

Enjeux éthiques de l'IA en santé : une humanisation du parcours de soin par l'intelligence artificielle ?

Fabrice Muhlenbach

Université de Lyon, UJM-Saint-Etienne, CNRS
Laboratoire Hubert Curien, UMR 5516
18 rue du Professeur Benoît Luras, F-42023 Saint Etienne, FRANCE
E-mail: fabrice.muhlenbach@univ-st-etienne.fr
ORCID: 0000-0002-1825-4290

Résumé. ¹ Envisager le recours à l'intelligence artificielle pour une plus grande personnalisation de la prise en charge du patient et une meilleure gestion des ressources humaines et matérielles peut sembler une opportunité à ne pas manquer. Afin de proposer une meilleure humanisation du parcours de soin, l'intelligence artificielle est un outil que les décideurs du milieu hospitalier doivent s'approprier en veillant aux nouveaux enjeux éthiques et conflits de valeurs que cette technologie engendre.

1 Introduction

Quiconque s'est déjà retrouvé dans un service des urgences en pleine nuit, un nourrisson pleurant dans les bras, rêverait d'être tout de suite pris en charge au lieu d'être confiné dans une salle avec d'autres patients dans la même situation de détresse : ici, à la terrible impuissance d'être le témoin de la souffrance de son enfant s'ajoute l'angoisse liée à l'ignorance du temps à patienter avant de se retrouver face à un médecin. À défaut de personnel médical disponible, on aurait souhaité qu'un robot d'accueil ² puisse nous orienter, accéder à notre dossier médical et nous aider à remplir le questionnaire d'entrée afin de collecter les données permettant à une intelligence artificielle de poser un prédiagnostic. En comparant les différents profils des personnes entrant au service des urgences, ce programme d'IA pourrait être en mesure de prioriser les patients, d'aider à la logistique, d'envoyer des alertes au personnel et de fournir à chaque patient des informations permettant de le rassurer sur son état et d'indiquer une estimation réaliste du temps d'attente avant d'être pris en charge.

Loin d'être de la science-fiction, il s'agit peut-être d'une réponse à la situation de crise que connaît l'hôpital où, faute de moyens et de ressources humaines disponibles, la douleur et le sentiment d'abandon qu'éprouvent certains patients les amènent à un mécontentement qui peut parfois dégénérer en incivilités ou agressions envers un personnel soignant déjà bien malmené par des conditions de travail épuisantes. Envisager le recours à l'intelligence artificielle pour

1. Pré-publication d'un article à paraître dans la revue *Soins Cadres*.
2. De tels robots d'accueil existent, à la manière du robot humanoïde Pepper, adopté par la RATP et la SNCF, qui est chargé de renseigner les voyageurs.

Enjeux éthiques de l'IA en santé

l'amitié	l'amour	l'art	l'authenticité	la beauté
le bien-être	la compétence	la confiance	la connaissance	la créativité
la dignité	l'égalité	l'enfance	l'érotisme	l'honneur
l'humilité	l'humour	l'impartialité	l'intéressant	la justice
la liberté	le luxe	le plaisir	le pouvoir	la propriété
le sacré	la santé	le savoureux	la solidarité	le sublime
la tradition	l'utilité	la vertu	la vie	la vie privée

TAB. 1 – Liste des 35 valeurs du *Petit Traité des valeurs* (Deonna et Tieffenbach, 2018)

optimiser les flux de patients et de matériel, réduire les coûts de mobilisations des ressources humaines et de matériels tout en améliorant l'expérience du patient (moins de temps d'attente, un meilleur parcours de soin), voilà de quoi réjouir les directeurs des soins et cadres de santé, mais un tel changement ne peut s'effectuer sans conséquences éthiques.

2 Au cœur de l'éthique : la gestion des conflits de valeurs

Qu'est-ce que l'éthique ? D'un point de vue philosophique, l'éthique concerne les réflexions sur les comportements à adopter pour rendre le monde humainement habitable, avec comme finalité la recherche d'un idéal de société et de conduite de l'existence. D'un point de vue pratique, le cœur de l'éthique concerne des conflits entre des valeurs. Ces valeurs – considérées comme l'idéologie d'un groupe d'individus – peuvent par exemple être celles proposées par Deonna et Tieffenbach (2018) dans le Tableau 1.

Le milieu hospitalier est dédié dans son ensemble au service d'une même cause : le *patient*. Les valeurs privilégiées dans le contexte médical sont ainsi la *vie*, la *santé* et le *bien-être* du patient. D'un point de vue managérial, les directeurs d'hôpitaux et cadres de soin ont pour fonction d'optimiser les ressources dont ils disposent et ont la charge, qu'elles soient humaines (personnel soignant ou fonctions de support), matérielles (nombre de lits, matériel médical, médicaments...) en cherchant à garantir les trois valeurs précédemment citées tout en veillant à une valeur propre à leur fonction : l'*utilité*, c'est-à-dire la gestion et l'organisation de ces ressources avec efficacité.

Avec la révolution numérique, c'est-à-dire le fait de coder toute forme d'information par des nombres (Serres, 2012; Berry, 2010), la société s'est retrouvée bouleversée par le développement de technologies issues de l'informatique et par l'arrivée d'Internet. Parmi ces technologies, l'intelligence artificielle (ou « IA ») est sortie des laboratoires et des domaines d'application très spécialisés pour s'adapter à tout type de terrain. Ainsi, le rapport Villani (Villani, 2018), dans son focus sur « la santé à l'heure de l'IA », débute par ces propos enthousiastes : « *L'intelligence artificielle en santé ouvre des perspectives très prometteuses pour améliorer la qualité des soins au bénéfice du patient et réduire leur coût – à travers une prise en charge plus personnalisée et prédictive – mais également leur sécurité – grâce à un appui renforcé à la décision médicale et une meilleure traçabilité. Elle peut également contribuer à améliorer l'accès aux soins des citoyens, grâce à des dispositifs de prédiagnostic médical ou d'aide à l'orientation dans le parcours de soin.* »

3 L'intelligence artificielle démystifiée

Pour prendre la mesure de ces changements en cours et à venir dans le milieu médical, il convient de préciser ce qu'est l'intelligence artificielle, ce qu'elle peut ou ne peut pas faire, et, au-delà du domaine du possible, ce qu'il est souhaitable ou non qu'elle fasse.

Par définition, l'intelligence artificielle est une discipline de l'informatique qui a pour objet de simuler par ordinateur des processus de la pensée. L'IA se propose d'imiter sur une machine des comportements qui, si on les rencontre, sont qualifiés d'intelligents. Il existe ainsi de multiples approches sur l'intelligence artificielle (Russell et Norvig, 2010), schématiquement réparties suivant quatre domaines en fonction de l'objectif poursuivi :

1. des systèmes qui pensent comme des humains : la modélisation cognitive qui cherche à étudier la nature de la pensée humaine ;
2. des systèmes qui agissent comme des humains : les programmes qui sont capables de passer le test de Turing, c'est-à-dire de faire croire qu'il s'agit d'êtres humains et non de machines ;
3. des systèmes qui pensent rationnellement : héritière des travaux de Blaise Pascal ou Charles Babbage sur la pensée logique, cette approche voit la pensée comme une forme de calcul ;
4. des systèmes qui agissent rationnellement.

Alors que la première approche, celle de la modélisation cognitive, est désignée par l'expression « intelligence artificielle forte », la dernière, celle des agents artificiels rationnels, n'a ni la prétention ni l'objectif de s'intéresser à la pensée ou de prendre l'être humain comme modèle. C'est pourtant cette dernière approche, appelée « intelligence artificielle faible », qui a vraiment fait connaître l'IA au grand public à partir des années 2010, avec le superordinateur *Watson* d'IBM qui avait réussi à battre les champions humains du jeu *Jeopardy!* en 2011, ou le programme *AlphaGo* de *Google DeepMind* qui avait largement remporté de nombreuses parties du jeu de go face aux joueurs humains champions d'Europe en 2015 ou du monde en 2016.

Pour comprendre comment un programme est capable de réaliser une tâche aussi complexe que de jouer à un jeu intellectuel (les échecs, le jeu de go) mieux qu'un être humain ou d'identifier, à partir d'une image d'une tache sur la peau ou d'un grain de beauté, s'il s'agit de tumeurs malignes ou de taches bénignes, et ceci avec de meilleurs taux de succès que les professionnels de la santé, il est nécessaire de donner quelques notions de base sur la manière dont sont élaborés de tels programmes. Un programme n'est rien d'autre qu'un algorithme mis sous une forme compréhensible par une machine. Un algorithme est tout simplement le découpage d'une certaine opération en tâches élémentaires, comme une recette de cuisine qui permet d'élaborer un plat ou un dessert (une mousse au chocolat) à partir d'ingrédients (des œufs, du sucre, du beurre et du chocolat) présents en une certaine quantité suivant un certain ordre bien défini (casser les œufs, séparer les blancs des jaunes, monter les blancs en neige, faire fondre le chocolat avec le beurre, etc.) ou comme un procédé mathématique, telle que la résolution d'une équation du second degré qui nécessite le calcul du discriminant pour trouver les éventuelles racines solutions de l'équation.

Un programme ne fait rien d'autre que suivre les instructions codées par le programmeur : on lui fournit des valeurs en entrée (par exemple les valeurs a , b et c de l'équation

Enjeux éthiques de l'IA en santé

$ax^2 + bx + c = 0$) et le programme effectue des actions en retour (le calcul des racines solutions). La particularité d'un programme d'intelligence artificielle doté de ce que l'on appelle « l'apprentissage automatique » (ou *machine learning*) est de parvenir à changer certains paramètres en fonction des exemples qui lui sont présentés afin d'améliorer ses réponses. En présentant à certains programmes d'apprentissage automatique des millions et des millions d'images de chats et d'images sur lesquelles il n'y a pas de chat, le système parviendra de lui-même à extraire des caractéristiques pertinentes des images et à modifier ses paramètres afin de répondre avec justesse « chat » à des images apprises, mais aussi à effectuer une généralisation lui permettant de répondre correctement s'il y a ou non la présence de chat sur des images non apprises.

À l'heure actuelle, les programmes les plus efficaces sur ce genre de problème se basent sur des modèles appelés « réseaux de neurones artificiels » (parce que leur fonctionnement copie certaines caractéristiques des neurones biologiques organisés en réseaux) et sur un procédé dénommé « l'apprentissage profond » (ou *deep learning*), c'est-à-dire des systèmes d'apprentissage automatique qui reposent sur une architecture en de multiples couches entre l'entrée (par exemple les différents pixels d'une image) et la sortie (la décision « chat » ou « non chat ») avec des quantités considérables de connexions entre ces couches, ces paramètres de connexions se modifiant petit à petit lors de l'apprentissage d'une tâche donnée (c'est-à-dire forcer le système à répondre « chat » quand l'image d'un chat est bien présente, et pas autrement). Au-delà de la reconnaissance d'objets sur des images, ces systèmes d'apprentissage automatique vont s'appliquer à tout type de problème, pourvu qu'il y ait suffisamment de données étiquetées pour faire un apprentissage – c'est-à-dire que des exemples doivent être associés à une étiquette à apprendre fournies par un expert – en « digérant » des informations issues d'un très grand nombre de variables pour établir des corrélations entre ces variables, en faisant l'hypothèse d'un lien causal entre ces variables en cas de corrélation.

Ce procédé d'apprentissage – qui n'a rien de magique, il s'agit juste de mathématique – ne fonctionne pas à tous les coups. Des variables peuvent en effet être fortement corrélées entre elles, positivement ou négativement, pour d'autres raisons : on remarque que si on se réveille après avoir dormi avec ses chaussures aux pieds, on a très souvent mal à la tête. Il n'y a pas de lien causal à rechercher entre ces deux variables (est-ce que le fait de passer une nuit sans enlever ses chaussures aurait un rôle perturbateur dans la circulation du sang qui induirait des maux de tête au réveil?), ces deux événements étant simplement tous deux la conséquence d'un autre phénomène : s'endormir après avoir trop bu d'alcool (et donc ne plus être en mesure d'enlever ses chaussures en se couchant et avoir la gueule de bois au réveil). Néanmoins, les découvertes de liens possibles entre certaines variables prises parmi une multitude par des procédés automatiques a permis de créer des systèmes de reconnaissance ou des modèles prédictifs dont les performances n'avaient jusqu'alors jamais été atteintes : un changement quantitatif a permis de réaliser un véritable saut qualitatif.

Historiquement, ce changement n'a pu se faire qu'à travers des avancées dans trois domaines au cours des décennies 2000 et 2010 :

1. la disponibilité en données servant d'exemples lors de la phase d'apprentissage,
2. le développement de nouveaux algorithmes d'apprentissage capables de traiter de telles quantités de données,
3. la mise au point de matériel électronique (tels les processeurs graphiques) permettant d'effectuer les calculs nécessaires à cet apprentissage en un temps raisonnable.

Les programmes d'intelligence artificielle, bien que parvenant à résoudre certaines tâches complexes avec brio, ne font pas vraiment preuve d'intelligence et ne comprennent absolument pas ce qu'ils font (Dessalles, 2019). Malgré cela, il est possible de donner à une machine une notion du « sens » associé aux mots grâce à une technique appelée « le plongement lexical » (en anglais, *word embedding*). Cette technique permet de représenter chaque mot d'un dictionnaire par un vecteur de nombres réels à travers l'analyse statistique d'un grand nombre de textes afin de retrouver les termes qui apparaissent dans des contextes similaires et, à partir de là, d'en déduire des significations apparentées. Grâce à cette représentation vectorielle des mots, sur laquelle peuvent s'appliquer des opérations arithmétiques, il est possible de procéder à des raisonnements par analogie : par exemple, la représentation vectorielle du mot « roi » moins celle du mot « homme » plus celle du mot « femme » donne la représentation vectorielle du mot « reine ».

4 Les entreprises de l'intelligence artificielle

Les géants du numérique (en particulier les GAFAM³ et NATU⁴ américains ou les BATX chinois⁵), disposant de quantités phénoménales de données sur leurs utilisateurs (le *big data*), sont aussi les entreprises leaders dans le domaine de l'intelligence artificielle parce que l'efficacité d'un programme d'apprentissage dépend de la quantité et de la qualité des exemples qu'on lui fournit. Il n'est par conséquent guère surprenant de voir une entreprise comme *Google LLC/Alphabet Inc.*, par le biais de sa filiale *Calico*, chercher à occuper le terrain de la santé de façon disruptive en misant sur la convergence des nanotechnologies, des biotechnologies, de l'informatique et des sciences cognitives avec comme objectif de « tuer la mort ». Les données sont devenues un enjeu crucial dans le domaine de l'intelligence artificielle, et les croisements d'informations issues de sources multiples amènent à des questions majeures dans le domaine éthique au sujet de la préservation de la vie privée. À titre d'exemple, la société de biotechnologie *23andMe*, qui propose une analyse du code génétique aux particuliers, a pour cofondatrice celle qui fut l'épouse du cofondateur de *Google*... Quand on sait qu'*Alphabet Inc.* (la maison mère de *Google*), investissant massivement dans la recherche en intelligence artificielle (avec *DeepMind Technologies*), dispose déjà de tant de sources d'informations (historiques du moteur de recherche, e-mails de la messagerie *Gmail*, système d'exploitation mobile *Android* avec ses fonctions de géolocalisation, outils de bureautique en ligne *Google Docs*, *Sheets* ou *Slides*, etc.), il y a de quoi vraiment s'inquiéter si cette entreprise possède en plus d'un accès à nos informations génétiques !

5 IA, données et conséquences éthiques

En tant que cadre du milieu hospitalier, quelle attitude adopter si une entreprise privée nous propose un système permettant d'assurer une meilleure santé de nos patients en échange des données sur ces derniers ? Il y a là affrontement entre les valeurs de *santé* et de *vie privée*.

3. *Google, Apple, Facebook, Amazon* et *Microsoft*.

4. *Netflix, Airbnb, Tesla* et *Uber*.

5. *Baidu, Alibaba, Tencent* et *Xiaomi*.

L'intelligence artificielle et la robotisation amènent au remplacement de tâches humaines répétitives par des automates matériels ou logiciels. Les ressources humaines du système de santé seront fortement impactées par ce changement, et en particulier les fonctions supports (administration, gestion, finances, fonctions logistiques générales ou médicotechniques) (Gru-son et Kirchner, 2019). Sur le sujet des attentes et des impacts du développement de l'intelligence artificielle dans les hôpitaux, le cabinet de conseil EY et le CHRU de Nancy ont publié récemment le résultat d'une enquête qui fait ressortir que l'IA est très majoritairement considérée comme un enjeu par les directeurs et présidents de commissions médicales d'établissements de CHU, que l'IA suscite de fortes attentes des professionnels mais également des craintes et des interrogations, en particulier au sujet du risque de déshumanisation du travail et de la perte des liens sociaux (EY et CHRU Nancy, 2019).

Un autre risque est lié aux caractéristiques propres de ces systèmes artificiels. Les programmes d'intelligence artificielle améliorent leurs performances à partir d'un apprentissage statistique. Ils détectent les coïncidences dans les données apprises et les extrapolent sur des données non apprises. Cela a pour conséquence de renforcer les stéréotypes rencontrés implicitement dans les usages en cours. À la différence de l'image d'un chat qui présente une forme prototypique et inamovible de petit félin à la tête légèrement arrondie et aux oreilles triangulaires, la société évolue, les pratiques médicales changent. Les systèmes artificiels apprennent du passé pour prédire le futur, en tenant compte d'un très grand nombre de variables, sans indiquer les raisons explicites qui les ont amenés à conclure de telle ou telle manière. Le résultat d'un apprentissage effectué sur des données déséquilibrées (c'est-à-dire quand il y a une surreprésentation d'une situation par rapport à une autre) amènera un renforcement des situations les plus présentes, et ceci de manière caricaturale. Un programme d'IA d'aide au tri des candidatures employé par l'entreprise *Amazon*, qui avait effectué son apprentissage sur des CV reçus pendant une dizaine d'années, a dû être abandonné car il discriminait les candidates, qu'il orientait sur des postes de secrétariat, alors que les candidatures d'hommes étaient dirigées vers des postes de cadres. Aux États-Unis, dans le domaine juridique, des logiciels d'aide à la décision censés mieux informer les juges et homogénéiser la manière dont la justice est rendue ont aussi montré des biais raciaux, en indiquant des risques de récidive plus élevée ou en proposant des peines plus lourdes pour les citoyens afro-américains que pour ceux à la peau blanche. Des biais racistes ou sexistes émergent aussi des textes a priori neutres (des pages *Wikipedia* ou des articles de *Google Actualités*) : en utilisant la technique du plongement lexical et en faisant le calcul sur les représentations vectorielles des mots « médecin » – « homme » + « femme », on obtient la valeur du mot « infirmière » et non celle de « femme médecin »...

6 Conclusion

À l'heure où le système de santé français traverse une crise, l'introduction de techniques issues de l'intelligence artificielle doit être vue comme une opportunité. La révolution numérique amène son lot de mutations professionnelles et sociétales, tout comme le firent la mécanisation et l'automation industrielle. L'arrivée des robots et de l'intelligence artificielle doit être vue positivement, leur utilisation doit se faire au service de l'humain, et les choix de valeurs à défendre sont de véritables enjeux civilisationnels. Alors que la vision des États-Unis est de considérer que les données sont la propriété des entreprises qui les collectent à des fins commerciales, que celle de la Chine est d'employer les données de ses citoyens à des fins de

contrôle à travers le système de crédit social, la France et l'Union européenne voient dans les données de la vie privée une ressource à protéger, comme le montre l'adoption et l'application du Règlement général sur la protection des données (RGPD).

Les directeurs des soins et cadres de santé ne doivent pas déroger à leurs fonctions managériales dédiées aux valeurs essentielles de *vie*, de *santé* et de *bien-être* des patients, en recherchant l'*utilité* dans la manière dont ils gèrent les ressources dont ils ont la charge, mais ils doivent aussi être les garants de la protection d'autres valeurs qui risquent d'être menacées par la révolution numérique : la préservation de la *vie privée* et des données sensibles des patients couvertes par le secret médical, la *compétence* dans l'usage de ces technologies numériques, la *confiance* dans les systèmes d'aide à la décision apportés par l'IA, la *connaissance* par la compréhension des raisons qui ont amené le programme d'IA à proposer tel ou tel résultat et que celui-ci pourra ou non suivre en toute conscience (transparence des algorithmes, impartialité des décisions, non-discrimination), et le *pouvoir* en bannissant les approches prescriptives et en laissant la maîtrise à l'utilisateur humain afin de demeurer un acteur éclairé et maître de ses choix (que ce soit le consentement du patient ou la validation par un expert humain d'un prédiagnostic posé par un programme d'IA).

Pour un directeur des soins ou cadre de santé, ce n'est qu'à travers la combinaison de la bonne connaissance de ce qu'est l'outil IA, de ses possibilités et de ses limites, et de son engagement dans le respect des valeurs fondamentales du patient, que l'intelligence artificielle pourra être employée à des fins d'une meilleure humanisation du parcours de soin.

Références

- Berry, G. (2010). *La numérisation du monde*. Paris : Éditions De Vive Voix.
- Deonna, J. et E. Tieffenbach (Eds.) (2018). *Petit Traité des valeurs*. Collection « Science & métaphysique ». Paris : Éditions d'Ithaque.
- Dessalles, J.-L. (2019). *Des intelligences TRÈS artificielles*. Paris : Éditions Odile Jacob.
- EY et CHRU Nancy (2019). Baromètre de maturité de l'IA dans les CHU.
http://www.chru-nancy.fr/images/colloqueIA/1911SG412_Barometre_AI_CHU_Nancy-VF3.pdf.
- Gruson, D. et C. Kirchner (2019). Éthique, numérique et intelligence artificielle : quels enjeux pour les cadres de santé ? *Soins Cadres* 28(115), 47–49.
<https://doi.org/10.1016/j.scad.2019.09.021>.
- Russell, S. J. et P. Norvig (2010). *Intelligence artificielle* (3 ed.). Montreuil : Pearson France.
- Serres, M. (2012). *Petite Poucette*. Collection « Manifestes ». Paris : Éditions Le Pommier.
- Villani, C. (2018). Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne. Mission confiée par le Premier Ministre Édouard Philippe, Mission parlementaire du 8 septembre 2017 au 8 mars 2018.
https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf.