichage des individus est immanquablement liée à celle du pouvoir et de son exercice. Le fichage en constitue un moment particulier, il vient après que les sociétés démocratiques aient renoncé à la force. Il n'empêche, quoique plus subtile que la décapitation en place publique, il correspond toujours à ce même besoin de coercition des sociétés sur les individus, besoin qui s'entend, dans le dialogue notamment entre « information » et « contrôle ».

Les gouvernants ont souvent fait un usage abusif des fichiers de personnes. En 1904, dans une période marquée par un fort anticléricalisme gouvernemental, le ministre de la guerre assurait la promotion des officiers les plus républicains à partir de fiches sur lesquelles étaient consignées les opinions politiques et religieuses des cadres de l'armée. Agressé à la tribune de l'Assemblée nationale, ce ministre dut démissionner.

L'informatisation des fichiers aggrave leurs dangers pour les libertés démocratiques comme devait le montrer en 1974, l'affaire SAFARI. Ce système automatisé pour les fichiers administratifs et le répertoire des individus, organisait à partir de l'identifiant unique que constitue le numéro INSEE attribué à chaque individu, l'interconnexion de l'ensemble des fichiers qui le concerne. Après une campagne de presse et une mobilisation citoyenne dénonçant le caractère liberticide d'une telle opération – et notamment la publication le 21 mars 1974 dans les colonnes du journal *Le Monde* d'un article intitulé : « SAFARI ou la chasse aux Français » – le gouvernement nomma une commission d'étude dont les travaux ont abouti au vote d'une loi « relative à l'informatique, aux fichiers et aux libertés » en 1978. Cette loi pose des limites au fichage et donne des garanties et de nouveaux droits aux personnes fichées. Dès lors, la création et la gestion des fichiers de personnes doivent respecter un certain nombre de règles et sont soumis au contrôle d'une commission indépendante : la commission nationale de l'informatique et des libertés (CNIL). Elle a défini des grands principes quant à la mise en place des fichiers de personnes : finalité, proportionnalité, durée de conservation des données, droit d'accès et de rectification, etc.

La polémique provoquée par la mise en place du système SAFARI a été d'autant plus vive que ce système faisait remonter à la surface, les souvenirs douloureux d'un fichage à peine passé. En effet, le numéro INSEE, pivot de l'opération, a été créé sous le régime de Vichy en 1941, après la création par ce même régime en 1940, d'une carte d'identité obligatoire pour tous les Français. Or ce numéro avait une dimension raciste, car la numérisation qu'il proposait, permettait de distinguer une population juive à l'intérieur de la population française. Ce numéro n'a pas servi la politique d'extermination nazie grâce aux actions de sabotage menées par son créateur, responsable du service national des statistiques. Il n'en a pas été de même dans des pays voisins comme les Pays-Bas où le responsable du service des statistiques a obéi aux ordres nazis. À l'aide d'une numérotation de la population qui permit une identification rapide et d'un parc important de machines mécanographiques, les trois quarts des juifs habitant dans ce pays furent exterminés.

1. **Des individus exposés et trop souvent consentants face à l'utilisation de leurs données**

Avec l'informatique et les techniques numériques, les fichiers de personnes se sont multipliés. Ces techniques offrent des possibilités de stockage et de traitement de l'information qui apparaissent sans limites ; les données collectées visent la connaissance intrusive des individus à travers leurs comportements. L'État-providence se mue au fil du temps en un État policier ; chaque individu est concerné, en moyenne, par plus de 500 fichiers et traitements. État pourvoyeur, État protecteur… les populations les plus fragiles sont fichées en priorité dans la mesure où toute aide sociale fait l'objet d'une collecte d'information auprès de son bénéficiaire. Depuis le 11 septembre 2001 et les multiples attentats terroristes qui ont suivi, on a assisté à une création ininterrompue de fichiers policiers. On en dénombre aujourd'hui environ une centaine. Le citoyen est, dès lors, considéré comme suspect. Des fichiers mis en place, souvent par décret, et visant au départ une population ciblée – par exemple les auteurs de crimes sexuels – étendent indûment leur champ d'application. Le Fichier National Automatisé des Empreintes Génétiques (FNAEG) en est un exemple flagrant.

Malgré l'abondance d'information sur les personnes apportée par les outils numériques, le fichage se poursuit dans tous les secteurs d'activité : aide sociale, police, mais aussi école, santé, transports, etc. La dernière création en date, le méga-fichier « Titres Electroniques Sécurisés » (TES) réalise la *biométrisation* de toute la population française, opération en faveur de laquelle tous les gouvernements oeuvraient sans succès depuis plus de 30 ans, en étendant aux cartes d'identité le fichage des données biométriques déjà en vigueur pour les passeports. On le savait déjà, un fichier potentiellement liberticide pour l'individu peut de surcroît en cacher un autre…

Les fichiers informatiques, leurs interconnexions et croisements, les divers traitements extrêmement personnalisés de données permettent un contrôle accru de la population. Par exemple, des profils à risques vont être réalisés pour identifier des cibles qui feront l'objet d'une surveillance particulière. Bien que cela ne soit pas nouveau, les traitements de données à finalité de prédiction des comportements, conformes à l'air du temps, se multiplient. La loi Renseignement de 2015 autorise, chez les opérateurs et hébergeurs, l'installation de « boîtes noires » à vocation prédictive. Elles sont considérées par plusieurs scientifiques, comme non fiables : les traitements prédictifs décèleront vraisemblablement plus à tort qu'à raison (plus de « faux positifs » que de « vrais positifs »).

