

Comité OMS

Comment le Big Data et l'intelligence artificielle peuvent-ils transformer les soins en santé ?

Rédigé par Line FARAH



Aperçu global

Contexte

L'utilisation de l'intelligence artificielle et du Big Data pose la question de la démocratisation et de l'acceptabilité sociale des changements qui sont ou seront introduits par ces technologies. A l'échelle planétaire, les gouvernements vont devoir réfléchir à des modes de régulation adaptés de ces technologies et de leurs applications qui se déploieront à travers le monde.

1

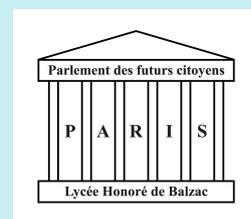
SANTÉ

2

INTELLIGENCE
ARTIFICIELLE

3

BIG DATA



Définitions des termes

Intelligence artificielle

L'« intelligence artificielle » éveille espoirs et peurs. Tous les secteurs sont concernés, et les perspectives économiques considérables : l'IA pourrait contribuer à hauteur de 15 700 Mds de dollars à l'économie mondiale en 2030, selon le cabinet de consultants Pricewaterhouse-Coopers, soit une hausse respective de 26 %, de 14,5 % et de 9,9 % des PIB de la Chine, des États-Unis et de l'Europe du Nord.

Émergeant lors de la conférence du Dartmouth College en 1956 portée par John McCarthy et Marvin Minsky, le concept d'intelligence artificielle (IA) a fait l'objet de près de soixante années de recherches et réalisations techniques et scientifiques. Ces dernières ont connu un bond technologique spectaculaire dans les années 2010. Dans l'intervalle, l'IA a connu différents coups d'arrêt, qui s'expliquent autant par les limites technologiques, des erreurs

scientifiques et des revers budgétaires que par l'arrêt de plusieurs programmes de recherche, le pessimisme des médias et la défiance des investisseurs. Certains ont alors parlé d'« AI Winter », pour caractériser ces périodes de stagnation.

« l'IA pourrait contribuer à 15 700 Mds de \$ à l'économie mondiale en 2030 »



LES RISQUES ET LES INCONVÉNIENTS DE L'INTELLIGENCE ARTIFICIELLE

Les machines n'ont pas de motivations, pas de désirs, pas de croyances, pas de système de représentation du monde. En fait, elles n'ont pas d'intelligence au sens propre du terme. Un algorithme, et par extension une intelligence artificielle, est toujours conçu par un être humain ou un groupe d'individus. Le code (donc les gènes) de l'IA n'est pas neutre car il reflète l'avis, l'intérêt, les finalités ou les biais intellectuels de son ou ses constructeurs.

Ces technologies peuvent donc comporter un risque majeur si elles sont monopolisées

par quelques conglomérats. C'est ce qui fait la difficulté de penser la technique aujourd'hui : elle n'est en rien réductible aux objets matériels qui sont offerts à notre perception et que nous avons sous la main. Elle réside essentiellement dans leur structure abstraite et formelle.

De plus, une question éthique et sociétale est introduite avec l'IA sur les rapports entre les hommes et les machines dans nos sociétés. Sommes-nous préparés à comprendre et à vivre avec les intelligences artificielles ? Des robots sont, par exemple, déjà utilisés dans les maisons de retraite japonaises, où les personnes âgées leur parlent et interagissent avec eux.



Définition du Big data

L'explosion quantitative des données numériques a obligé les chercheurs à trouver de nouvelles manières de voir et d'analyser le monde. Il s'agit de découvrir de nouveaux ordres de grandeur concernant la capture, la recherche, le partage, le stockage, l'analyse et la présentation des données.

Le volume

Le volume correspond à la masse d'informations produite chaque seconde. Selon des études, pour avoir une idée de l'accroissement exponentiel de la masse de données, on considère que 90 % des données ont été engendrées durant les années où l'usage d'internet et des réseaux sociaux a connu une forte croissance. L'ensemble de toutes les données produites depuis le début des temps jusqu'à la fin de l'année 2008, conviendrait maintenant à la masse de celles qui sont générées chaque minute. Dans le monde des affaires, le volume de données collecté chaque jour est d'une importance vitale.

La vitesse

La vitesse équivaut à la rapidité de l'élaboration et du déploiement des nouvelles données. Par exemple, si on diffuse des messages sur les réseaux sociaux, ils peuvent devenir « viraux » et se répandre en un rien de temps. Il s'agit d'analyser les données au cours de leur lignée (appelé parfois analyse en mémoire) sans qu'il soit indispensable que ces informations soient entreposées dans une base de données.

La variété

LES 5 V DU BIG DATA

Volume

Vélocité

Variété

Valeur

Variété

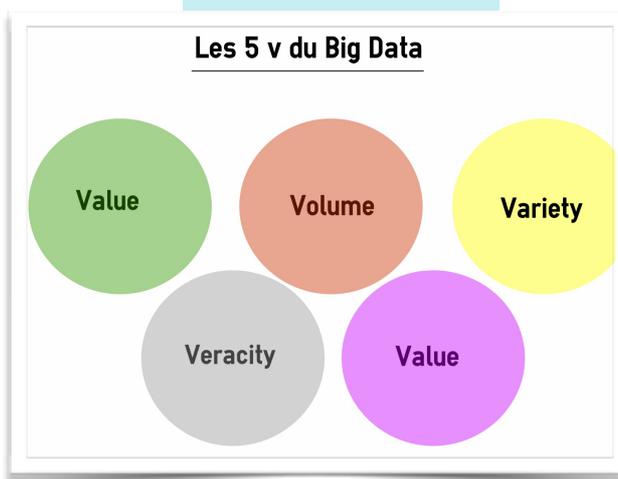
Seulement 20% des données sont structurées puis stockées dans des tables de bases de données relationnelles similaire à celles utilisées en gestion comptabilisée. Les 80% qui restent sont non-structurées. Cela peut être des images, des vidéos, des textes, des voix, et bien d'autres encore... La technologie Big Data, permet de faire l'analyse, la comparaison, la reconnaissance, le classement des données de différents types comme des conversations ou messages sur les réseaux sociaux, des photos sur différents sites etc. Ce sont les différents éléments qui constituent la variété offerte par le Big Data.

La véracité

La véracité concerne la fiabilité et la crédibilité des informations collectées. Comme le Big Data permet de collecter un nombre indéfini et plusieurs formes de données, il est difficile de justifier l'authenticité des contenus, si l'on considère les post Twitter avec les abréviations, le langage familier, les hashTag, les coquilles etc. Toutefois, les génies de l'informatique sont en train de développer de nouvelles techniques qui devront permettre de faciliter la gestion de ce type de données notamment par le W3C.

La valeur

La notion de valeur correspond au profit qu'on puisse tirer de l'usage du Big Data. Ce sont généralement les entreprises qui commencent à obtenir des avantages incroyables de leurs Big Data. Selon les gestionnaires et les économistes, les entreprises qui ne s'intéressent pas sérieusement au Big Data risquent d'être pénalisées et écartées.

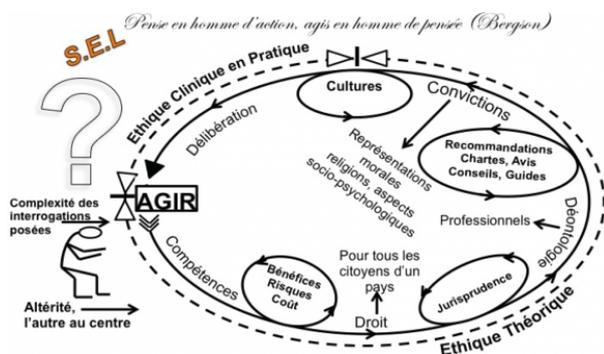


Puisque l'outil existe, ne pas s'en servir conduirait à perdre un privilège concurrentiel.

Risques de l'utilisation du Big Data en Santé

Quels sont les problèmes soulevés par les données de santé stockées dans un espace numérique ou *cloud* hébergé à l'étranger?

1. D'un point de vue **médico-légal** : le secret médical existe-t-il toujours ? Quelles responsabilités ont les soignants dans l'utilisation de ces données ?
2. D'un point de vue **éthique** : Comment sélectionner les données confidentielles ou non? Comment certifier les organismes qui pourront exploiter ou non des données nominatives pour des études scientifiques cliniques et épidémiologiques en garantissant la protection de ces données ?



3. D'un point de vue **épistémologique et sociale** : le Big Data n'est-il pas l'ouverture vers le Big Brother et l'anéantissement de l'individu par sa dilution dans la société ? En effet, les atteintes à la vie privée (tentative de chantage, rackets, harcèlement moral et sexuel notamment des enfants, harcèlement au travail, atteinte à la réputation...) n'ont jamais été aussi nombreux depuis le développement des réseaux sociaux. Ils conduisent à une dilution des frontières entre le travail et soi, les amis et soi, la famille et soi qui deviennent omniprésents, envahissants, intrusifs et occupent bientôt tout l'espace de votre vie. Des moyens informatiques de protection des données confidentielles et du respect de la vie privée doivent être renforcés. Une meilleure information du public sur l'usage qui peut être fait de leurs données personnelles doit être mise en place. Aujourd'hui, des

clauses de contrat souvent illisibles et pléthoriques, incluses dans les licences d'utilisation des réseaux sociaux et même de certains systèmes d'exploitation, obligent tacitement les usagers à renoncer à la propriété de leurs données personnelles à partir du moment où ils utilisent ces services ou logiciels.

