



HAL
open science

Technologies numériques, intelligence artificielle et responsabilité

Christine Balagué

► **To cite this version:**

Christine Balagué. Technologies numériques, intelligence artificielle et responsabilité. 28èmes Rencontres Santé Société Georges Canguilhem (Euro Cos Humanisme & Santé), Oct 2019, Strasbourg, France. hal-02907065

HAL Id: hal-02907065

<https://hal.science/hal-02907065>

Submitted on 2 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Rencontres EURO COS & Humanisme Santé, 10 octobre 2019

Technologies numériques, intelligence artificielle et responsabilité

Christine Balagué

Professeur et Titulaire de la Chaire Good in Tech

Institut Mines-Télécom Business School

Les innovations technologiques dans le secteur de la santé permettent des progrès considérables, mais le développement des objets connectés, des robots ou des algorithmes utilisant des méthodologies d'apprentissage machine de plus en plus complexes, est source d'impacts sociétaux majeurs. Quels sont donc les enjeux éthiques d'une médecine s'appuyant sur des technologies d'intelligence artificielle ? Comment repenser l'humain au cœur de la e-santé ? Comment concevoir des technologies responsables et éthiques ? Quelles sont les courants de pensée actuels et les actions menées pour penser une médecine du futur responsable ?

E-santé et médecine s'appuyant sur des données : une réalité

La e-santé et la médecine orientée par les données n'évoquent plus un monde futur et lointain, mais une réalité aujourd'hui. Une multitude d'objets connectés permettent à partir de capteurs de collecter des données sur les patients. Pour certaines maladies chroniques comme le diabète, ces dispositifs connectés améliorent significativement la vie des patients, en apportant des mesures plus régulières (par exemple sur leur taux de glycémie) leur permettant de mieux gérer leur maladie et de partager ces données avec des professionnels de santé. Pour d'autres, ces objets connectés leur rappellent constamment la maladie chronique alors qu'ils préféreraient plutôt l'oublier le plus possible dans leur vie quotidienne, ou suscitent également plus de stress pour le patient et son entourage. L'ambivalence de l'impact de ces technologies pose donc des questions éthiques aujourd'hui. Parallèlement à l'essor des dispositifs connectés médicaux, on assiste également au développement d'algorithmes prédictif à l'origine d'une médecine personnalisée, tant sur le diagnostic que sur une recommandation de traitement. Dans certaines applications déjà présentes dans certains hôpitaux, les questions éthiques sont importantes. Par exemple, dans certains services de cardiologie aux Etats-Unis, un algorithme calcule, à l'arrivée du patient, la probabilité que celui-ci meure dans l'heure qui suit. En fonction du résultat de l'algorithme, le patient est traité prioritairement ou non. Or, tout algorithme présente des taux d'erreur, les deux principaux étant le taux de faux positif (le patient est diagnostiqué comme devant être traité en priorité mais c'est faux) ou de faux négatif (le patient n'est pas diagnostiqué comme devant être traité en priorité et c'est faux). Il semble clair que ces erreurs, en particulier l'existence des faux négatifs, pose dans ce cas des enjeux majeurs d'éthique. Ceci, l'utilisation d'algorithmes fondés sur des techniques complexes d'apprentissage machine (machine learning ou deep learning) sont aussi à

l'origine de progrès médicaux considérables. Par exemple, le séquençage du génome permet de détecter des marqueurs génétiques prédictifs de maladies, et d'adapter en conséquence la vie de certains individus ou de faire de la prévention. Les algorithmes de traitement d'images permettent d'individualiser les doses de radiothérapie dans le cadre de certains cancers. D'autres technologies numériques impactent le secteur de la santé. Les robots chirurgiens dans les salles d'opération, grâce à leur précision et à leur résistance (un robot est rarement fatigué alors que le chirurgien peut l'être), sont de plus en plus répandus. Les chatbots, machines répondant automatiquement à des patients, que ce soit par texte ou par voix, sont opérationnels, comme celui développé par la NHS en Angleterre pour répondre aux urgences médicales téléphoniques. Les serious game (ou jeux sérieux) ou encore les simulations sont intégrées dans la formation des étudiants en médecine (Paris Descartes par exemple). La télémédecine est en cours de déploiement en France, tandis que se développent des robots empathiques pour l'accompagnement des personnes âgées à domicile. L'ensemble de ces technologies numériques posent des enjeux éthiques majeurs en 2019.

Qu'est-ce-que l'éthique en e-santé ?

L'éthique en e-santé repose sur le respect de quatre principes fondamentaux. Ces principes sont ceux partagés et défendus en Europe, mais ne s'appliquent pas forcément à des cultures en dehors de l'Union Européenne, comme en Chine ou aux Etats-Unis par exemple. Le premier principe, qui relève de la macro-éthique, est le principe de justice, ou encore d'équité. Appliqué aux technologies, ce principe signifie que les technologies numériques en santé ne devraient pas générer d'injustice ou d'inégalités entre les patients. Le deuxième principe (qui relève comme les deux suivant de la micro-éthique), est le principe d'autonomie, ou encore du respect de la vie des individus. Le troisième est le principe de bienfaisance : les technologies devraient apporter une utilité, une valeur ajoutée, et être responsable. Enfin, un quatrième principe est le principe de non malfaisance. Il ne suffit d'apporter des bénéfices (principe de bienfaisance), il est également nécessaire de ne pas être mal faisant, donc de ne pas générer des effets négatifs, que ce soit sur les individus comme sur la société plus globalement. Cette exigence de non malfaisance génère parfois l'application du principe de précaution : on ne va pas déployer la technologie car ses effets peuvent potentiellement être négatifs, ou on ne sait pas si les effets peuvent être négatifs.

A l'ère de la e-santé et de la médecine des données, ces quatre principes peuvent être menacés, en particulier par quatre enjeux. Le premier porte sur le biais des données et des algorithmes, à l'origine de potentiels effets de discrimination, qui enfreindrait le principe d'équité entre les individus. Le deuxième est lié aux effets de bulles d'enfermement des algorithmes, qui limiterait l'autonomie de l'individu. Le troisième concerne l'opacité du fonctionnement des technologies numériques, en particulier les algorithmes, qui interrogerait la question de l'utilité et de la valeur ajoutée, ainsi qu'une faible garantie sur la non malfaisance. Enfin, au sein des algorithmes sont encapsulées des opinions, l'application des technologies algorithmiques posant alors des interrogations concernant les prises de décision fondées sur les résultats de ces systèmes.

Les biais des algorithmes

L'un des potentiels biais des algorithmes est le biais des données utilisées au départ. Plusieurs caractéristiques de ces données peuvent générer des résultats d'algorithmes biaisés ou discriminant, posant donc problème par rapport au principe d'équité et de justice. Les données peuvent en effet être de mauvaise qualité, incomplètes et ne pas représenter la population souhaitée, elles peuvent également être vulnérables, c'est-à-dire peu sécurisées, ce qui dans le cas des données de santé peut avoir un impact négatif sur certains patients. Un autre enjeu porte sur l'origine des données permettant la phase d'apprentissage de l'algorithme. Le courant actuel des sociologues sur le « digital labor » dénonce un travail de codage des données effectué par des populations dans des pays en développement, très peu payées. Un rapport récent de France Stratégie a montré que le phénomène existait aussi en France.

Les effets de discrimination des algorithmes ont été montrés dans de nombreuses publications, dont une récente dans Sciences montrant que les minorités noires aux Etats Unis sont discriminées relativement aux minorités blanches par les algorithmes calculant l'accès aux soins. Dans un autre domaine que la santé, le recrutement, les équipes du MIT ont montré une discrimination des femmes lorsqu'un algorithme est utilisé.

Une fracture numérique chez les patients

Ces effets potentiellement négatifs des technologies numériques et des algorithmes dans le domaine de la santé sont accentués de l'accroissement de la fracture numérique des usages au sein des patients. Une recherche que nous avons menée en 2018 avec ICA ? collectif de 14 associations de malades chroniques, nous a permis de croiser l'usage des technologies avec des variables de gestion de la maladie, et d'identifier ainsi trois groupes de malades chroniques. Le premier groupe largement majoritaire (72%) est *hypoconnecté*, à savoir utilisateur moins de trois fois/mois d'internet, quasiment pas utilisateur d'applications mobiles ou d'objets connectés. Le deuxième groupe, les *biconnectés*, représente 19% des malades chroniques, et utilisent principalement l'internet et les applications mobiles. Seuls 9% des malades chroniques sont *hyperconnectés*, utilisant régulièrement les trois technologies dans la gestion de leur pathologie. Cette recherche dévoile une grande fracture numérique au niveau des usages, et donc une plus grande fragilité des patients dans un monde où la e-santé et la médecine des données devient une réalité.

Opacité des algorithmes et risques de manipulation

Le deuxième enjeu majeur des technologies numériques en santé porte sur l'opacité des algorithmes et la faible compréhension de leur fonctionnement, que ce soit par les professionnels de santé ou par les patients. Ce problème est d'autant plus important que des grands acteurs de l'internet, entreprises privées, ont un développement exponentiel sur le marché de la santé. Google vient à ce propos d'annoncer un nouveau partenariat avec plus de cent cinquante hôpitaux américains avec accès aux données de leurs patients (sans

consentement éclairé de ceux-ci, mais le Règlement Européen des Données n'est pas en vigueur aux Etats-Unis). L'opacité des algorithmes génère une faible garantie sur la bienfaisance ou la non malfaisance de la technologie. Mais l'opacité des algorithmes pose également des enjeux sur l'impact potentiel sur les patients. Par exemple, les tests génomiques révélant des marqueurs génétiques de plus forte probabilité d'avoir un cancer (par ex la présence de BRCA1 ou BRCA2 pour le cancer du sein) sont de plus en plus pratiqués. Cependant, ces tests fournissent des probabilités et le patient est désormais caractérisé par des probabilités d'avoir telle ou telle maladie, et l'effet épée de Damoclès peut profondément impacter la vie quotidienne du patient et ses relations avec son entourage. Enfin, un algorithme peut présenter des erreurs (faux positifs et faux négatifs) qu'il faut savoir expliquer aux patients. Par rapport à ces enjeux, un courant pluridisciplinaire international de chercheurs travaille activement à l'explicabilité, l'interprétabilité, la redevabilité voire la transparence des algorithmes.

Enfermement des individus et perte d'autonomie

Un troisième effet potentiel des algorithmes se nomme les bulles d'enfermement. Dans le domaine de la santé, les données d'objets connectés sont généralement utilisées dans des algorithmes de recommandations (d'activité physique, d'alimentation, de gestion de la glycémie etc.). Ces algorithmes de recommandations ont un double effet : d'un côté ils permettent aux patients de suivre régulièrement un critère important pour eux (ex : le taux de glycémie chez les diabétiques), mais ils enferment d'un autre côté les patients dans leur choix d'activités, d'action, en limitant leur autonomie ou leur choix de pensée, entraînant parfois une certaine déresponsabilisation du patient qui s'en remet à la machine.

Prise de décision médicale fondée sur des algorithmes comportant des opinions

Un algorithme est un programme informatique qui comprend des opinions et des choix de son ou ses concepteurs. Par exemple, les machines d'imagerie médicale dont l'objectif est de détecter des parties du corps comportant une pathologie, sont généralement conçues par des ingénieurs en sciences informatiques. Dans leur domaine, leur priorité peut être de faire avancer la science et donc de minimiser les faux positifs de l'algorithme de reconnaissance d'image. Un médecin au contraire a plutôt pour objectif de minimiser les faux négatifs, à savoir des patients détectés non porteurs d'un cancer (si l'algorithme vise à détecter une zone cancérogène) par l'algorithme, alors qu'ils ont un cancer qu'il faut soigner. D'autres algorithmes sont porteurs d'opinion ou de norme sociale. Par exemple, les assistants vocaux (de type Alexa ou Google Home) utilisent majoritairement des voix féminines, portant une norme sociale selon laquelle la femme (et pas l'homme) serait toujours au service des autres.

Comment adresser les questions d'éthique et de responsabilité des technologies en santé ?

Les technologies numériques en santé présentent aujourd'hui une double vision. D'un côté ces technologies sont source de progrès considérables : elles améliorent le parcours patient, elles permettent une aide à la décision des médecins pour à la fois le diagnostic ou le traitement, elles sont source de précision au travers des robots chirurgiens, elles sont source d'incroyables avancées en recherche médicale (le séquençage du génome par exemple), elles permettent également de mieux accompagner des personnes âgées atteintes de pathologies à domicile.

Cependant ces technologies présentent des risques, notamment sur la protection des données personnelles, sur les biais des algorithmes, leur opacité, les bulles d'enfermement, ou encore les opinions encapsulées par les algorithmes.

Face à ces enjeux, plusieurs mouvements se sont lancés. En novembre 2018, l'Université de Montréal a lancé la déclaration de Montréal, prônant la nécessité de développer des technologies d'intelligence artificielle responsables et éthiques. Dans le même objectif, l'OCDE a également travaillé sur les critères d'une intelligence artificielle éthique. L'Unesco a également lancé ses groupes de travail sur le même thème, avec un objectif de formation dans le monde des populations à ces enjeux.

Au niveau européen, la commission européenne a créé un « High Level Group on Artificial Intelligence », qui a publié fin 2018 un guide éthique sur l'intelligence artificielle autour de dix principes : redevabilité, gouvernance des données, design pour tous, gouvernance de l'autonomie de l'intelligence artificielle, non-discrimination, respect pour l'autonomie de l'humain, respect des données personnelles, robustesse, sécurité et transparence des algorithmes.

En France, la loi pour une république numérique de 2016 a instauré dans son article 49 l'obligation de loyauté des plateformes numériques, et la transparence des algorithmes publics dans son article 4.

Dans le domaine de la santé, le Comité Consultatif National d'Éthique (CCNE), qui dépend du premier ministre et qui a travaillé sur les nouvelles lois de bioéthique, a rendu avec la Cerna (Comité d'Éthique de la Recherche sur le Numérique d'Allistène) le rapport Numérique et Santé, sur les enjeux des technologies numériques en santé, qui souligne la nécessité d'une garantie humaine. En décembre 2019, un comité d'éthique issu du CCNE a été créé pour travailler sur ces enjeux.

Le rapport du Conseil National de l'Ordre des médecins de janvier 2018 va dans le même sens et prône le développement de technologies numériques en santé responsables et éthiques, affirmant que les décisions issues de la machine ne remplaceront pas la décision médicale, qu'il est nécessaire d'anticiper l'impact des technologies sur les plans éthique, social et légal, et lançant un appel pour un label des technologies qui seraient évaluées sur leur niveau d'éthique.

La CNIL avait déjà affirmé une vision similaire dans son rapport de 2017, en émettant six recommandations :

1. Former à l'éthique tous les acteurs-maillons de la « chaîne algorithmique » (concepteurs, professionnels, citoyens) : l'alphabétisation au numérique doit permettre à chaque humain de comprendre les ressorts de la machine
2. Rendre les systèmes algorithmiques compréhensibles en renforçant les droits existants et en organisant la médiation avec les utilisateurs
3. Travailler le *design* des systèmes algorithmiques au service de la liberté humaine, pour contrer l'effet « boîtes noires »
4. Constituer une plateforme nationale d'audit des algorithmes
5. Encourager la recherche sur l'IA éthique et lancer une grande cause nationale participative autour d'un projet de recherche d'intérêt général
6. Renforcer la fonction éthique au sein des entreprises (par exemple, l'élaboration de comités d'éthique, la diffusion de bonnes pratiques sectorielles ou la révision de chartes de déontologie peuvent être envisagées)

Enfin, nous avons lancé en 2019 la chaire de recherche Good in Tech. Cette Chaire, portée par Institut Mines-Télécom Business School et l'école de Management et d'Innovation de Sciences Po, en partenariat avec Telecom Paris et Telecom Sud Paris, regroupe des chercheurs pluridisciplinaires dans le cadre d'un programme de recherche visant à repenser l'innovation et les technologies comme moteurs d'un monde meilleur pour et par l'humain. Il s'articule autour de quatre axes de recherche :

Axe 1 : innovation numérique responsable : quelles mesures ?

Quelles sont les dimensions et mesures de l'innovation numérique responsable ?

Comment intégrer l'innovation numérique responsable dans la responsabilité sociale numérique des entreprises (RSE 4.0) ?

Axe 2 : comment développer des technologies responsables « by design » ?

Comment éviter les biais liés aux données ?

Comment développer des technologies responsables by design (plus interprétables, explicables, redevables, équitables et respectant la régulation sur les données personnelles) ?

Axe 3 : réinventer les futurs : quelle société pour demain dans un monde numérique ?

Comment créer une société numérique acceptable pour tous ?

Comment réinventer les futurs dans une perspective fidèle aux Lumières, de préservation du principe d'égalité appliqué au monde connecté ?

Axe 4 : gouvernance de l'innovation et des technologies responsables

Quels sont les mécanismes de gouvernance de l'innovation numérique responsable ?

Quelles sont les niveaux pertinents de cette gouvernance : Europe, Nation, Entreprise ?

En conclusion, l'ensemble de ces initiatives semblent tout à fait encourageantes pour mieux repenser les enjeux de responsabilité et d'éthique des technologies numériques dans le secteur de la santé, et les prochaines années seront riches en avancées pour une réflexion collective sur le monde futur acceptable et désirable que nous souhaitons construire dans une société connectée.