La loi de 1978 et la directive européenne de 1995 (transposée en droit français en 2004), au nom de la préservation de la vie privée, avaient posé des limites à ces possibilités. Cependant une obsession voire une stratégie sécuritaire des gouvernants, rend de plus en plus difficile les compromis réalisés jusqu'alors grâce à l'intervention de la Commission informatique et libertés dont les pouvoirs ont été sur ce point sévèrement revus à la baisse. Avec la mondialisation de la communication numérique, la réglementation française ou européenne relative aux données personnelles n'est pas, ou peu, respectée par les grandes entreprises monopolistiques américaines sur Internet. La donnée personnelle dont l'individu, parfois consentant, se découvre de fait dépossédé, est devenue une marchandise échangeable et vendue sur un marché qui explose avec un Internet marchand et des objets communicants. Les fichiers et applications prélèvent et exploitent massivement, entre

autres, des informations d'un genre nouveau, des traces prélevées sur le contexte. Leur utilisation permet de préciser des profils d'usagers, pour ainsi anticiper des comportements.

Cependant comme l'a révélé Edward Snowden en juin 2013, ces données peuvent faire, outre l'exploitation commerciale, l'objet d'une appropriation étatique.

2. **Exploitation à des fins marketing et de propagande contre gratuité d'un service**

De nombreux « services » sont proposés *via* Internet aux utilisateurs et présentés comme « gratuits ». Mais force est de constater, avec les majors du Net, que « la gratuité n'est pas gratuite ! ». Le modèle économique des GAFAM (Google, Amazon, Facebook, Apple, Microsoft) et de tous ceux qui se sont lancés à leur suite s'appuie en partie sur l'extraction et l'exploitation des données personnelles. Pour offrir recherches et services personnalisés, il faut connaître l'utilisateur, ses habitudes, ses « amis », sa localisation, ses horaires de travail, etc. Tout devient prétexte à collecter des données, à les traiter, à les fusionner et bien sûr … à les vendre. Des algorithmes, souvent compliqués et opaques, traitent du *big data*. La réussite des services en ligne tient, pour une large part, à la personnalisation qui s'opère devant l'individu et qui stipule qu'il est au centre du jeu. Il en oublie la collecte, souvent à son insu, et la marchandisation de ses données et s'habitue d'une part à ce qu'elles ne soient plus son exclusive propriété et, d'autre part, à ne plus trop y porter attention. Il y est aidé par des chartes peu lisibles d'utilisation des données et une offre particulièrement peu symétrique qui dépossèdent celui ou celle qui, afin de bénéficier du service, en aura accepté les Conditions Générales d'Utilisation (CGU). Pourtant, il s'agit là souvent d'informations très personnelles sur nos habitudes de consommation, notre façon de penser, notre vie sociale… Il se pourrait bien alors que demain nous ne sachions plus rien de ce que l'on sait de nous, de nos clones informatiques.

Profilé, le consommateur est aussi manipulé. Son traçage de navigation, par exemple, va orienter les affichages des bandeaux publicitaires et pour l'ensemble des internautes les suggestions faites lors d'une saisie faite sur un moteur de recherche. Il faut faire entendre au « client » ce qu'il a envie d'entendre. Et tant pis pour l'objectivité, voire la véracité des réponses. Ce qui fait la « pertinence » d'un document n'est pas son contenu mais le nombre de fois où il a été visité. L'évaluation se fait au nombre de clics. Pour orienter une opinion, divers acteurs font appel à des sociétés « fermes à clics » où des salariés mal payés n'ont d'autre choix que de cliquer pour n'importe quelle « cause ». Agrégateurs de données, *data brokers* (courtiers en données), travaillent non seulement pour le marketing mais aussi pour la propagande politique. Ce sont l'autonomie de l'individu, ses choix voire la démocratie qui sont ici en jeu.

L'Internet des objets et la multiplication des applications embarquées vont accélérer, si on n'y prend garde, la collecte de données présentée pour apporter un mieux-être, faciliter notre vie. Les applications d'e-bien-être vont, prétendument pour notre bien, nous barder de prothèses, afin de collecter des informations sur notre rythme cardiaque, nos différents paramètres biologiques, nos habitudes alimentaires, etc. Qui aura accès à ces informations ? Qu'en sera-t-il fait ?

CHAPITRE 2 – Aspects juridiques et rapport à la connaissance

1. Une différence culturelle fondamentale entre la France et les États-Unis

Le rapport aux données est plus sensible en France qu'aux États-Unis comme exposé au chapitre 1 du fait de l'histoire et des heures sombres du gouvernement de Vichy. C'est l'*opt-in* qui prévaut à savoir le consentement préalable de l'internaute (*via* une case à cocher par exemple) pour l'envoi d'informations publicitaires, d'une newsletter, etc. par courrier électronique. Il peut même exister un double *opt-in* où après avoir coché une case pour acceptation, un message est à adresser à une adresse mél convenue ou un clic est à opérer sur un lien reçu dans un mél pour signifier l'intérêt de l'internaute. Sans consentement préalable, tout envoi massif non sollicité est considéré comme du spam. Aux États-Unis, *a contrario*, mais aussi en France pour l'e-mailing en B2B, c'est l'*opt-out* (option de retrait) qui prévaut. Dans ce cas, c'est après avoir reçu une sollicitation que l'internaute peut manifester le souhait de ne plus recevoir de messages à l'avenir, *via* une case à cocher ou l'envoi d'un mél pour se désinscrire. Les États-Unis sont plus dans une logique de marchandisation des informations.

Les conditions générales d'utilisation (CGU) définissent les droits et les obligations réciproques d'une entreprise qui fournit une application sur Internet ou une application sur smartphone (ordiphone) d'une part et les utilisateurs d'autre part. Pour utiliser un service, il convient d'accepter l'ensemble des clauses. Si une d'entre elles ne nous convient pas, une utilisation partielle ou avec réserve du service ne peut être faite. Soit on accepte le tout comme le vote pour un candidat politique soit on n'utilise pas le service. Cette logique booléenne facilite le fait d'imposer des clauses souvent léonines en faveur de l'application. En l'échange de l'utilisation des données de l'internaute par l'entreprise qui fournit le service, celui-ci peut être utilisé soit gratuitement soit en formule freemium (service gratuit pour tous et service optionnel payant avec des fonctionnalités additionnelles comme le fait LinkedIn par exemple). Toutefois il est possible pour des services de désactiver des fonctions comme la géolocalisation qui est souvent activée par défaut. Néanmoins, certaines informations sont produites par les utilisateurs eux-mêmes sans leur consentement préalable, comme le fait de taguer une photo sur Facebook où l'on apparaît. L'intérêt pour les plateformes qui proposent ces services est bien évidemment d'inciter les utilisateurs à produire le plus de contenu possible qui pourra faire l'objet d'exploitation ultérieurement à des fins marketing. Par ailleurs, les CGU évoluent avec le temps et ne sont pas figées. Des services jadis gratuits peuvent devenir payants et il est parfois difficile pour l'utilisateur de migrer ses données produites (textes, images, photos, sons, etc.) d'une plateforme à une autre, chaque outil veillant judicieusement pour assurer sa pérennité à s'ouvrir à l'extérieur *via* des APIs tout en limitant ou contraignant la migration des données de l'internaute vers un autre outil.

2. De nécessaires droits à instaurer

La société numérique dans laquelle nous vivons emporte de nouveaux droits qu'il convient d'avoir présents à l'esprit et à faire valoir.

Face aux excès des informations collectées et stockées, certaines informations pourraient ou devraient être effacées pour ne pas causer préjudice à l'internaute sur le long terme. On pense à des maladresses d'un adolescent sur les réseaux sociaux qui pourraient se retourner contre lui lorsqu'il sera à la recherche d'un premier emploi. Un équilibre est, bien sûr, à trouver entre respect de la vie privée et droit à l'oubli d'une part et devoir de mémoire de l'autre même si celui-ci concerne plus des groupes de personnes que des individus.

Soumis à davantage de pression le salarié doit pouvoir se déconnecter pour instaurer de nouvelles frontières entre sa vie personnelle et sa vie professionnelle alors même que les nouveaux outils comme le smartphone et la possibilité de consulter ses méls professionnels souvent en tout lieu, à tout moment et sur tout type d'outil, ne lui permet plus de sas de décompression pour profiter de la vie avec des amis ou en famille sans être sollicité. Le droit à la déconnexion est une des dispositions de la loi El Khomri entrée en vigueur au 1er janvier 2017 pour les employés forfaitisés ou oeuvrant en télétravail. Celui-ci est malheureusement sans véritable impact pratique contrairement à d'autres droits comme celui lié à la portabilité des numéros de téléphone mobile.

Enfin la non-géolocalisation est le fait de pouvoir désactiver la géolocalisation souvent permise par ses smartphones, tablettes et de nombreuses applications associées. Et le silence des puces est le fait de pouvoir désactiver les puces ambiantes qui se développent avec l'Internet des objets et dans le cadre de l'habitat connecté et plus largement la ville intelligente (*smart city*).

3. L'importance de la question de l'enseignement avant même la majorité numérique

Pour les défenseurs de la vie privée et plus généralement pour l'ensemble des citoyens, un vaste chantier s'ouvre, celui de la formation aux enjeux des outils numériques et les bouleversements qu'ils induisent dans nos comportements et les conséquences pour nous-mêmes et le futur. Les acteurs, les promoteurs, les enjeux de l'informatisation ont changé. Comment faire de chaque citoyen, dont les jeunes adolescents rivés dans leur chambre sur l'écran de leur smartphone, un utilisateur pleinement conscient des risques pour ses libertés et sa vie privée ? Comment penser collectivement ce qu'il adviendra du déluge technologique en cours ? L'enseignement, l'éducation ont bien sûr leur rôle à jouer. L'éducation au numérique, introduit dans les programmes scolaires, ne devrait pas se limiter à la seule approche technique indispensable comme l'algorithme avec L'Informatique et Sciences du Numériques (ISN) qui se généralise progressivement dans les cursus – et c'est une bonne chose. Mais elle doit aussi sensibiliser aux enjeux de protection de la vie privée et des données personnelles pour faire des citoyens éclairés. De nombreux outils sont utilisables dès 13 ans comme Facebook qui fixe à cet âge-là la majorité numérique. Or, l'utilisateur peut tricher quant à son année de naissance (près de 30 % des 11 à moins de 13 ans possèdent en effet un compte Facebook et même un peu plus s'agissant de Snap ou d'Instagram). Des solutions techniques plus protectrices des données personnelles, souvent dans le monde du logiciel libre, sont disponibles, mais il faut les faire connaître et les vulgariser. Là encore, la formation et la sensibilisation sont indispensables en mettant en exergue les impacts qui en découlent et le bagage juridique minimum à acquérir par rapport aux risques : fraude à l'identité, détournement d'information et de données personnelles, etc.

CHAPITRE 3 – L'explosion de la quantité de données

1. Explosion de la quantité de données avec le big data

Le big data désigne l'explosion de la quantité de données. Elles résultent de données de transaction (par exemple sur les sites de ventes en ligne), de comportements (publications nombreuses sur les médias sociaux), d'enregistrements (météo, astronomie, physique des particules), liées à la géolocalisation (par exemple badge Navigo pour la RATP). Ces données proviennent tant du réseau Internet que de la part d'objets intelligents (par exemple puces RFID) avec deux facteurs structurants, le développement des mobiles et la multiplication des capteurs. Les informations abondent de toute part et elles sont d'une grande diversité et présentent un intérêt quant à leur exploitation par des grands groupes, des banques et des États.

