INTELLIGENCE ARTIFICIELLE ET PROTECTION DE LA VIE PRIVÉE : comment identifier et résoudre les problèmes

Yves-Alexandre de Montjoye [a], Ali Farzanehfar [a], Julien Hendrickx [b], Luc Rocher [b]
[a] Imperial College London, Data Science Institute and Dept. of Computing
[b] Université catholique de Louvain, ICTEAM Institute

We are unlikely to see any 'general AI'—machines that could learn the way we do and successfully perform a large range of task—anytime soon

Yves-Alexandre de Montjoye est professeur à l'Imperial College London, où il dirige le Computational Privacy Group et le directeur de l'Algorithmic Society Lab.

Julien M. Hendrickx est professeur d'ingénierie mathématique à l'Université catholique de Louvain (Ecole Polytechnique de Louvain) depuis 2010 et directeur du département d'ingénierie mathématique.

Luc Rocher est doctorant au département d'ingénierie mathématique à l'Université de Louvain et chercheur au F.R.S.-FNRS.

Ali Farzanehfar est doctorant à l'Imperial College London (Computational Privacy Group).

MOTS CLÉS

- VIE PRIVÉE
- ANONYMISATION DES DONNÉES
- PSEUDONYMISATION
- DÉSIDENTIFICATION
- VASTES JEUX DE DONNÉES
- IDENTITÉ
- K-ANONYMAT
- UNICITÉ

L'intelligence artificielle (IA) a le potentiel de modifier en profondeur nos manières de travailler, de vivre et d'interagir. Il n'existe pourtant pas d'IA générale et la précision des modèles d'apprentissage automatique actuels dépend largement des données avec lesquelles ils ont été formés. Dans les décennies à venir, le développement de l'IA sera conditionné par l'accès à des ensembles de données médicales et comportementales plus vastes et plus riches. Nous avons aujourd'hui des preuves solides du fait que l'outil que nous avons utilisé par le passé pour établir un équilibre entre l'utilisation des données agrégées et la protection de la vie privée des personnes, la désidentification, ne fonctionne pas avec de vastes ensembles de données. Le développement et le déploiement des « technologies renforçant la protection de la vie privée » (PET), permettant aux contrôleurs de données de rendre les données disponibles de manière sûre et transparente, seront essentiels pour que puisse s'exprimer tout le potentiel de l'IA.

INTRODUCTION

Un monde qui était une simple vision il y a quelques années devient aujourd'hui réalité. Les voitures apprennent à conduire seules, l'analyse prédictive transforme les soins médicaux et la recherche, et le monde de la finance utilise des algorithmes pour identifier les fraudeurs. Les sociétés de carte bancaire collectent et surveillent déjà toutes les transactions[1] pour détecter la fraude en temps réel. Les voitures autonomes vont transformer les espaces de stationnement, les habitudes liées au trajet domicile-travail, et devraient fortement réduire les accidents de la circulation. En médecine, on commence à utiliser des algorithmes pour identifier les molécules à fort impact dans le développement pharmaceutique, et pour accélérer le diagnostic du cancer de la peau avec une exactitude au moins égale à celle des experts en dermatologie[2]. De quelle manière ces changements vont-ils affecter nos sociétés ? Un récent rapport McKinsey indique que 45 % de

1 Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017), Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542* (7639); 115-118.

2 McKenna, F. (2017), 11 Companies That Teach Machines To Detect Fraud, Frankonfraud.com, http://*frankonfraud.com*/fraud-products/10-companies-that-use-machine-learning-to-solve-financial-fraud.

toutes les activités professionnelles pourraient prochainement être automatisées avec l'intelligence artificielle (IA) [3]*. L'intelligence artificielle change notre économie à une vitesse sans précédent, avec un impact radical sur notre façon de vivre, d'interagir, de produire des biens et des services. Développer des solutions permettant aux algorithmes d'IA d'apprendre à partir de jeux de données à grande échelle, souvent sensibles, tout en préservant la vie privée des personnes, est l'un des grands défis auxquels nous sommes aujourd'hui confrontés.*

3 McKinsey Global Institute (2016), The age of analytics: Competing in a data-driven world, *McKinsey*.

Cependant, malgré ce que la presse populaire voudrait nous faire croire, l'IA ressemble très peu à l'intelligence humaine. Les experts de sa branche la plus populaire, l'apprentissage automatique (machine learning en anglais), ont passé des décennies à entraîner un grand écosystème de modèles statistiques, conçus pour des tâches spécifiques comme l'inférence des émotions humaines à partir de messages textuels ou la détection et la classification des lésions cutanées sur photographie. Ces modèles sont très spécifiques à leur domaine, et contrairement aux humains, ils ont rarement la capacité à transférer les connaissances d'un domaine à l'autre. Un modèle entraîné sur les messages de téléphone portable ne fera probablement pas mieux sur des messages Facebook que le modèle entraîné directement sur ce lot de données. Nous avons peu de chances de voir dans les prochaines décennies un type quelconque de ce que les experts appellent l'IA générale : des machines qui réussiraient toute sorte de tâches intellectuelles[4].

En revanche, un récent progrès de l'IA concerne autant la mise au point de nouveaux algorithmes que l'accès à une grande quantité de données. Les techniques telles que l'apprentissage profond (deep learning en anglais), c'est-à-dire au moyen de réseaux neuronaux profonds, ne sont pas nouvelles. Le concept des réseaux neuronaux date des années 1950, et un grand nombre des révolutions algorithmiques ont eu lieu dans les années 1980 et 1990. Mais ce sont les gigantesques stocks de données qui existent aujourd'hui et l'énorme puissance de calcul dont nous disposons[5] qui nous permettent d'enfin exploiter tout le potentiel de cette technologie. Des corpus de millions d'enregistrements de la parole, d'images haute résolution et de métadonnées humaines ont conduit à la révolution de l'IA.

Par exemple, pour la première fois, la collecte de données à cette échelle a été cruciale pour modéliser finement les opinions politiques, ce qui a constitué un outil déterminant pour les dernières élections aux États-Unis. En 2017, la société Cambridge Analytica[6] a pu dresser le profil de plus de 220 millions de citoyens américains et construire un portrait psychométrique intime des électeurs, afin de trouver et de cibler leurs déclencheurs émotionnels. Elle a utilisé les Likes sur Facebook, une masse de données librement accessibles, pour identifier le genre, l'orientation sexuelle, les croyances politiques et la personnalité des

4 Etzioni, O. (2016), No, the Experts Don't Think Superintelligent AI is a Threat to Humanity, *MIT Technology Review*.

5 Roger Parloff (2016), Why Deep Learning is Suddenly Changing Your Life, *Fortune*, http://fortune.com/ai-artificial-intelligence-deep-machine-learning.

6 Green, J. and Issenberg, S. (2017), Trump's Data Team Saw a Different America—and They Were Right, *Bloomberg*, bloom.bg/2eEWfeO.

"DÉVELOPPER DES SOLUTIONS PERMETTANT AUX ALGORITHMES D'IA D'APPRENDRE À PARTIR DE JEUX DE DONNÉES À GRANDE ÉCHELLE, SOUVENT SENSIBLES, TOUT EN PRÉSERVANT LA VIE PRIVÉE DES PERSONNES, EST L'UN DES GRANDS DÉFIS AUXQUELS NOUS SOMMES AUJOURD'HUI CONFRONTÉS."

individus, exploitant ainsi le pouvoir que donne l'accès en temps réel à des données personnelles par millions[7].

Toutefois, alors que les données alimentent des progrès fantastiques en IA, leur utilisation soulève des questions profondes sur la vie privée et la propriété des données. La plupart de ces données sont produites par des individus dans leur quotidien : des métadonnées téléphoniques sont créées par les appels et l'envoi de messages, les enregistrements médicaux par la surveillance des patients suivis en hôpital ou en clinique, les statistiques d'encombrement routier par les traces GPS des conducteurs. Ces données contiennent des informations détaillées et souvent sensibles sur le comportement des personnes, leur état de santé, leurs habitudes de déplacement et leur style de vie, qui peuvent être utilisées pour glaner d'autres connaissances comme les croyances religieuses et l'adhésion à un syndicat.

L'IA possède véritablement un potentiel immense, mais la construction d'une meilleure IA nécessite des bases de données à grande échelle ; leur constitution et l'accès à ces informations personnelles et privées exigent des solutions qui protègent la vie privée. C'est un des grands défis auxquels nous faisons face actuellement.

Historiquement, l'équilibre entre l'utilisation des données et la protection de la vie privée reposait sur l'anonymisation des données. Du point de vue pratique et législatif, aux États-Unis, en Europe et dans le monde, l'anonymisation était le moyen de permettre l'utilisation des données en garantissant le respect de la vie privée. En effet, si les données ne peuvent pas être associées à l'individu dont elles proviennent, alors les informations qu'elles contiennent ne peuvent pas lui nuire.

En pratique, les jeux de données sont anonymisés par une combinaison de pseudonymisation et de dé-identification. La pseudonymisation consiste à remplacer un identifiant clair, comme le nom, par un pseudonyme. Historiquement, c'est la première ligne de défense. Mais la simple utilisation de pseudonymes pour la protection de la vie privée a été remise en question à la fin des années 1990, lorsque la Group

7 Thompson-Fields, D. (2017), Did artificial intelligence influence Brexit and Trump win?, *Access AI*, http://access-ai.com/news/21/artificial-intelligence-influence-brexit-trump-win.

Insurance Commission (GIC) du Massachusetts a diffusé des données « anonymisées » concernant toutes les visites hospitalières des agents de l'État. Le gouverneur du Massachusetts de l'époque, William Weld, a garanti que la GIC avait protégé la vie privée des patients en effaçant les identifiants. Mais alors, au moyen de la liste électorale de la ville de Cambridge, publique, l'étudiante au MIT Latanya Sweeney a pu révéler le dossier médical du gouverneur à partir de sa date de naissance, de son sexe et de son code postal. Elle a ensuite démontré que 87 % des Étatsuniens pourraient être identifiés de manière unique par la combinaison de ces trois données[8]. Cet incident a mis en évidence les limites de l'anonymat obtenu par la simple utilisation de pseudonymes.

La deuxième ligne de défense, la dé-identification, a ensuite été développée pour prévenir la réidentification (restauration du lien entre les données et les personnes réelles), ce qui a permis d'utiliser à nouveau les données tout en protégeant la vie privée. La première solution contre la réidentification, le k-anonymat[9], a été proposée juste après l'attaque de Latanya Sweeney. Une base de données est dite k-anonyme si toute combinaison d'informations (par exemple l'année de naissance, le sexe et le code postal) peut être trouvée chez au moins k individus. Il est alors impossible d'identifier une personne spécifique dans la base de données, car toute information collectée conduit à un groupe d'au moins k individus. De nombreux algorithmes basés sur les principes de généralisation et de suppression ont été proposés pour k-anonymiser les bases de données. Un attribut peut être généralisé en dégradant la résolution de son information (par exemple en remplaçant l'âge exact par une tranche d'âge). La valeur d'un attribut peut être supprimée parce qu'elle donne trop d'informations sur l'identité d'une personne. Des extensions du k-anonymat ont été développées, comme la l-diversité[10] et la t-proximité[11], qui protègent contre des attaques par inférence plus complexes.

Cette combinaison de pseudonymisation et de dé-identification a très bien fonctionné depuis que le k-anonymat a été proposé en 1995. Cependant, les jeux de données modernes, et en particulier ceux utilisés par l'IA, sont très différents de ceux de la moitié des années 1990. Aujourd'hui, la plupart des données possèdent un grand nombre de dimensions et sont même longitudinales : pour chaque individu, il existe des dizaines, des centaines ou des milliers d'éléments d'information différents. Par exemple, chaque personne est associée à des milliers de points visités via les métadonnées de téléphonie mobile, à des milliers de clics et de sites via les données de navigation web, et les jeux de données génétiques contiennent souvent plus d'informations par personne que de personnes.

Ce changement fondamental dans nos données se traduit dans les capacités de nos techniques d'anonymisation à protéger la vie privée. En 2013, le concept d'unicité a été introduit pour évaluer l'efficacité de l'anonymisation dans les jeux de données modernes possédant un grand nombre de dimensions. L'unicité quantifie le nombre de personnes identifiées de manière unique au moyen d'un nombre p d'informations aléatoirement choisies et auxquelles l'adversaire pourrait avoir accès. Une étude basée sur les métadonnées de téléphonie mobile d'1,5 million de personnes montre que 4 points (approximatifs dans le temps et l'espace) suffisent à identifier de manière unique 95 % des personnes[12]. Cela signifie qu'il suffit de connaître où et quand un individu interagit avec le réseau de téléphonie mobile, seulement 4 fois sur 15 mois, en moyenne, pour le réidentifier dans un jeu de données téléphoniques simplement anonymisé, et donc révéler tout l'historique de ses trajets.

Ces résultats, initialement obtenus dans un pays européen, ont été reproduits avec un jeu de données d'un million de personnes en Amérique latine[13] et avec les données d'un demi-million de personnes dans un autre pays non nommé[14]. En 2015, la même méthodologie a été appliquée à des données de transactions bancaires. Cette étude publiée dans le magazine Science conclut que 4 points (date et lieu d'un achat) suffisent à identifier de manière unique 90 % des personnes parmi un million d'utilisateurs de carte bancaire[15].

Mais est-il possible de brouiller à nouveau les pistes ? Peut-on à nouveau généraliser les données, ou leur ajouter du bruit ? Malheureusement, la réponse est non pour les données de téléphonie mobile et pour celles des cartes bancaires. Les études ci-dessus montrent que l'ajout de bruit ou la dégradation de la résolution spatiale ou temporelle des données provoque une augmentation marginale de la difficulté d'identification. En effet, même dans un jeu de métadonnées téléphoniques de très basse résolution (dégradées d'un facteur 15 pour la date et pour le lieu), 10 points suffisent à trouver une personne dans 50 % des cas[16]. Il peut être surprenant de constater que, dans l'étude des données bancaires, la connaissance d'à peine 10 visites d'un individu dans l'un des 350 magasins sur deux semaines permet une réidentification correcte dans 80 % des cas[17]. Considérés ensemble, ces résultats suggèrent que d'autres grands jeux de données possédant un grand nombre de dimensions et utilisés par l'IA présentent probablement un haut niveau d'unicité, facilitant la réidentification.

Aujourd'hui, non seulement les jeux de données modernes sont extrêmement difficiles à anonymiser, mais leur richesse accentue leur sensibilité. Autrefois, un aperçu des données suffisait à évaluer le préjudice potentiel d'une réidentification (par exemple selon qu'il s'agissait d'un dossier médical ou de données assez anodines). Parfois,

8 Sweeney, L., 2000. Simple demographics often identify people uniquely. *Health (San Francisco), 671*, pp.1-34.

9 Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557-570.

10 Machanavajjhala, A., Gehrke, J., Kifer, D., & Venkitasubramaniam, M. (2006, April). l-diversity: Privacy beyond k-anonymity. In *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on* (pp. 24-24). IEEE.

11 Li, N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on* (pp. 106-115). IEEE.

12 De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, *3*, 1376.

13 U.N. Global Pulse. Mapping the risk-utility landscape of mobile phone data for sustainable development & humanitarian action, 2015.

14 Yi Song, Daniel Dahlmeier, and Stephane Bressan. Not so unique in the crowd:a simple and effective algorithm for anonymizing location data. ACM PIR, 2014.

15 De Montjoye, Y. A., Radaelli, L., & Singh, V. K. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science, 347*(6221), 536-539.

16 De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, *3*, 1376.

17 De Montjoye, Y. A., Radaelli, L., & Singh, V. K. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science, 347*(6221), 536-539.

l'information sensible pouvait même être retirée pour que les données deviennent non sensibles (par exemple le fait qu'une personne ait vu certains films). Cela est devenu impossible avec les jeux de données modernes possédant un grand nombre de dimensions. La raison est que, pour évaluer la sensibilité des données, il faut prendre en considération non seulement ce qui est visible directement, mais aussi ce qu'un algorithme pourrait inférer à partir des données, aujourd'hui ou demain. Par exemple, il a été démontré que les traits de personnalité[18], les données démographiques[19], la situation socio-économique[20] [21] ou même le taux de remboursement des emprunts[22] peuvent tous être prédits à partir de données de téléphonie mobile, apparemment anodines. Ce « risque de l'inférence » dans les grands volumes de données rend l'évaluation complète des risques incroyablement difficile, certains diraient impossible.

Si l'anonymisation est défaillante sur les grands jeux de données possédant un grand nombre de dimensions, quelle solution permet d'avancer ? Dans les années 1990, lorsque le transfert de données était très coûteux, l'anonymisation était la seule solution pour diffuser les données à un coût minimal. Ce n'est plus le cas, et de nouvelles solutions existent. Les contrôleurs de données peuvent facilement autoriser un accès distant à une partie des données, pour analyse, au lieu de publier les enregistrements bruts : on rapproche les algorithmes des données et non l'inverse.

Le projet OPAL (Open Algorithms)[23] par exemple, récemment créé, est basé sur ce principe. Mené sur le plan technique par le Computational Privacy Lab de l'Imperial College London, en partenariat avec le MIT et les sociétés de télécommunication Telefonica et Orange, Data-Pop Alliance, MIT, WEF, OPAL permet à des tiers d'utiliser en toute sécurité les données de géolocalisation collectées par les sociétés de télécommunication, selon un modèle de questions-réponses. En résumé, la plateforme autorise des tiers tels que des chercheurs et des médecins à soumettre des requêtes (par exemple combien de personnes se sont-elles déplacées du point A au point B dans une journée donnée). À partir des sources de données disponibles dans cet environnement fiable, la plateforme valide le code, calcule les résultats du modèle et renvoie uniquement des résultats agrégés, ce qui garantit qu'aucun individu ne peut être identifié ou profilé. La totalité des interactions est enregistrée dans un registre non falsifiable, afin de permettre les audits et les vérifications. La combinaison des mécanismes de contrôle d'accès, de bac à sable pour le code et de l'agrégation nous permet de garantir que les données sont utilisées de manière anonyme, y compris par les algorithmes d'apprentissage automatique et même si les données sont seulement pseudonymisées.

Il existe de nombreux cas d'utilisation de tels outils de partage des données. Dans les pays en voie de développement où les statistiques nationales détaillées sont soit périmées, soit totalement inexistantes, les données de téléphonie mobile peuvent contribuer à guider les politiques publiques, en fournissant des statistiques à jour et fiables pour l'allocation de ressources en temps réel en cas de catastrophe[24], pour chiffrer avec exactitude la densité de population[25] en temps réel, ou pour construire des modèles de propagation d'une épidémie. C'est dans cet esprit que le lancement d'OPAL aura lieu en Colombie et au Sénégal, en collaboration étroite avec les gouvernements locaux et les bureaux de statistiques.

OPAL n'est pas la seule plateforme de partage des données qui se soucie de la vie privée. En France, le centre d'accès sécurisé aux données (CASD)[26] est une autre expression de cette technologie. Il fournit aux chercheurs un accès distant, au moyen de cartes à puce, à un ordinateur où les enquêtes publiques et les recensements nationaux peuvent être analysés. DeepMind de Google est aussi engagé dans la mise au point d'un système auditable pour utiliser les données médicales individuelles du système de santé (NHS)[27] du Royaume-Uni. Un « Verifiable Data Audit » garantit que toute interaction avec les données est l'objet d'un enregistrement accessible, afin de modérer le risque d'acte illicite.

Le potentiel de l'intelligence artificielle dépend de manière cruciale de la qualité des données sur lesquelles les algorithmes sont entraînés. Cependant les jeux de données apparemment anodines et anonymisées permettent à des tiers d'inférer une quantité surprenante d'informations sensibles sur les individus concernés, et les techniques d'anonymisation sont inefficaces car un petit nombre d'informations externes suffit à identifier les individus. L'avenir de l'IA nous impose de repenser notre approche de la protection des données. Des solutions comme OPAL sont à la pointe de cet effort, et forment le socle du partage des données personnelles et privées pour le bien public. Un objectif auquel nous pouvons tous adhérer.

18 de Montjoye, Y. A., Quoidbach, J., Robic, F., & Pentland, A. (2013, April). Predicting Personality Using Novel Mobile Phone-Based Metrics. In *SBP* (pp. 48-55).

19 Felbo, B., Sundsøy, P., Pentland, A. S., Lehmann, S., & de Montjoye, Y. A. (2015). Using deep learning to predict demographics from mobile phone metadata. *arXiv preprint arXiv:1511.06660*.

20 Jahani, E., Sundsøy, P., Bjelland, J., Bengtsson, L., & de Montjoye, Y. A. (2017). Improving official statistics in emerging markets using machine learning and mobile phone data. *EPJ Data Science*, *6*(1), 3.

21 de Montjoye, Y. A., Rocher, L., & Pentland, A. S. (2016). Bandicoot: a python toolbox for mobile phone metadata. *Journal of Machine Learning Research*, *17*(175), 1-5.

22 Bjorkegren, D., & Grissen, D. (2015). Behavior revealed in mobile phone usage predicts loan repayment.

23 Open Algorithms (2017), OPAL, www.opalproject.org/.

24 Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., et al. (2016). Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 Nepal Earthquake. *PLoS currents*, *8*.

25 Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., et al. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, *111*(45), 15888-15893.

26 Centre d'accès Sécurisé aux Données, CASD, https://casd.eu/en.

27 Suleyman, M., Laurie, B, (2017), Trust, confidence and Verifiable Data Audit, *DeepMind Blog*, https://deepmind.com/blog/trust-confidence-verifiable-data-audit.

"LA COMBINAISON DE MÉCANISMES DE CONTRÔLE D'ACCÈS, DE SANDBOX POUR LE CODE ET DE L'AGRÉGATION PERMET À OPAL DE GARANTIR QUE LES DONNÉES SONT UTILISÉES DE MANIÈRE ANONYME, Y COMPRIS PAR LES ALGORITHMES D'APPRENTISSAGE AUTOMATIQUE ET MÊME SI LES DONNÉES SONT SEULEMENT PSEUDONYMISÉES."