



APPARIEMENT DES DONNÉES : LE POTENTIEL DE L'IA DANS LE PARCOURS DE SOINS



ISABELLE LAFORGUE (X01)
directrice digital, transformation & innovation,
AstraZeneca France

L'intelligence artificielle trouve des applications déterminantes dans le traitement de la donnée industrielle, mais elle est encore limitée dans le domaine de la santé, en particulier à cause d'une dispersion de la donnée entre les différents acteurs. Tout en préservant la confidentialité de la situation individuelle du patient, il serait possible de faire bénéficier tous les patients de progrès inestimables pour leurs soins grâce à un partage de la donnée.

Dans tous les secteurs et toutes les industries, il est un lieu commun que de dire que l'émergence de données massives couplées aux nouvelles capacités de calcul ouvre des perspectives formidables. La grande consommation, le transport... nourrissent jour après jour leurs algorithmes de données générées par leur activité et celle de leurs utilisateurs, et bâtissent des chefs-d'œuvre de modèles d'intelligence artificielle. À la fois pour plus d'efficacité dans leurs opérations et également pour prédire et anticiper tous les souhaits de leurs utilisateurs, qu'ils

soient collaborateurs ou consommateurs. Le potentiel de ce phénomène dans le secteur de la santé est encore plus prometteur. Parmi les données de tous les secteurs, la donnée de santé est très particulière.

Une protection particulière en France

Elle est particulière tout d'abord par sa nature. En France, elle est très protégée et son accès ainsi que son traitement sont encadrés par des règlements et obligations très protecteurs pour l'individu, qui, dans ce cas précis, est souvent un patient. Les données de santé sont protégées par la loi informatique et libertés, le RGPD et le Code de la santé publique, et les autorisations de traitement sont délivrées par la Cnil (Commission nationale de l'informatique et des libertés), le régulateur français. L'ensemble de ces protections garantit la nature des traitements de la donnée de santé, permettant à tous, patients comme organisations, d'évoluer avec des règles connues et respectueuses de la vie privée des personnes.

La richesse de la donnée de santé

Une caractéristique forte de la donnée de santé, c'est son incomparable richesse. La donnée dans le domaine de la santé, c'est avant tout la donnée générée tout au long du parcours de soins d'un patient, depuis les médecins qu'il a consultés, les examens qui lui ont été prescrits, leurs résultats, les médicaments qu'il a pris, son passage à l'hôpital... Des données extrêmement riches, non seulement par la multiplicité des acteurs qui participent à les générer, mais aussi par leur nature elle-même. Ce sont des données tabulaires (donc structurées) de parcours de soins, des données non structurées comme de l'imagerie médicale, des résultats d'examens cliniques, biologiques, génomiques, protéomiques... une véritable richesse d'informations.

Une exceptionnelle base de données en France

La structuration de cette donnée de santé est aussi propre à la France : notre système de soins centralisé nous a permis d'établir un recueil des données de soins des patients unique au monde. Le SNDS (Système national des données de santé) rassemble toutes les données de consultation, de prescription, d'actes et remboursement de soins de 66 millions de Français depuis vingt ans, en un unique endroit. Et, au-delà, les nouvelles bases constituées au cours de la pandémie comme la base des

dépistages Covid SI-DEP, système d'informations de dépistage. Cette centralisation permet aujourd'hui d'avoir toutes ces données sur une seule plateforme, accessible dans le respect des obligations réglementaires propres aux données personnelles de santé.

Mais une dispersion persistante

Toutes les promesses de la donnée de santé ne tiennent que si cette donnée est accessible à des projets de recherche. En France, l'accès à ces données hautement sensibles est aujourd'hui régi par des règles bien définies, et le chemin réglementaire, s'il est contraignant, n'en est cependant pas moins clair. Délais et dispersion des données de santé sont aujourd'hui les freins majeurs à l'exploitation de cette donnée, et donc à la recherche, à l'innovation et à l'amélioration des parcours de santé. Si le SNDS centralise les données de remboursement de la Sécurité sociale, en revanche les résultats d'analyse biologique, d'examen médical... sont des données disponibles dans les établissements de soins (privés et publics) et les laboratoires (privés) d'analyse. Ces données sont aujourd'hui dispersées dans différents lieux physiques et sous la responsabilité de différents acteurs. Accéder à ces données est possible, mais au prix de temps de mise à disposition technique et réglementaire souvent longs de plusieurs mois. L'éclatement des acteurs rend le système inefficace, et l'accès, quand il est possible, est souvent limité à une seule base de données à la fois, dans l'espace de stockage de la structure qui a généré cette donnée (l'entrepôt de données d'un hôpital, le serveur d'un laboratoire d'analyse, etc.).

Le rêve d'une base de données intégrale

Pourtant, si l'on parvenait à réconcilier toutes ces données, on aurait accès à des parcours de soins exhaustifs, une richesse et une complétude d'informations qui permettraient ensuite de nourrir puissamment des modèles d'intelligence artificielle pour le suivi longitudinal des patients. De manière très schématique, on pourrait imaginer pouvoir réconcilier les données du SNDS (les informations sur les soins reçus par le patient) détenues par la Caisse nationale de l'assurance maladie avec les résultats des analyses biologiques prescrites hébergées par le laboratoire biologique qui les a réalisées, et également avec des résultats d'examens du centre d'imagerie médicale où le patient s'est ensuite rendu. Une « vraie vie » →

→ médicale de bout en bout reconstituée au service des patients, pour leur permettre individuellement par la suite de bénéficier de meilleurs traitements et suivis, et collectivement de nourrir de nouvelles études cliniques à partir de ces données de vie réelle.

Des initiatives publiques et privées

Cette réconciliation, qui s'appelle de l'appariement de données et que l'IA permet aussi de faciliter *via* des modèles de corrélation statistique, nécessite de rassembler la donnée en un unique lieu. Des initiatives publiques et privées de regroupement des données se développent en France, menées en tête par le *Health Data Hub*, la plateforme des données de santé lancée par le gouvernement français en 2019. Des initiatives privées telles que celles de l'Inca-Ariis ou d'Agoria Santé (auxquelles participe AstraZeneca) complètent cette démarche avec une pleine responsabilité d'émulation et d'accélération. Au-delà de l'enjeu national, toutes ces initiatives devront permettre à la France de tirer le plein potentiel de son système de soins centralisé, ce fameux SNDS unique en son genre, et de se distinguer au niveau européen. Dans un cadre réglementaire clair et constant, c'est à présent aux acteurs de la donnée de santé, producteurs de données comme porteurs de projet d'intelligence artificielle, acteurs privés et publics, de collaborer pour permettre l'émergence de dispositifs de centralisation. Les intérêts particuliers doivent laisser place à une dynamique commune pour atteindre un équilibre collectif supérieur à l'ensemble des équilibres individuels.

L'application de l'intelligence artificielle

L'intelligence artificielle appliquée à la donnée de santé est porteuse d'une grande promesse et d'un grand défi pour le secteur de la santé et ses acteurs, publics comme les centres de soins ou privés comme les start-up ou les laboratoires pharmaceutiques. Une grande promesse, car la donnée de santé est d'une richesse inégalée et les perspectives de ce que pourra y apporter l'intelligence artificielle sont très excitantes. Ce potentiel ne s'exprimera cependant pas encore pleinement, et ne s'exprimera que si une nécessaire collaboration entre les acteurs, publics et privés, se met en place. Un grand défi également, car ce potentiel est aujourd'hui peu exploité dans une industrie qui a pourtant fait de la science et des découvertes scientifiques les piliers de son développement.

Le domaine des essais cliniques

Un exemple éclairant est celui des essais cliniques. Tandis que les industries plus techniques comme l'aéronautique ou l'automobile ont depuis longtemps utilisé la modélisation informatique pour limiter le recours aux essais « en vie réelle », l'industrie pharmaceutique commence seulement à travailler sur le sujet des bras de contrôle synthétique dans ses essais cliniques. L'idée est simple : en disposant de données médicales suffisamment nombreuses de parcours de soins de patients, on peut dans un essai clinique se dispenser d'une cohorte de patients à qui l'on délivre normalement un placebo, par une modélisation du comportement de ces patients. La démarche a aujourd'hui prouvé son efficacité pour certaines applications, et elle va continuer à se perfectionner techniquement et scientifiquement. Cependant elle reste encore complexe à déployer car les autorités sanitaires doivent accepter cette nouvelle méthode pour permettre aux industriels de l'intégrer dans leurs dossiers d'autorisation. Les premières autorisations ont été accordées par la FDA (Food and Drug Administration) aux USA et l'EMA (Agence européenne des médicaments) en Europe, les réflexions avancent, mais sont encore frileuses au regard des enjeux pour les patients. Des patients acceptent de rejoindre des essais cliniques, avec l'espoir d'un traitement plus prometteur, et certains reçoivent aujourd'hui un placebo ou le traitement standard, là où demain l'intelligence artificielle pourra leur permettre de bénéficier du traitement. FDA et EMA en particulier sont des acteurs majeurs de cette collaboration entre acteurs, régulateurs et industriels, nécessaire pour permettre l'émergence de tout le potentiel de la donnée de santé dans les essais cliniques.

Construire le parcours de soins

Un autre domaine dans lequel une collaboration entre acteurs publics et privés, industriels et start-up est nécessaire est le parcours de soins du patient. L'intelligence artificielle peut permettre à un patient d'être diagnostiqué plus efficacement et plus tôt, traité et accompagné au cours de sa maladie, et là encore, si la technologie existe, elle n'a cependant pas encore pu exprimer tout son potentiel. Les capacités prédictives des modèles d'intelligence artificielle, couplées à la grande quantité de données disponibles, permettent aujourd'hui de développer de nombreux outils d'aide à la décision et au diagnostic. Ces modèles permettent



“Libérer grâce à l’IA toutes les promesses de la donnée au service du patient.”

par exemple d’accompagner la lecture d’une image d’examen (radiographie, IRM...) ou d’un résultat d’analyse médicale. Ils permettent également d’anticiper des événements de soins dans la vie du patient, comme une hospitalisation ou un risque de récurrence. Les dispositifs existent ; un grand nombre d’entre eux sont développés en France par des start-up à partir d’une cohorte de patients en réponse à une question médicale (survie, diagnostic, par exemple). Un modèle d’intelligence artificielle a besoin pour être efficace et pertinent d’être nourri d’une grande quantité de données – le plus possible – et de données de qualité, d’où l’importance de l’accès à cette donnée de santé. Développer un modèle d’intelligence artificielle pertinent est une première étape technologique nécessaire ; la suivante est réglementaire, avec l’éventuelle autorisation d’être un dispositif médical, soit un outil autorisé dans la pratique d’un professionnel de santé. Enfin reste l’ultime et décisive étape de l’adoption.

Former les professionnels de santé

Ces dispositifs doivent pour être utilisés en première étape être acceptés par les autorités de santé, dont on a déjà vu la nécessaire collaboration dans le développement des produits de l’intelligence artificielle en santé. Les étapes pour devenir un dispositif médical sont aujourd’hui claires et connues, à défaut d’être simples. L’adoption par les professionnels de santé est une seconde étape clé pour leur usage. Des freins existent sur la manière d’intégrer ces nouveaux outils

dans les usages médicaux, à la fois sur le volet pratique et sur le volet théorique. Le constat en France est que nos professionnels de santé ne sont pas formés systématiquement à ces nouvelles technologies au cours de leur cursus universitaire, ni au cours de leur formation continue. Former, accompagner les professionnels de santé à l’intelligence artificielle et à son impact dans la détection, le parcours de soins et le suivi du patient devient nécessaire, au risque qu’un écart ne se creuse entre le potentiel de l’intelligence artificielle en santé et la réalité dans la pratique de nos professionnels de santé.

Acteurs de santé, encore un effort !

Autorités de santé et professionnels de santé ont une responsabilité de collaboration aux côtés des industriels pour permettre le déploiement des solutions d’intelligence artificielle au service des patients. Cette articulation doit se mettre en place afin de laisser la technologie de l’intelligence artificielle s’installer dans la pratique médicale et le parcours de soins. Notre système de santé est robuste, mais éclaté entre les acteurs historiques des soins, souvent détenteurs de la donnée de santé, et des nouveaux acteurs qui, eux, sont détenteurs de la technologie pour tirer tout le potentiel de cette donnée. Le secteur de la santé est peut-être conservateur par nature, prudent car intervenant sur l’être humain ; il doit cependant rassembler et articuler ses acteurs pour permettre de libérer grâce à l’IA toutes les promesses de la donnée au service du patient. X