Problèmes éthiques dans l'utilisation des mégadonnées

Nicolae Sfetcu

23.02.2020

Sfetcu, Nicolae, « Problèmes éthiques dans l'utilisation des mégadonnées », SetThings (23 février 2020), MultiMedia Publishing (ed.), URL = https://www.telework.ro/fr/problemes-ethiques-dans-lutilisation-des-megadonnees/

Email: nicolae@sfetcu.com

Une traduction partielle de :
Sfetcu, Nicolae, « Etica Big Data în cercetare », SetThings (6 iulie 2019), DOI: 10.13140/RG.2.2.27629.33761, MultiMedia Publishing (ed.), ISBN: 978-606-033-228-2, URL = https://www.telework.ro/ro/e-books/etica-big-data-in-cercetare/

Prise de conscience

La prise de conscience du type de données fournies lors d'une inscription en ligne (pour la création d'un compte ou d'un abonnement, par exemple) est un fait rare, d'autant plus qu'il existe la possibilité d'utiliser une identité numérique existante (profil Facebook par exemple) au lieu d'une liste séparée pour un accès plus rapide. De telles situations créent une opacité concernant les données partagées entre le fournisseur d'identité et le service utilisé. (European Economic and Social Committee 2017)

Consentement

Afin d'utiliser les données personnelles d'une personne, son consentement informé et explicite est requis concernant qui, quand, comment et dans quel but elles sont utilisées. Lorsque les données doivent être partagées, ces utilisations doivent être portées à la connaissance de la personne. Il devrait toujours être possible de retirer son consentement pour une utilisation future.

Dans l'analyse des mégadonnées, très peu de choses peuvent être connues sur les utilisations futures prévues des données, ainsi que sur les avantages et les risques impliqués. Ici, il existe des procédures de consentement « large » et « générique » pour partager des données génomiques, par exemple, et à des fins différentes. Même lorsqu'il est fait correctement, il existe des défis pratiques spécifiques : obtenir un consentement éclairé peut être impossible ou très coûteux, et la validité du consentement est contestée lorsque l'accord est nécessaire pour accéder à un service.

Contrôle

Dans le monde d'aujourd'hui, les données personnelles peuvent être échangées comme n'importe quelle devise dans la mise en œuvre des mégadonnées. Il y a différentes opinions sur la mesure dans laquelle cette situation est éthique, y compris sur qui participer au profit obtenu de ces transactions.

Dans le modèle d'échange des données personnelles, la transmission des données personnelles est un cadre qui donne aux gens la possibilité de contrôler leur identité numérique et de créer des accords de partage des données granulaires.

L'idée des données ouvertes, centrée sur l'argument selon lequel les données devraient être disponibles gratuitement, est en train d'émerger. La volonté de partager des données varie selon la personne.

Dans le cas des enfants, les parents ou tuteurs ont la responsabilité de leurs données, qui ne peuvent pas être échangées contre des avantages financiers.

Au niveau national, un gouvernement est souverain sur les données générées et collectées. Le 26 octobre 2001, la Loi patriotique est entrée en vigueur aux États-Unis, et le 25 mai 2018, le Règlement Général de la Protection des Données 2016/679 (RGPD) au niveau de l'Union européenne, pour les questions liées à la protection des données personnelles.

Dans les mégadonnées, la relation homme-données est asymétrique, basée sur le contrôle des données. Le « droit à l'oubli », adopté au niveau de l'UE, est l'un des éléments fondamentaux du contrôle d'un individu sur ses données personnelles.

Transparence

Les algorithmes utilisés dans les mégadonnées peuvent conduire à des biais qui affectent systématiquement les droits de l'individu. Par conséquent, la conception de l'algorithme doit être transparente et inclusive.

La gouvernance anticipative implique une analyse prédictive basée sur les mégadonnées pour évaluer les comportements potentiels, avec des implications éthiques qui peuvent encourager les préjugés et la discrimination.

Une personne qui accepte l'inclusion de ses données personnelles dans les mégadonnées a le droit de savoir pourquoi les données sont collectées, comment elles seront utilisées, combien de temps elles seront stockées et comment elles pourront être modifiées.

Confiance

La confiance dans les systèmes des mégadonnées est liée à l'interdépendance avec la confidentialité et la sensibilisation. Jusqu'à présent, la confiance a été considérée d'un point de vue strictement technologique. On espère que des architectures matérielles et logicielles seront

développées qui pourraient accroître la confiance entre les êtres humains et les objets, et donc une plus grande acceptation de l'utilisation des données personnelles.

Propriété

Une question fondamentale dans l'éthique de la recherche sur les mégadonnées est de savoir à qui appartiennent les données ? Cela implique le sujet des droits et obligations de propriété. En droit européen, le RGPD indique que les gens détiennent leurs propres données personnelles.

La somme des données personnelles d'un individu forme une identité numérique.

La protection des droits moraux (le droit d'être identifié comme source de données et de les contrôler) d'un individu repose sur l'opinion que les données personnelles sont une expression directe de sa personnalité et ne peuvent être transférées qu'à une autre personne, éventuellement, par succession au décès de l'individu.

La propriété implique l'exclusivité, c'est-à-dire la restriction implicite d'autrui concernant l'accès à la propriété. Une propriété efficace des données personnelles implique la portabilité, la possibilité d'utiliser des alternatives sans perdre de données. La normalisation aiderait également à nettoyer les données personnelles.

En fait, à l'heure actuelle, les données sont détenues par le propriétaire des capteurs, celui qui effectue l'enregistrement ou l'entité propriétaire du capteur.

Dans l'UE, la possibilité que les données des citoyens de l'UE soient stockées en dehors de ce que l'on appelle l'« Euro cloud » a été progressivement réduite, mais le problème des données déjà stockées et traitées ailleurs n'a pas été résolu et « ne résout pas le dilemme éthique de la façon dont la propriété des données est définie philosophiquement, avant de passer à une approche plus juridique et politique. » (European Economic and Social Committee 2017)

Surveillance et sécurité

De plus en plus de sources de données sont disponibles à l'aide de technologies avancées telles que la vidéosurveillance, le GPS, les appareils mobiles, les cartes de crédit, les distributeurs automatiques de billets. De plus, la surveillance active est une méthode de collecte de données, mais en même temps limitant les libertés des citoyens. Une telle surveillance permanente détermine l'augmentation du stress des personnes et crée leur tendance à se comporter d'une certaine manière conforme aux normes attendues.

Identité numérique

L'identité numérique a l'avantage d'un accès rapide au contenu en ligne et aux services associés. L'utilisation de l'identité numérique a le potentiel de générer une discrimination fondée sur la représentation d'une personne en fonction de ses données en ligne, qui peut souvent ne pas correspondre à la situation réelle, dans un processus appelé « dictature des données » dans lequel « nous ne sommes plus jugés sur la base d'actions mais sur la base de ce qui indique toutes les données nous concernant comme nos actions probables », (Norwegian Data Protection Authority 2013) l'interaction personnelle n'est pas placée dans un plan secondaire.

Réalité ajustée

Toute interaction que nous avons avec Internet implique la possibilité de stocker nos données personnelles. Le traitement et l'analyse de ces données déterminent les résultats personnalisés qui apparaissent ultérieurement sur Internet, à travers les résultats de nos recherches, l'affichage des produits dans les boutiques en ligne, l'affichage des publicités, etc. Cela génère une version plus étroite et plus personnalisée de l'expérience en ligne précédente d'un utilisateur (ce que l'on appelle le « ballon filtre ». (Pariser 2011) Un avantage est que l'utilisateur trouvera rapidement ce qu'il recherche habituellement, mais l'exclusion de certains aspects, perspectives et

idées peut conduire à une restriction de la créativité et au développement d'une attitude tolérante par isolement politique et social des autres aspects, par le manque de vues pluralistes. (Crawford, Gray, and Miltner 2014)

De-anonymisation

La désidentification implique la suppression ou la dissimulation d'éléments qui pourraient immédiatement identifier une personne ou une organisation. La législation de différents pays sur la protection des données définit différents traitements pour les données identifiables. L'identifiabilité est de plus en plus considérée comme un continuum et non comme un aspect binaire. Les risques de divulgation augmentent simultanément avec le nombre de variables, les sources de données et la puissance de l'analyse des données. Les risques de divulgation peuvent être atténués mais non éliminés. La désidentification reste un outil essentiel pour garantir une utilisation sûre des données. (UK Data Service 2017)

Des informations parfaitement anonymes prises séparément peuvent être combinées avec d'autres données pour identifier de manière unique une personne avec différents degrés de certitude. Le profilage peut devenir un outil puissant, suscitant des inquiétudes quant à la mesure dans laquelle l'intrusion dans la vie d'un individu est autorisée, la possibilité d'assurer la sécurité et la supervision.

Inégalité numérique

Les avantages d'une grande taille de données sont évidents, mais certains pensent également que l'accumulation de données à grande échelle présente des risques spécifiques. De ce fait, peu d'entités ont accès, via l'infrastructure et les compétences, aux systèmes des mégadonnées. Dans ce contexte, les coûts et les compétences nécessaires à l'accès conduisent à des inégalités numériques spécifiques traitées par l'éthique.

Confidentialité

Dans les transactions de données, il est très important de garantir la confidentialité :

> « Nul ne sera l'objet d'immixtions arbitraires dans sa vie privée, sa famille, son domicile ou sa correspondance, ni d'atteintes à son honneur et à sa réputation. Chacun a droit à la protection de la loi contre de telles ingérences ou attaques. » - *Déclaration des Droits de l'homme des Nations Unies*, Article 12.

Dans de nombreux pays, la surveillance publique des données par le gouvernement pour observer les citoyens nécessite une autorisation explicite par le biais d'un processus judiciaire approprié. La confidentialité ne consiste pas à garder des secrets, mais à choisir, à respecter les droits de l'homme et la liberté.

Souvent, la confidentialité est considérée à tort comme un choix binaire entre l'isolement et le progrès scientifique. La protection de l'identité dans les données est technologiquement possible, par exemple en utilisant un cryptage homomorphe et une conception algorithmique.

La confidentialité en tant que limitation de l'utilisation des données peut également être considérée comme contraire à l'éthique, (Kostkova et al. 2016) en particulier dans les soins de santé, mais il convient de garder à l'esprit qu'il est possible d'extraire la valeur des données sans compromettre la confidentialité.

La confidentialité est reconnue comme un droit de l'homme par de nombreuses réglementations nationales et internationales. La confidentialité dans la recherche est obtenue grâce à une combinaison d'approches : limiter les données collectées, les anonymiser ; et réglementer l'accès aux données. Dans le cas de la recherche des mégadonnées, des problèmes spécifiques se posent : l'ambiguïté entre les termes « vie privée » et « confidentialité »; la déclaration des espaces sociaux publics ou privés; l'ignorance des risques de confidentialité par les utilisateurs; la distinction floue entre les utilisations publiques et privées. Il existe actuellement des

différends quant à savoir si la science des données doit être classée comme une recherche sur des sujets humains, et donc non soumis aux règles habituelles de confidentialité.

Recherche des mégadonnées

A travers les nouveaux concepts de « dommages algorithmiques », « analyse prédictive », etc., les algorithmes actuellement utilisés dans les opérations avec les mégadonnées dépassent la vision traditionnelle de la confidentialité. Selon le Conseil national pour la science et la technologie des États-Unis,

> « Les « algorithmes analytiques » sont des algorithmes de priorisation, de classification, de filtrage et de prédiction. Leur utilisation peut créer des problèmes de confidentialité lorsque les informations utilisées par les algorithmes sont inadéquates ou inexactes, lorsque des décisions incorrectes se produisent, lorsqu'il n'existe aucun moyen d'appel raisonnable, lorsque l'autonomie d'un individu est directement liée au résultat algorithmique ou lors de l'utilisation d'algorithmes prédictifs encourage d'autres atteintes à la vie privée. » (NSTC (National Science and Technology Council) 2016, 18)

La recherche sur les mégadonnées est ce que l'éthicien James Moor qualifierait de « marché conceptuel » en raison de « l'incapacité de conceptualiser correctement les valeurs éthiques et les dilemmes du jeu dans un nouveau contexte technologique ». (Buchanan and Zimmer 2018) Dans cette situation, la confidentialité est assurée par une combinaison de différentes tactiques et pratiques (environnements contrôlés ou anonymes, limitation des informations personnelles, anonymisation des données, restrictions d'accès, sécurité des données, etc.). En général, tous les concepts associés deviennent confus dans le cas des mégadonnées. Ainsi, les publications sociales sont considérées comme publiques sur les réseaux sociaux en cas de paramétrage approprié. Mais les réseaux sociaux sont des environnements complexes d'interactions sociotechniques où les utilisateurs ne comprennent pas toujours la fonctionnalité des paramètres et des conditions d'utilisation. Ainsi, il existe une incertitude quant aux intentions et aux attentes des utilisateurs, et

ces lacunes conceptuelles dans le contexte de la recherche sur les mégadonnées conduisent à des incertitudes quant à la nécessité d'un consentement informé.

Bibliographie

Buchanan, Elizabeth A., and Michael Zimmer. 2018. "Internet Research Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2018/entries/ethics-internet-research/.

Crawford, Kate, Mary L. Gray, and Kate Miltner. 2014. "Big Data| Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction." *International Journal of Communication* 8 (0): 10. https://ijoc.org/index.php/ijoc/article/view/2167.

European Economic and Social Committee. 2017. "The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context." European Economic and Social Committee. February 22, 2017. https://www.eesc.europa.eu/en/our-work/publications-other-work/publications/ethics-big-data.

Kostkova, Patty, Helen Brewer, Simon de Lusignan, Edward Fottrell, Ben Goldacre, Graham Hart, Phil Koczan, et al. 2016. "Who Owns the Data? Open Data for Healthcare." *Frontiers in Public Health* 4. https://doi.org/10.3389/fpubh.2016.00007.

Norwegian Data Protection Authority. 2013. "Big Data – Privacy Principles under Pressure." https://www.datatilsynet.no/globalassets/global/english/big-data-engelsk-web.pdf.

NSTC (National Science and Technology Council). 2016. "National Privacy Research Strategy." https://obamawhitehouse.archives.gov/sites/default/files/nprs_nstc_review_final.pdf.

Pariser, Eli. 2011. *The Filter Bubble: What The Internet Is Hiding From You*. Penguin Books Limited.

UK Data Service. 2017. "Big Data and Data Sharing: Ethical Issues." https://www.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf.