Ethique de la donnée.

Camille Bibard, Elise Nahuet, Jean-Charle Bournot,
Karine Goarand, Pierre Allee, Alice Pennors

8 juillet 2016

1 Introduction (pitch 1)

1.1 Ethique, morale, déontologie :

Il existe une confusion entre éthique morale, déontologie. La morale est liée à la notion de bien/mal. Celle-ci est liée à une échelle de valeurs liées à une culture, une religion. Elle est plutôt de l'ordre de l'individuel.

La déontologie, quant à elle est liée plutôt à des valeurs concernant un corps de métiers. C'est un ensemble de régles généralement corrélées avec la loi.

1.2 Ethique médicale :

L'éthique est une tentative de faire mieux. La question de l'éthique s'est surtout posée dans le domaine médical. Celui-ci a la particularité de toucher le domaine de santé et de poser des questions de vie ou de mort et appelle bien souvent une réponse rapide.

Dans cette perspective une discussion est née entre Tom Beauchamp et Jame Childress, ceux-ci étant de confessions réligieuses différentes les discussions ont donné lieu à des débats consignés dans le livre *Les principes de l'éthique biomédicale* qui évolue et en est à sa 5e version (unique traduction française). Ils créent une éthique principiste (correspondant à un certain nombre de principes), le principe d'autonomie, le principe de bienfaisance, le principe de non malfaisance.

1.2.1 Autonomie :

Le principe d'autonomie est le principe fondamental dans la prise de décision médicale. Celui-ci s'attache à entendre la parole du patient et à en

prendre compte quand cela est possible. Dans le cas d'une personne incapable d'exprimer sa décision, ou de raisonner (folie, démence), il est généralement possible de prendre une décision basée sur un supposé souhait de sa part. Dans ce cas il semble intéressant d'articuler ce principe et bouger les curseurs en fonction des autres principes.

1.2.2 Bienfaisance :

Cette idée de bienfaisance est de tenter de faire bien.

1.2.3 Non malfaisance :

L'idée de non malfaisance est elle en lien avec le fait de ne pas faire mal. Il est dans les soins médicaux d'avoir à faire des examens, ou à utiliser des traitements invasifs, qui mettent à mal le corps (chimiothérapie, par exemple). Ce principe est directement en lien avec la bienfaisance. Il est parfois difficile d'articuler les deux et il s'agit alors de faire un arbitrage sur un potentiel mieux lié au mal subit.

1.2.4 Justice / Equité :

Ce topic concerne le fait de ne pas faire de discrimination entre les individus. Il est lié au fait qu'un individu est à l'intérieur d'une communauté potentiellement inégalitaire, mais que dans la prise en charge celui-ci doit être traité avec ses spécificités.

1.3 La donnée :

Une donnée est une description élémentaire d'une réalité. C'est par exemple une observation ou une mesure.

La donnée est dépourvue de tout raisonnement, supposition, constatation, probabilité. Étant indiscutable ou indiscutée, elle sert de base à une recherche, à un examen quelconque.

Les données sont généralement le résultat d'un travail préalable sur les données brutes qui permettra de leur donner un sens et ainsi d'obtenir une information. Les données sont un ensemble de valeurs mesurables en fonction d'un étalon de référence. La référence utilisée et la manière de traiter les données (brutes) sont autant d'interprétations implicites qui peuvent biaiser l'interprétation finale (limites des sondages).

Par exemple, des données dans un graphique permettront à un être humain d'y associer un sens (une interprétation) et ainsi créer une nouvelle information.

2 Transposition vers l'éthique de la donnée :

Nous ne proposerons pas ici une transposition point à point de ces principes. Ces principes ne sont ni plus ni moins qu'une méthodologie de traitement éthique de ces questions. Nous proposons ici, plus proche du monde numérique, une éthique hacker. L'éthique hacker dans un premier temps décrite et introduite par Steven LEVY puis théorisée plus largement par Peka HIMANEN. Considérons notre présence dans ce hackathon de mettre en pratique nos intuitions sur ce qu'est le hack, nous proposons pour l'éthique de la donnée de reprendre :

L'opensource : pour qu'un logiciel soit considéré open source il faut qu'il remplisse dix critères :

- La redistribution doit être libre
- Le programme doit être distribué avec le code source, sinon il doit y avoir un moyen très médiatisé pour l'obtenir sans frais
- La licence doit autoriser les modifications et les œuvres dérivées, et doit leur permettre d'être distribuées sous les mêmes termes que la licence du logiciel original
- Pour maintenir l'intégrité du code source de l'auteur, la licence peut exiger que les œuvres dérivées portent un nom ou un numéro de version différent de ceux du logiciel original
- La licence ne doit discriminer aucune personne ou groupe de personnes
- La licence ne doit pas défendre d'utiliser le programme dans un domaine d'activité spécifique
- Les droits attachés au programme doivent s'appliquer à tous ceux à qui il est redistribué, sans obligation pour ces parties d'obtenir une licence supplémentaire
- La licence ne doit pas être spécifique à un produit
- La licence ne doit pas imposer des restrictions sur d'autres logiciels distribués avec le logiciel sous licence. Par exemple, la licence ne doit pas exiger que tous les autres programmes distribués sur le même support doivent être des logiciels open source
- La licence doit être technologiquement neutre

Selon Richard Stallman, l'un des premiers penseurs du logiciel libre : *l'open source est une méthodologie de développement ; le logiciel libre est un mouvement social.*

L'opensource permet de se passer du tiers de confiance, si chacun·ne peut consulter librement le code, si celui-ci n'a plus de secret, l'utilisateur n'a plus besoin de faire confiance.

La sécurisation des données : Les paquets transmis sur le réseau doivent être signés et chiffrés de bout en bout. Cela afin de garantir la confidentialité et l'intégrité de l'information transmise.

Permissionless : C'est la possibilité donnée à l'utilisateur de collecter, reccueillir, exploiter ses données sans demander la permission à un tiers.

P2P : Littéralement le pair à pair. L'échange de pair à pair des informations sans passer par un tiers, permet de se passer de la question du tiers de confiance. Les échanges sont recueillis sur un annuaire. C'est le principe du blockchain.

3 Etude de cas : Le cas Volswagen (pitch 2)

Le 20 septembre dernier, le scandale Volkswagen éclate. Les capteurs posés sur la voiture pour mesurer le taux d'émission de CO2 a été réglé de manière à donner des valeurs inférieures à celles émisent réellement. En dehors d'une véritable fraude, c'est aussi un cas pratique de l'intêret de l'open-source dans ce genre de système, système prévu pour participer à l'effort sur la diminution de la pollution, ayant un véritable intêret collectif.

Peu après la sortie de ce scandale, l'Electronic Frontier Foundation [1], prend position sur ce scandale en affirmant qu'avec un logiciel Open Source, il n'y aurait pas eu besoin de la prise de risque d'un lanceur d'alerte, et que la communauté ayant possibilité d'auditer le code source aurait pu faire remonter la triche.

D'autre part, il est probable que l'obligation de rendre public les codes sources aurait empêché l'entreprise de tricher, l'obligation d'ouverture du code aurait, dans ce cas là, été vertueux.

4 Etudes cas : Fraude

En 2014, la cour des comptes déplore que Pôle Emploi n'ait pas suffisamment d'informations pour évaluer l'ampleur de la fraude. Aujourd'hui Pôle Emploi teste, dans plusieurs régions françaises, la possibilité de traquer les demandeurs d'emploi en scannant les IP afin de connaître d'où se connecte le demandeur d'emploi. En mai 2015, un projet de loi est déposé par le gouvernement sur le projet de loi relatif au dialogue social, comprenant la possibilité donnée à Pôle Emploi d'avoir accès aux relevés banquaires, relevés

1. Fondation américaine qui défend les droits des citoyens dans le monde numérique

téléphoniques et données de géolocalisation détenues par certaines entreprises privées. Finalement celui-ci a retiré l'amendement concernant le relevé banquaire (le tout sans contrôle judiciaire).

4.1 Comment s'en sortir ?

Les entreprises privées sont coincées entre leur client, qui souhaiterait plus d'intimité et être moins *"fliqué"*, et la loi qui impose à l'opérateur une transmission d'informations plus fine.

Le cas Gandi est un exemple intéressant car, au moment où sort le projet de loi renseignement visant à installer des boîtes noires chez les opérateurs du réseau informatique afin de recueillir toute les informations qui passent sur le réseau. Face à ce projet d'obligation légale, Gandi monte au créneau et s'oppose frontalement à la loi. Cette opération de communication a marqué les esprits et a consolidé la confiance que les consommateurs avaient dans leur hébergement.