Dans une définition basique issue du Gartner, le big data se définit selon trois dimensions, les 3 V : volume, variété et vitesse.

Le volume des données stockées sur support magnétique s'accroît à un rythme exponentiel : les données numériques créées dans le monde étaient estimées à 2,8 zettaoctets (10^{21} octets) en 2012 et elles s'élèveront à 40 zettaoctets en 2020. L'essentiel de cette masse d'informations proviendrait des bases de données, des systèmes informatiques et de tous les dispositifs chargés de capter, puis de conserver, les traces de nos actions. Le cabinet IDC estime par ailleurs que la masse des informations relatives aux individus a déjà dépassé le volume de données qu'ils sont eux-mêmes capables de produire.

La variété des données rend leur exploitation plus complexe qu'avec les traditionnelles bases de données relationnelles (avec des tables relationnelles et des langages de requêtes de type SQL). Elles comprennent des données issues du Web selon différents formats (textes, images, vidéos, etc.), etc. Des outils spécifiques ont été créés pour exploiter ces données non structurées. Ainsi aujourd'hui MapReduce popularisé par Google et la plateforme Hadoop règnent. Pour autant, les techniques sont appelées à évoluer, le marché n'étant pas encore stabilisé. Concrètement, de façon récursive, un problème est découpé en sous-problèmes et distribué à des nœuds qui vont pouvoir effectuer des calculs en parallèle (*map*) et les résultats sont remontés et récupérés par des nœuds (*reduce*). Il existe des alternatives à Hadoop telles MongoDB, Cassandra ou CouchDB. Une complexité découle de la nécessité de disposer de machines travaillant en parallèle sur les données. Ceci induit de nouveaux paradigmes pour les développeurs jusqu'alors habitués à des langages séquentiels.

La vitesse est propre à la génération, au partage et à la mise à jour des données qui peuvent nécessiter pour être pertinentes des analyses en quasi temps-réel (par exemple données boursières).

Le big data est né de la conjonction de plusieurs évolutions tant dans le domaine du matériel que dans celui des logiciels :

- Accroissement exponentiel de la vitesse des microprocesseurs selon la loi de Moore qui fait que leur puissance double tous les 18 mois ainsi que des capacités de stockage de l'information.
- Développement des moteurs de recherche à la fin des années 1990 rendu possible par la construction de « fermes de PC » et l'apparition de nouveaux systèmes de bases de données distribuées (noSQL) et de framework (par exemple Hadoop) permettant aux applications de travailler avec des milliers de nœuds et des quantités de données « quasi infinies ».
- Apparition et généralisation des réseaux sociaux surfant sur la vague des smartphones à l'aide des technologies précédentes.
- Nouveaux algorithmes (provenant de recherches en intelligence artificielle basées sur les réseaux de neurones) permettant d'extraire massivement des informations pertinentes de masses de données non structurées

2. De nouveaux usages induits

Le cloud computing (informatique dans les nuages) rend possible le stockage tant des données que des applications sur des serveurs. Celles-ci sont accessibles partout dans le monde à partir d'un ordinateur ou d'un smartphone. Les nouveaux modes relationnels entre individus dans le monde passent par les réseaux sociaux avec plus de 3 milliards d'individus présents. En outre, on assiste à un déclin des médias classiques (presse, radio, télévision) qui ne sont plus dominants et qui se réinventent avec Internet avec de nouveaux modèles économiques (par exemple paiement au-delà de la consultation d'un certain nombre d'articles dans le mois) alors que les réseaux sociaux deviennent des sources d'informations privilégiées, par exemple Twitter pour avoir des scoops. Les achats en ligne, les connexions des utilisateurs de réseaux sociaux sont évidemment enregistrés et sont exploités systématiquement par la publicité. Elle peut ainsi quasiment individualiser chaque message et chaque annonce de nouveaux produits selon les algorithmes utilisés.

Les données peuvent désormais faire l'objet d'études statistiques et de segmentations plus fines. L'utilisation de statistique inférentielle sur des données à faible densité en information donne dès lors (dans certaines limites) au big data des capacités prédictives. Le grand volume d'informations disponible permet d'inférer sous certaines conditions des lois. La simple étude des seules métadonnées (numéro appelé ou appelant, borne d'accès, durée de la communication) des communications entre smartphones, sans utiliser ni le contenu des messages ni les données GPS, est déjà l'objet de centaines de recherches permettant d'évaluer les risques d'épidémie dans une région du monde (grippe, dengue, etc.), les déplacements de population, les besoins en énergie ou en moyen de déplacement. Des applications prétendent même ainsi déduire les lieux des futures activités criminelles dans les grandes villes.

3. De nouveaux risques pour les données personnelles

De nouveaux risques sont les corollaires des nouveaux usages. L'affaire Snowden a révélé l'ampleur de la surveillance des communications sur Internet par la NSA et les services de renseignements américains. D'autres pays ne sont pas en reste dans ce domaine à des degrés divers. En Chine par exemple, la surveillance institutionnalisée est de mise avec le

« bouclier doré ». Elle va de pair avec des dénonciations des citoyens eux-mêmes, ce qui génère une auto-régulation liberticide. On assiste à un renforcement de la surveillance étatique des réseaux informatiques et téléphoniques justifiée sous couvert de tensions internationales et de la lutte contre le terrorisme. Les entreprises interviennent sur deux fronts : la surveillance de leurs salariés d'une part, de leur clientèle qu'il faut fidéliser d'autre part.

Chacun devient aussi acteur de cette surveillance en exportant des images de « soi » et produit des traces exploitables à son insu (et exploitées). Toutes ces informations se déversent dans les serveurs des fournisseurs d'accès à Internet et des géants du Web, au premier rang desquels les GAFAM (Google, Apple, Facebook, Amazon, Microsoft). Elles constituent la matière première aux analyses des trop fameux « algorithmes » qui font la une de l'actualité.

CHAPITRE 4 – Précautions et conduite à tenir pour faire face à la cybersurveillance à l'ère du numérique

L'informatisation d'une large partie de nos vies nous rend indubitablement de grands services : facilitation de nos communications et de notre accès à l'information, mise en pratique d'une réelle liberté d'expression, possibilités de travailler collectivement plus facilement, etc.

Ces évolutions positives impliquent toutefois la numérisation d'une partie de nous-mêmes. Pour pouvoir utiliser la machine, il est nécessaire de s'y soumettre en parlant son langage. Cela passe par une « traduction » de nos actions et informations en « données » normalisées, compréhensibles par les programmes informatiques. Cette numérisation, de nos personnes et nos actes, peut être volontaire, mais elle est également bien souvent « subie », quand nous naviguons sur Internet où utilisons divers programmes, ceux-ci peuvent analyser nos comportements et les transformer en données qui pourront être comprises et traitées automatiquement. L'amélioration des méthodes d'analyse des données a, entre autres, démultiplié les possibilités de surveillance et de manipulation publique comme privée.

Il est désormais possible pour Facebook de « prédire » la future rupture d'un couple, de modifier l'expression de nos sentiments en jouant sur les contenus que nous montre la plateforme. Des sociétés se trouvent également en capacité d'analyser des millions de profils de citoyens et diffuser une propagande politique ultra-ciblée pour influer leurs votes futurs ou encore de profiter d'action d'une multitude d'utilisateurs pour générer de la valeur sans indemniser qui que ce soit (voir en ce sens les travaux d'Antonio Casilli sur le *Digital labor)* et au passage pratiquer une évasion fiscale très agressive. Ces actions peuvent aussi bien être réalisées par des sociétés privées, notamment les géants du numérique, mais également les « courtiers en données », que par des États. Les États disposent par ailleurs de leurs propres outils de surveillance tels que, en France, ceux légalisés par la loi du 24 juillet 2015 relative au renseignement et qui peuvent de surcroit s'appuyer sur la coopération plus ou moins volontaire de ces sociétés privées comme ont pu le montrer les révélations d'Edward Snowden.

Si, malheureusement, il reste encore des personnes pour considérer que cela n'est pas grave, car celles-ci n'ont « rien à cacher », beaucoup ont pris conscience de ces enjeux et souhaitent limiter ces possibilités de surveillance, d'exploitation numérique et d'influence.

Néanmoins, pour beaucoup la tâche n'apparaît pas aisée et même parmi les citoyens les plus militants beaucoup négligent de mettre en place des processus solides de sécurisation de leurs données.

Pourtant, l'informatique est sur ce point très ambivalente : si elle permet des méthodes de surveillance extrêmement intrusives et peu coûteuses, elle offre aussi des possibilités de protection individuelle très poussées. Il est ainsi nécessaire de ne pas toujours céder aux sirènes de la facilité et de services « trop » connus qui nous prennent totalement par la main et en main. On peut à la place se diriger vers des solutions et outils potentiellement moins attrayants, mais plus respectueux de notre vie privée et de nos données personnelles.

Cela passe par ce que l'on peut appeler de « l'autodéfense numérique » : le développement de compétences de base pour limiter les atteintes des grands prédateurs du numérique. C'est un enjeu important pour les défenseurs des libertés : permettre à tous de se réapproprier une certaine maîtrise sur les outils qu'ils utilisent au quotidien dans une approche d'autodétermination informationnelle.

Un préalable à l'apprentissage de ces méthodes est celui d'une « hygiène numérique », autrement dit des pratiques respectueuses des libertés.

1. Une nécessaire hygiène numérique

L'apprentissage de l'informatique est souvent très « opérationnel ». On apprend à utiliser les logiciels (éditeur de texte, navigateur, processus d'échange de courriel, tableur, logiciel de lecture de médias…) sans forcément expliquer les mécanismes sous-jacents et les enjeux de ces choix ni toujours insister sur l'importance des bonnes pratiques.

Réaliser correctement des sauvegardes, protéger les accès à ses appareils (verrouillage, choix des mots de passe complexes difficiles à deviner pour un attaquant éventuel qu'il soit humain ou une machine, chiffrement), avoir une gestion sécurisée de ses mots de passe, segmenter ses usages et ses comptes, limiter la transmission d'informations sur Internet, mettre à jour ses applications, savoir détecter un message ou un site frauduleux, vérifier et maîtriser les paramètres de ses applications, utiliser des outils de protection de base (pare-feu, antivirus, protection de sa connexion Internet, HTTPS…), comprendre ce qu'est un ordinateur, un navigateur, etc.

Toutes ces actions qui paraissent évidentes sont bien souvent négligées par beaucoup et facilitent grandement le travail d'attaquants qui souhaiteraient mettre la main sur vos données. Au-delà des mécanismes de protection spécifiques contre la surveillance, il est important de se former à de bons usages du numérique et même pour un usage limité. Pour ce faire, il ne faut pas hésiter à se diriger vers les « Espaces Publics Numériques » et les autres associations compétentes en la matière, notamment les associations promouvant l'utilisation de logiciels libres.

Il s'agit d'étapes fondamentales dans un « processus de sécurité ». La sécurité n'est pas un produit fini, c'est un processus que l'on peut améliorer au quotidien en allant toujours vers une meilleure compréhension et maîtrise de ses outils sachant que les techniques d'attaques évoluent par ailleurs. Une fois cette maîtrise de base mise en place, il sera facile pour quiconque de se diriger, petit pas par petit pas, vers des solutions et pratiques toujours plus protectrices de ses libertés.

2. Un peu d'autodéfense numérique

De nombreux documents (le guide d'autodéfense numérique : guide.boum.org, le livre Surveillance:// de Tristan Nitot, le site prism-break.org, etc.) existent pour orienter l'utilisateur vers des outils qui vont permettre à chacun et chacune de se réapproprier, dans une certaine mesure, la défense de ses données sur Internet. Ainsi, le CECIL édite un « guide de survie à destination des aventuriers d'Internet ». Il s'agit de fiches pratiques visant à faire les premiers pas dans l'univers de la protection des données à l'ère d'Internet.

Il s'agit notamment d'y promouvoir des outils respectueux des libertés des utilisateurs. Cela passe tout d'abord par des outils développés dans le respect des principes du « logiciel libre ». En effet, seul le logiciel libre garantit, à court comme à long terme, la possible maîtrise par la communauté du contenu véritable d'un logiciel évitant ainsi l'introduction d'outils de surveillance (ou « portes dérobées ») ainsi qu'un monopole abusif. Malheureusement en matière de « cloud computing », où nous allons utiliser des logiciels et des serveurs appartenant à des tiers, le seul respect de ces principes ne suffit pas. Il est également important de s'interroger sur l'éthique et les pratiques des entités à qui nous confions nos données sachant que le droit diffère selon l'endroit où sont stockées les données (Europe *vs* États-Unis par exemple).

La transition vers un usage principal de logiciel libre est capitale, il n'est néanmoins pas forcément évident pour un utilisateur ayant toujours utilisé Windows ou Mac OS de faire le saut d'utiliser un système d'exploitation libre (tels qu'Ubuntu, Linux Mint, Mageia…), il reste possible de les essayer avec des clefs « Live USB ». Pour les découvrir, des « fêtes d'installation » (cf. www.agendadulibre.org) sont proposées.

Il est déjà possible de troquer son navigateur propriétaire pour utiliser « Mozilla Firefox » ou un autre navigateur libre sans perdre en fonctionnalités ni en efficacité tout en gagnant en respect de ses libertés.

En effet, le navigateur est clef dans notre usage d'Internet, c'est par lui qu'on accèdera notamment au moteur de recherche qui reste un passage très fréquent au cours de nos navigations. Le moteur de recherche est un outil central sur Internet où il est malheureusement nécessaire de faire confiance à un service tiers. Heureusement des alternatives fiables n'exploitant pas les données, très personnelles, de nos recherches existent désormais. On pourra ainsi avantageusement passer à Qwant ou à Duckduckgo pour réaliser l'essentiel de nos recherches sans tout confier à Google même si les résultats fournis sont moins personnalisés et abondants.

L'étape supplémentaire consiste en la maîtrise d'une partie des traces laissées sur Internet. En effet, les entreprises de surveillance, notamment publicitaires ont développé des techniques de traçage et profilage extrêmement performantes permettant de pister et traquer les internautes au fil de leurs navigations sur Internet en exploitant entre autres les cookies. Heureusement, il est possible de limiter, contrôler ou supprimer ces enregistrements en maîtrisant tout d'abord les paramètres de son navigateur ainsi qu'en employant des petits modules complémentaires qui vont venir bloquer un grand nombre de traceurs sur Internet tels que uBlock Origin et/ou Disconnect.

Il s'agit là de premiers pas, plutôt faciles et accessibles sans effort, qui auront vocation à être complétés par la suite notamment au niveau des services utilisés en ligne. Il convient en effet, pour limiter la surveillance en ligne, de boycotter les entreprises réalisant des exploitations abusives de nos données. On valorisera à l'inverse les entités pour qui la protection des libertés et le respect de la vie privée ne sont pas négociables. On peut ainsi remplacer avantageusement un certain nombre d'outils en ligne par des outils plus respectueux : par exemple le service « doodle » par « Framadate » géré par l'association Framasoft. Cette association s'affaire chaque jour à soutenir et proposer de nombreux outils sans exploitation corrélative des données personnelles des internautes.

Cela passe également par le recours à un hébergeur de courriel dont la sécurité et la protection des données sont une priorité, tels que Netcourrier, Posteo ou Protonmail par exemple.

Il s'agit là de limiter l'hégémonie des grands acteurs pour participer au développement d'alternatives décentralisées dont le modèle d'affaires n'est pas centré sur la surveillance et le profilage.

Il s'agit également d'apprendre plus largement à protéger ses données et la confidentialité de ses communications, là encore des outils existent tels que le « réseau TOR » qui permet d'améliorer la protection de son identité en ligne où les outils de chiffrement qui vont protéger radicalement ses données et ses communications contre des tentatives d'intrusions.

3. Au-delà de l'autodéfense, un engagement militant

Il est toutefois capital de rappeler que si chacun peut participer à développer un système plus vertueux et à limiter les possibilités de surveillance en ligne en recourant à des solutions et des pratiques plus responsables, cela ne peut suffire pour améliorer durablement la situation. Même la « non-utilisation » de ces outils n'est pas non plus une solution parfaite, en effet elle n'empêche pas nos contacts ou des tiers de transmettre des données sur nous et elle n'est pas suffisante pour lutter contre les effets sociétaux de ces pratiques. Il est nécessaire que l'action collective et politique prenne le relai pour parvenir à édicter des règles contraignantes face à ces pratiques prédatrices. Il est ainsi aussi important d'adopter des pratiques responsables que de faire valoir ses droits existants en dénonçant et en s'opposant aux abus ainsi que de participer (activement ou en les soutenant) aux organisations qui luttent au quotidien pour une meilleure protection de nos libertés à l'ère numérique.

CHAPITRE 5 – Les 10 axes de CREIS-Terminal et du CECIL

CREIS-Terminal et le CECIL s'inscrivent dans une approche du long terme, du temps long pour cerner les évolutions de l'informatisation de la société, des choix scientifiques et techniques dans le domaine du numérique. Ainsi, les deux associations, qui agissent depuis plus de 20 ans dans l'espace public (analyses, journées d'étude, colloques, publications, communiquées de presse...) et prennent également en compte l'évolution des cadres légaux, s'inscrivent dans un temps moins rapide que celui de l'évolution des pratiques et de l'économie numérique. Elles interviennent, souvent en partenariat, pour analyser les risques du numérique sur les libertés.

Les États et les grands acteurs de l'économie numérique améliorent les instruments du traitement et de l'exploitation des données, pour les premiers avec notamment un objectif de surveillance et de contrôle et pour les seconds également pour programmer des formes d'obsolescence et s'approprier les marchés. Ces évolutions mettent en exergue des tensions entre sécurité et libertés, entre contrôle et ouverture, des injonctions à se conformer à des pratiques numériques et plus généralement à des subordinations face à l'économie numérique. On observe cependant, des usages sociaux qui ouvrent d'autres voies : de subtiles déconnexions ou la recherche d'une forme d'acceptabilité du numérique sur d'autres critères (sociaux, d'ouverture…). Le Web 3.0 – conjonction du Web sémantique et de l'Internet des objets selon la définition consacrée dans le livre *Web 2.0 et au-delà* – va continuer d'augmenter le poids des enjeux du numérique et notamment les possibilités d'exploitation des données personnelles. Face à cela, l'activisme et l'ambivalence de certains usages donnent lieu à des formes diversifiées d'émancipation qui visent à s'octroyer des libertés numériques dans une approche d'autodétermination informationnelle.

Les actions et recherches de CREIS-Terminal et du CECIL s'articulent autour des 10 axes qui suivent et qui intègrent l'analyse, la diffusion et la communication indispensable au citoyen soucieux de l'exploitation de ses données personnelles par des tiers et du respect de ses libertés.

#1 Le fichage social, le profilage et la prévention des risques pour les citoyens

Le fichage bénéficie de plus d'un siècle de sophistication technique pour consigner opinions et comportements. Les objectifs politiques et administratifs ont été rejoints par ceux des acteurs économiques. Cela nous conduit, de plus en plus rapidement, à une généralisation du fichage social des citoyens dans une opposition fictive entre « sécurité et libertés ». Sous couvert de prévention des risques et de services plus performants, le citoyen est de plus en plus contrôlé, surveillé voire manipulé.

#2 Les surveillances illégitimes

Les seules données personnelles telles l'identité, l'adresse, les numéros de téléphones, etc. ne suffisent plus aux acteurs politiques et économiques. La surveillance des pratiques et usages, et donc des données comportementales sur les réseaux informatiques et sociaux, devient permanente et nourrit des instruments de plus en plus perfectionnés au service de leurs objectifs. Certains États augmentent conséquemment leurs attirails de surveillance pour chercher à rejoindre celui des acteurs privés et s'arrogent des pouvoirs de plus en plus grands pour surveiller les citoyens et les réseaux.

#3 La protection de la vie privée par le droit

Face à une évolution des techniques, les cadres réglementaires ou légaux, censés protéger les droits fondamentaux, doivent évoluer pour trouver un bon équilibre entre sécurité, droits et libertés. Les pratiques alternatives, la prise de conscience de certaines conséquences fâcheuses des fichiers et traitements, l'action des groupements ou associations qui oeuvrent pour la défense des libertés sont autant d'éléments qui peuvent accompagner l'évolution du cadre légal. Ils provoquent polémiques, débats et mobilisations dans l'espace public.

#4 Les mécanismes d'acceptation sociale de la surveillance

Une idéologie de la modernisation et de l'innovation se retrouve fréquemment dans le débat public. Elle porte sur des activités de toutes natures, par exemple, celles qui exploitent le big data. Ces évolutions sont souvent présentées comme inéluctables et comme un tsunami numérique qui emporterait tout le monde sans réelle alternative que de « nager dans le même sens ». Pourtant, certaines personnes sont déconnectées (temporaires ou peu nombreux, voire « fracturés » sans équipement ou sans culture technique) bien qu'ils soient continuellement poussés vers une acceptabilité du numérique.

#5 Les équilibres entre sécurité et libertés

Le numérique est réellement ambivalent sur ces aspects, tout en permettant un réel exercice de certaines libertés fondamentales (liberté d'expression, d'opinion, accès à la culture, protection des communications…), il offre également des possibilités d'atteintes terrifiantes à ces mêmes libertés (surveillance de masse, analyse prédictive des comportements, manipulation des populations…). Ce problème est souvent présenté dans les débats publics par une opposition caricaturale entre sécurité et libertés.

#6 Les modèles d'affaires du numérique

Les services en ligne et applications numériques, gratuits, peu onéreux ou en modèle freemium, ont gagné les faveurs des citoyens-consommateurs aux sociabilités connectées, friands de produits et services (incluant les publicités) prétendument personnalisés. Ces modèles où le client et son attention deviennent le produit ont des conséquences en termes de surveillance, mais également sur la démocratie et les équilibres des marchés.

#7 L'évolution du numérique : algorithmes, Internet des objets, réseaux neuronaux, machine learning…

Les pratiques du numérique évoluent vite et le recoupement de petites innovations peut avoir des conséquences sociétales colossales : Internet des objets, développement du poids et de l'ampleur des algorithmes, des mécanismes d'apprentissage profond, des réseaux neuronaux, informatique quantique, robotique, automatisation poussée…

#8 L'ambivalence des usages du numérique

Si le numérique a des apports indéniables, il emporte avec lui des conséquences dommageables. Au-delà des seuls risques de contrôle social et du rapport entre « liberté et sécurité », si les usages de certains outils peuvent nous faciliter la vie, ils peuvent provoquer certaines pertes de compétences comme avec l'usage systématique d'un outil de navigation GPS aboutissant à une détérioration des capacités spatiales. En ce même sens, la perte d'une forme de simplicité, une imposition technique ou le développement d'une dépendance technique doivent être questionnés sur leurs conséquences notamment sociales ou environnementales.

#9 Les résistances citoyennes

Les révélations d'Edward Snowden et d'autres lanceurs d'alertes ont eu un rôle important en faveur d'une prise de conscience de l'ampleur des possibilités de surveillance. Ces révélations couplées aux campagnes et actions de sensibilisation semblent avoir permis aux militants actifs sur ces questions d'être plus écoutés et de renforcer la prise de conscience de ces enjeux. Néanmoins, si un activisme en réseau existe sur ces questions, il ne concerne qu' une minorité face à des pratiques majoritaires qui continuent de se soumettre. Pourtant, les attentes des citoyens en termes de protection de leur vie privée et de leurs droits sont fortes. Ce paradoxe apparent doit être décrypté pour comprendre les blocages à une amélioration réelle des pratiques et des droits.

#10 Une réappropriation du numérique : libertés, émancipation et autodétermination informationnelle

La protection de la vie privée est conçue comme une liberté fondamentale dans le cadre des sociétés démocratiques, mais malheureusement rien n'est acquis tant les risques sont permanents, les technologies évolutives et les velléités économiques de certains géants du numérique fortes. Si les autorités de contrôles « Informatique et Libertés » poursuivent leurs tâches sur ces questions, elles ne doivent pas être les seuls acteurs à se sentir concernés. Les individus et leurs associations doivent s'adapter grâce à l'acquisition d'une nécessaire culture numérique au fil de la sophistication technologique des dispositifs potentiellement liberticides. Sur ce point, si certains usages sont fortement conditionnés par les choix d'acteurs économiques ou étatiques, des individus et groupes, notamment issus des milieux dits « hacker » ou du logiciel libre, se réapproprient certains outils et usages pour ne conserver que certains apports ou en développer de nouvelles pratiques non « programmées » qu'ils considèrent comme plus pertinentes au regard de leurs attentes et besoins.

Bibliographie

Contrôle social, surveillance et dispositifs numériques, dossier commun des revues Tic & société (Vol 10 N°1) et Terminal (N°118), 2016

Géopolitique d'Internet – qui gouverne le monde ?, David Fayon, Economica, 2013

Guide de survie à destination des aventuriers d'Internet, CECIL en partenariat avec la LDH, 2017

L'incertaine révolution numérique, André Vitalis, ISTE éditions, 2016

Les libertés à l'épreuve de l'informatique, Terminal 108-109, 2011

Qu'est-ce que le digital labor ?, Dominique Cardon et Antonio Casilli, INA, 2015

Société de l'information, société du contrôle, 13ème colloque international du CREIS-Terminal, Paris, 2004 www.lecreis.org/?p=188

Surveillance:// Les libertés au défi du numériques : comprendre et agir, Tristan Nitot, C&F Éditions, 2016

Ils en parlent

« *L'explosion du volume de données personnelles, combiné à une capacité croissante de les traiter et à des business models très lucratifs fait que chacun de nous est de plus en plus surveillé sans que le grand public comprenne comment tout cela est devenu possible. Ce livre blanc sera très utile pour comprendre l'étendue du problème et liste des pistes pour le résoudre. Que les auteurs en soient remerciés !* » (Tristan Nitot, @nitot)

« *La question n'est plus de savoir si l'on n'a "*rien à cacher*", mais de savoir si ceux avec qui l'on communique ont quelque chose à cacher : les services de renseignement ont en effet le droit de surveiller toute personne "*susceptible d'être en lien avec une menace*"... Un livre pour alimenter les réflexions* » (Jean-Marc Manach, @manhack)

« *Ce livre blanc de CREIS-Terminal et du CECIL livre une analyse approndie des enjeux liés à l'or transparent, la donnée. Il donne des pistes de reflexion et d'action que chaque citoyen devrait avoir present à l'esprit pour agir en tant qu'internaute éclairé… alors que Big Brother plane sur chacun d'entre nous.* » (David Fayon, @fayon)

« *Lorsque le règlement européen des données fut voté, j'eu le secret espoir que le débat à l'égard de ce sujet complexe et passionnant qu'est la donnée prendrait un tour nouveau. Si le débat a en effet pris en consistance, je dois cependant concéder qu'il n'est pas au niveau où il devrait être. Des sujets essentiels comme celui de la régulation de l'IA, l'émergence des données de santé, les enjeux de sécurité des données personnelles restent confinés dans le cercle des spécialistes et plus souvent encore des régulateurs. Puisse cet ouvrage très fouillé participer à une meilleure compréhension de ces sujets et à une dynamique renouvelée du débat à l'égard de la donnée. Il mériterait de prendre cette place et nous aussi* » (Gilles Babinet, @babgi)

Rédaction collective CREIS-Terminal et CECIL

www.lecreis.org & https://terminal.revues.org

@creis_terminal

&

www.lececil.org

@le_cecil

Novembre 2017