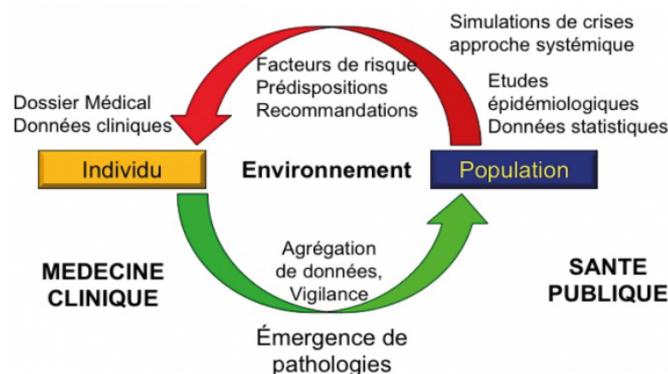
« La protection des données confidentielles et du respect de la vie privée doit être renforcée »

Les systèmes d'information médicaux dans le monde

La plupart des systèmes d'aide à la décision médicaux ont d'abord été développés dans les pays occidentaux.

En Amérique et en Europe, puis récemment en Asie où il existe un développement exponentiel de l'ingénierie informatique (Corée, Chine, Singapour, Japon). Entre les années 1964 à 2000 plus de 10000 systèmes d'aide à la décision en médecine (SADM) ont été développés. Sur un échantillon de 110 systèmes la répartition géographique de ces systèmes est représentée figure 6. Elle montre que les USA et l'Europe ont été leaders dans le développement des SADM de 1964 à 2000. L'échantillon de systèmes n'est pas nécessairement très représentatif de la population globale de systèmes car il a surtout été recueilli pour étudier le types de modèles utilisés. On assiste aujourd'hui à de nombreuses applications des SADM en Afrique afin d'apporter une aide sanitaire de terrain.

En Asie, à Singapour par exemple, les SIAD et les entrepôts de données cliniques progressent de manière explosive depuis 2002 dans le but d'améliorer la qualité des soins (Ling et al., 2014). L'utilisation des SIAD en Afrique améliore la structuration des systèmes de soin et



participe au progrès de l'organisation de la santé dans de nombreux pays.

Conséquences de l'utilisation du Big data et de l'IA sur la santé et la vie privée

Risque du fichage des personnes

La mise en fiche des populations est ancienne, Joseph Fouché, les régimes totalitaires (les Nazis, la Stasi... l'utilisaient mais, heureusement, ils ne disposaient pas des moyens informatiques actuels. La constitution de bases de données nominatives sous couvert de besoins de santé publique a ému les médecins, en particulier, le système GAMIN qui organisait la discrimination a priori d'enfants socialement défavorisés. La loi du 15 juillet 1970 obligeait les médecins à leur déclaration, ce qui constituait une discrimination a priori de ces enfants. Sur le plan technique, la notion de base de données locale administrée et protégée dont la vocation est connue, déclarée et contrôlée par la Commission Nationale Informatique et Liberté (CNIL) devient obsolète du fait des flux transfrontières de données.

Conséquences du développement du « cloud computing » dans le domaine de la santé

Le Cloud peut se définir comme un espace de stockage externe virtuel et partageable des données. Il se présente comme un disque externe qui est administré par des tiers inconnus et qui est mis à votre disposition sur un serveur à l'aide d'une adresse Internet et d'un protocole de connexion. Le principal avantage du Cloud Computing est d'offrir un vaste espace de mémoire de sauvegarde distribué accessible de n'importe où, pourvu que l'on dispose d'une connexion Internet pour un coût assez faible. Les professionnels de santé ont des besoins

importants de stockage de gros volumes de données liées aux examens paracliniques : images (IRM, Scintigraphies, TDM, Radios), films (échographie 3D, Signaux (MEG, EEG, ECG,...) et textes (observations, Dossier Médical Informatisé, examens biologiques, prescriptions. Ces besoins augmentent de manière exponentielle avec les nouveaux systèmes d'imageries et la politique de numérisation systématique des examens pratiquée par les établissements. La télémédecine qui fonctionne en mode « *Workflow* » bénéficie du partage facile des données entre soignants s'occupant d'un même patient. Les données sont accessibles de partout : à l'hôpital au lit du patient ou lors d'une visite à domicile. En médecine tout retard de décision pouvant être fatal, la rapidité d'accès est donc essentielle. Actuellement, l'accès au dossier du patient peut même avoir lieu à partir des unités de soin mobile, ce qui constitue un atout essentiel pour agir rapidement de manière adaptée. La sous-traitance offre une solution économique pour les médecins et les hôpitaux.

Solutions proposées

Risque d

Le patient dispose d'un droit à l'opacité de ses données personnelles intimes liées à sa santé ce afin de garantir qu'elles soient uniquement utilisées pour le parcours de soin en garantissant la sécurité du patient et en le préservant de l'exploitation de ces données par des personnes malintentionnées. Les usages du Big data pour la consommation courante le sont aussi pour la santé, il faut donc avertir et former le public sur les risques des réseaux sociaux afin qu'ils puissent garder le contrôle sur leurs données. Le législateur devrait interdire, dans les licences d'utilisation des logiciels, les clauses léonines concernant le droit que s'arroge l'opérateur de disposer tacitement des données de l'utilisateur dès lors qu'il utilise un logiciel, un service, un site, ou un système.



Il faut développer un Intranet spécifique mieux sécurisé, à l'usage des professionnels de santé pour le transfert, le partage des données médicales.

Les Clouds comportant des données médicales doivent être hébergés uniquement sur le territoire Français et administrés par des informaticiens tenus au secret professionnel sous contrôle de la CNIL et du Ministère de la santé. Des sociétés françaises fournissent des Datacenter sur le sol français afin de stocker des données médicales. Elles doivent obtenir un agrément en déposant un dossier ASIP Santé auprès de l'Agence des Systèmes d'Information Partagés de Santé. La cryptographie des données médicales doit avoir lieu à l'aide d'algorithmes de cryptage efficaces et performants permettant de ne décrypter les données nominatives que de manière éphémère et parcellaire et uniquement à l'usage des seules personnes autorisées. Cette solution est proposée par la couche CRYPTO de l'architecture proposé par le W3C. Cela nécessite une protection spécifique de chaque donnée individuellement, ce qui est actuellement techniquement possible. Il convient donc de mieux réglementer le périmètre des moteurs de recherche et des réseaux sociaux en renforçant la législation à l'échelle Européenne. Il faut standardiser les licences ou contrats d'utilisation en rejetant les clauses inacceptables obligeant les utilisateurs à abandonner la propriété de leurs données au profit des fournisseurs de logiciels ou de services sur Internet.

Conclusion

L'utilisation des nouvelles technologies et le développement d'Internet au niveau mondial nécessite de faire converger les forces de plusieurs grands puissances internationales afin de se doter de nouveaux outils techniques et juridiques pour contrôler ces usages et permettre aux patients de conserver la maîtrise

de leurs informations personnelles et intimes confiées aux soignants. Cette confiance demeure essentielle à la relation de soin. Les dangers et les risques sociaux, éthiques et médicaux de la divulgation de données personnelles existent réellement. L'objectif des réflexions au sein du comité OMS est donc d'encourager l'élaboration de moyens pour aider les soignants et les patients à conserver la confidentialité de leur données tout en optimisant les parcours de soin dans un contexte où les avancées technologiques rendent cette tâche à la fois de plus en plus facile à réaliser mais difficile à maîtriser. Au-delà du Big data et l'intelligence artificielle, il s'agit de la préservation de la vie privée et l'amélioration des techniques de soins pour les acteurs du systèmes de santé afin de mieux soigner les patients.

Sitographie

- Bastien L. *Le Big Data, l'IA et le Machine Learning transforment les soins de santé.* 28/09/2016. Disponible sur : <https://www.lebigdata.fr/big-data-soins-de-sante> (Consulté le 12/12/2017)
- Terra nova. *La santé à l'heure de l'intelligence artificielle.* 5/12/2017 Disponible sur : http://tnova.fr/system/contents/files/000/001/489/original/Terra-Nova_Sante-Intelligence-Artificielle_051217.pdf?1512401730 (Consulté le 12/12/2017)
- Disponible sur : <https://www.lebigdata.fr/definition-big-data> (Consulté le 12/12/2017)
- Joël Colloc. *Health and Big Data: the State and the Individuals, Powerless in Front of Powers of Networks* Disponible sur : <http://journals.openedition.org/espacepolitique/3493?lang=en> (Consulté le 12/12/2017)