Il s'agirait alors pour les entreprises privées d'être vigilantes sur les projets et propositions de loi, de communiquer sur leur opposition et d'activer le lobbying contre ces lois visant à restreindre un certain nombre de libertés de leurs utilisateurs.

Glossaire

Anonymisation : Ne peut être que relatif dans l'environnement numérique. Le terme renvoie à la possibilité de dissocier les données de l'identité personnelle et des données sans identité associée.

Authentification : Vérification de l'identité déclarée.

Base de donnée : Ensemble organisé d'informations structurées (numérique, personnelle, bibliographique, etc.), accessible et interrogeable par diverses méthodes, et utilisé pour générer l'information ou les contenus dynamiques de sites Internet populaires.

Cryptologie : Conversion de données numériques en texte illisible. Une clé est nécessaire pour reconvertir des données chiffrées dans un format lisible.

Cybersurveillance : Ensemble des règles régissant la mise en place de moyens électroniques de surveillance des personnes dans une organisation (entreprises, collectivités publiques), qu'il s'agisse des personnels ou du public reçu.

Data mining : Traitement mathématique qui permet d'analyser et d'interpréter de gros volumes de données, afin d'en extraire des connaissances (identifier des tendances, rassembler les éléments similaires, formuler des hypothèses).

Données personnelles : Les données personnelles (ou nominatives) correspondent généralement aux nom, prénom, adresse électronique, numéro de téléphone, date de naissance, etc., qu'un individu peut transmettre par courrier électronique, inscrire sur un formulaire en ligne ou sur un site Web.

Gestion de l'identité : Outils et méthodes permettant de créer et de maintenir une présence numérique, avec les droits et restrictions qui l'accompagnent.

Identification : Déclaration d'identité, qui peut nécessiter une authentification (ou une vérification).

Identité numérique : Représentation numérique d'un utilisateur individuel dans un environnement réseau.

Intelligence artificielle : Science des « machines intelligentes » correspondant aux efforts tendant à produire des équivalents informatiques de l'intelligence humaine.

Métadonnées : Données permettant d'accéder à des ressources.

Opt in / opt out : Possibilité donnée au consommateur de laisser utiliser ses données personnelles ou de ne pas l'autoriser.

Peer to peer ou p2p : Protocole de mise en réseau des ordinateurs qui maximise le partage et la diffusion des fichiers numériques.

Platform for Privacy Preferences ou p3p : Système de protection des données personnelles.

Privacy ou vie privée : Catégorie déterminée culturellement. Dans l'environnement numérique, elle désigne le droit qu'a l'utilisateur de gérer ses informations confidentielles et de limiter leur diffusion et leur réutilisation par des tiers.

Redocumentarisation : Processus consistant à traiter à nouveau un document ou une collection de documents en réarticulant les contenus selon l'usage ou l'interprétation de l'utilisateur.

Sécurité : Protection des données numériques et garantie de leur intégrité par contrôle d'accès, chiffrement ou autres méthodes du même ordre.

Tiers de confiance : Un tiers de confiance est un organisme habilité à mettre en œuvre des signatures électroniques reposant sur des architectures d'infrastructure à clés publiques.

Source : Glossaire *, Hermès, La Revue 1/2009 (n 53) , p. 185-186*

Références

[1] ~~~~= :) - datalove.me - (:=~~~~.

[2] Big data : le respect de la vie privée doit être assuré avant que la révolution commence.

[3] Code de bonne conduite – GRDF, le distributeur de gaz naturel.

[4] Descripteur : TRAITEMENT AUTOMATISE DE DONNEES a CARACTERE PERSONNEL (TADCP) | legifrance.

[5] Ethique big data.

[6] Glossaire. (53) :185–186.

[7] Insee - définitions, méthodes et qualité - IRIS.

[8] Les données numériques : il faut concilier usage commercial et usage citoyen ! [vidéos].

[9] Numérique. Page Version ID : 124466423.

[10] Pourquoi contrôler mes données ?

[11] Smart meter security : 43 areas of concerns (by ethical hackers).

[12] white hat. Page Version ID : 124015393.

[13] Christophe Aguiton and Dominique Cardon. Web participatif et innovation collective. (50) :75–82.

[14] Michel Arnaud. Authentification, identification et tiers de confiance. (53) :127–136.

[15] Dominique Cardon. Regarder les données. (49) :138–142.

[16] Emmanuel Kessous and Bénédicte Rey. Économie numérique et vie privée. (53) :49–54.