

17 Intelligence artificielle et santé



Marguerite BRAC DE LA PERRIÈRE,
Avocat associé du cabinet *Lerins & BCW*,
spécialisée en santé numérique

1 - L'intelligence artificielle est définie comme un système « soit fondé sur des logiciels, soit intégré dans des dispositifs matériels, et qui fait preuve d'un comportement intelligent, notamment en collectant et traitant des données, en analysant et en interprétant son environnement et en prenant des mesures, avec un certain degré d'autonomie, pour atteindre des objectifs spécifiques¹ ». Elle constitue un outil d'aide puissant, qui assiste l'homme dans des environnements complexes, en lui permettant de prendre la décision la plus éclairée, dans un contexte donné, grâce à sa puissance de calcul. Aux qualités du cerveau humain s'ajoutent ainsi les forces et la puissance de calcul d'un cerveau digital. En ce sens, une intelligence humaine augmentée.

2 - Dans le domaine de la santé, l'intelligence artificielle doit permettre une meilleure prise en charge du patient, au service de la médecine dite « des quatre P » : prédictive, préventive, personnalisée, précise.

3 - Nombreuses sont les illustrations médicales des progrès qu'elle permet, diagnostiques et thérapeutiques, notamment en cancérologie et analyse des tumeurs, anatomopathologie, dermatologie, chirurgie, ophtalmologie, radiologie et imagerie médicale, etc.

4 - Les logiciels d'apprentissage peuvent aujourd'hui se comparer aux meilleurs spécialistes dans certains domaines tels que l'analyse des mammographies suspectes, la détection des rétinopathies diabétiques, ou le diagnostic des mélanomes.

5 - Il apparaît essentiel de favoriser le développement des outils d'intelligence artificielle en matière de santé, et pour ce faire, en premier lieu, la lisibilité du cadre éthique et juridique ayant trait à l'accès aux données massives.

1. La nécessité de données massives

6 - L'intelligence artificielle, les algorithmes, ont besoin de données massives pour apprendre, fonctionner, s'améliorer et, le plus souvent, de données non pas anonymisées mais pseudonymisées de sorte à les chaîner entre elles et à établir des corrélations et des liens.

7 - Or, les données de santé ont longtemps été considérées comme des attributs intimes des patients à l'usage exclusif des producteurs de données pour les seuls besoins de leur prise en charge.

8 - En 2018, Cédric Villani a révolutionné cette vision désuète et sclérosante au bénéfice d'un nouveau paradigme fédérateur,

parfois qualifié de « datasolidarité » : les données de santé financées par la solidarité nationale constituent un patrimoine commun devant être mis au service du patient et du système de santé dans le respect de l'éthique et des droits fondamentaux.

9 - Dans le même temps, l'insuffisance du recours au numérique et à l'intelligence artificielle dans notre système de santé est apparue comme un problème éthique² au regard de la qualité des soins, des moindres exigences éthiques de conception des outils par les acteurs étrangers, et de la compétition internationale.

10 - S'appuyant sur cette nouvelle philosophie et ce constat d'insuffisance, et conscients que, pour développer le potentiel de l'intelligence artificielle dans le domaine de la santé, la France devait collecter et structurer les données de santé et permettre leur accès et utilisation. C'est pourquoi, les pouvoirs publics ont souhaité libéraliser l'accès aux données de santé, et créer une plateforme d'accès et de mutualisation des données pertinentes pour la recherche mais aussi l'innovation en matière de santé.

2. Les sources de données

A. - Le Health Data Hub

11 - C'est dans ce contexte que la Loi relative à l'organisation et à la transformation du système de santé du 24 juillet 2019 a créé la plateforme des données de santé³ en charge du Health Data Hub⁴ destiné à rassembler à la fois les données médico-administratives [SNDS] et les données cliniques, et à constituer un dispositif de partage entre producteurs et utilisateurs de données de santé pseudonymisées.

12 - Aux fins de mise en œuvre d'un projet mobilisant des données de santé, les utilisateurs de données peuvent ainsi s'adresser au Health Data Hub, guichet unique déclenchant l'instruction par le comité scientifique et éthique national, le CESREES, et par la CNIL (Commission Nationale de l'Informatique et des Libertés), seule habilitée à autoriser le traitement. Ces deux étapes visent à garantir la sécurité des données de santé et la pertinence de leur utilisation, laquelle doit répondre à une finalité d'intérêt public.

13 - La notion d'intérêt public est appréciée comme principalement articulée autour de « la garantie de normes élevées de

1. Proposition de règlement A9-0186/2020 contenant des recommandations à la Commission concernant un cadre d'aspects éthiques en matière d'intelligence artificielle, de robotique et de technologies connexes du 8 octobre 2020, disponible en ligne : [https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_FR.html].

2. Avis 129 relatif à la « Contribution du Comité consultatif national d'éthique à la révision de la loi de bioéthique 2018-2019 », septembre 2018, disponible en ligne : [https://www.ccne-ethique.fr/sites/default/files/avis_129_vf.pdf].

3. Art. L. 1462-1 du Code de la santé publique.

4. Disponible en ligne sur le site du Health Data Hub : [https://www.health-data-hub.fr].

qualité et de sécurité des soins de santé et des médicaments ou des dispositifs médicaux », et des études, recherches et évaluations⁵, étant précisé qu'un bénéfice direct ou indirect du projet pour les individus, la société ou la communauté scientifique doit être démontré, et qu'un intérêt commercial n'est pas incompatible avec l'intérêt public

14 - À ce jour, le répertoire de projets du Health Data Hub compte d'ores et déjà quelques études rétrospectives de données pour le développement d'algorithmes d'intelligence artificielle répondant aux conditions de la MR004 ou autorisées par la CNIL, dont certaines à l'initiative d'industriels en matière d'intelligence artificielle s'appuyant, le cas échéant, sur un laboratoire de recherche ou un bureau d'études.

15 - Les caractéristiques des projets retenus par le Health Data Hub gagneront à être précisées et publiées à des fins de lisibilité et de sécurité juridique.

B. - Les entrepôts de données de santé

16 - Dans le même temps, les entrepôts de données de santé se sont multipliés aux fins de mise en œuvre de traitements de données massives.

17 - Les entrepôts sont fondés sur le consentement des personnes concernées, ou sur une demande d'autorisation à la CNIL – bientôt remplacée par un engagement de conformité au référentiel « entrepôt de données de santé » en concertation⁶.

18 - Les données qu'ils contiennent peuvent être réutilisées dans le respect de l'une des méthodologies de référence de la CNIL, notamment les MR003 et les MR004, ou d'une demande d'autorisation recherche, la CNIL visant une réutilisation « principalement à des fins d'études, de recherches et d'évaluations dans le domaine de la santé »⁷.

19 - Dans ce contexte, la réutilisation des données aux fins de développement de l'intelligence artificielle est marginale.

C. - Les bases de données anonymes

20 - Certains projets d'intelligence artificielle peuvent être réalisés à partir de bases de données anonymes, notamment s'agissant du développement d'outils algorithmiques en matière d'imagerie médicale, par exemple de détection de lésions.

21 - Dans le même sens, il y a lieu de mentionner l'utilité des données synthétiques, créées artificiellement à l'aide d'algorithmes, mais représentatives, et utilisées en tant que données de test pour les besoins de l'intelligence artificielle en exploration de données et en apprentissage automatique⁸.

22 - Si le RGPD (Règlement Général sur la Protection des Données) ne s'applique pas au traitement d'informations anonymes, il s'applique au traitement consistant à anonymiser les données, de sorte que pour bénéficier d'une base de données

anonymes, les acteurs sont confrontés à une double difficulté articulée autour de la qualité de l'anonymisation des données – laquelle doit permettre pour être parfaite d'empêcher toute individualisation, inférence et corrélation – et de la définition des conditions juridiques de l'anonymisation.

23 - S'agissant de la qualité de l'anonymisation, surtout en matière d'imagerie, une appréciation fine *in concreto* s'impose, étant précisé qu'un ensemble de données pour lequel au moins un des trois critères n'est pas respecté ne pourra être considéré comme anonyme qu'à la suite d'une analyse détaillée des risques de réidentification.

24 - Ainsi, l'appréciation de la notion d'anonymisation est centrale, subjective et doit être appréciée *in concreto*, le RGPD prévoyant :

- Il n'y a pas lieu d'appliquer les principes relatifs à la protection des données aux informations anonymes, à savoir les informations ne concernant pas une personne physique identifiée ou identifiable⁹ ;

- Pour déterminer si une personne physique est identifiable, il convient de prendre en considération l'ensemble des moyens raisonnablement susceptibles d'être utilisés (.) les facteurs objectifs, tels que le coût de l'identification et le temps nécessaire à celle-ci, en tenant compte des technologies disponibles au moment du traitement et de l'évolution de celles-ci.¹⁰

25 - S'agissant des conditions juridiques applicables aux traitements d'anonymisation, la définition des bases de licéité, la qualification des acteurs y procédant, les conditions de respect des droits et libertés des personnes concernées par les données, pourraient être clarifiées afin de favoriser la sécurité juridique, et donc la pérennité des projets initiés. À cet égard, le défaut de lisibilité est susceptible de conduire d'une part à des dérives, d'autre part à une distorsion de concurrence entre les acteurs. Dans cette attente, les acteurs procédant à l'anonymisation de données de santé doivent veiller à constituer une documentation solide, et/ou procéder à une demande d'autorisation à la CNIL.

3. Perspectives pour l'intelligence artificielle

26 - Dans cet environnement réglementaire complexe, l'effort de lisibilité doit être poursuivi afin de favoriser un accès effectif aux données de santé nécessaires au développement d'outils d'intelligence artificielle en matière de santé par les organismes privés, et ce au service des ambitions nationales.

27 - S'il est nécessaire que les outils d'intelligence artificielle en matière de santé développés fassent l'objet d'une étude, recherche ou évaluation afin de déterminer si une prise en charge médicale incluant leur utilisation présente un bénéfice par rapport à une prise en charge médicale sans leur utilisation, pour autant le préalable est que ces outils d'intelligence artificielle soient développés à partir de données issues du « patrimoine » français par un recours large au Health Data Hub.

28 - Nombre d'industriels recourent encore aujourd'hui à des jeux de données provenant de l'étranger, moyennant de lourds investissements financiers, et une qualité ou représentativité possiblement moindres.

29 - Rappelons l'importance des bases de données, tant à l'égard de la qualité des algorithmes, que du bénéfice médical

5. Health Data Hub, « Qu'est-ce que l'intérêt public ? », disponible en ligne : [https://www.health-data-hub.fr/interet-public].

6. CNIL, Projet de référentiel relatif aux traitements de données à caractère personnel mis en œuvre à des fins de création d'entrepôts dans le domaine de la santé, 8 mars 2021, disponible en ligne : [https://www.cnil.fr/fr/entrepots-de-donnees-de-sante-la-cnil-lance-une-consultation-publique-sur-un-projet-de-referentiel].

7. CNIL, « Traitements de données de santé : comment faire la distinction entre un entrepôt et une recherche et quelles conséquences ? », 28 novembre 2019, disponible en ligne : [https://www.cnil.fr/fr/traitements-de-donnees-de-sante-comment-faire-la-distinction-entre-un-entrepot-et-une-recherche-et].

8. Documentation du SNDS, « Données de synthèse du lab santé de la DREES », 9 décembre 2020, disponible en ligne : [https://documentation-snds.health-data-hub.fr/formation_snds/donnees_synthetiques.html#donnees-de-synthese-du-lab-sante-de-la-drees].

9. Règlement (UE) 2016/679 du parlement européen et du conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données), considérant 26.

10. *Ibid.*

pour le patient. En effet, des données biaisées donneront des résultats non fiables. Pour illustrer ce propos, les algorithmes construits et entraînés sur des jeux de données étrangers pourraient ne pas être adaptés aux patients français aux profils de maladies, habitudes de soins et politiques de santé nationales, différents. Ils devraient donc à tout le moins être validés sur un échantillon français de population.

30 - L'équilibre reste à parfaire entre une réglementation protégeant les patients, respectueuse du secret médical, de la vie privée et de la protection des données, et un cadre juridique favorisant les initiatives industrielles sans pénaliser les chercheurs français et européens face au reste du monde. ■