



ENJEUX ÉTHIQUES DE L'IA EN SANTÉ

Joé T. Martineau, Frédérique Romy Godin,
Janine Badr, Alexandre Castonguay, Martin Cousineau,
Philippe Després, Aude Motulsky, Jean Noel Nikiema,
Cécile Petitgand, Catherine Régis

Québec, Canada
Mars 2022

Mandat réalisé par un groupe de travail de l'Axe Santé durable de l'OBVIA pour la Direction de la planification stratégique et de la performance (DGPS) du MSSS



**OBSERVATOIRE INTERNATIONAL
SUR LES IMPACTS SOCIÉTAUX
DE L'IA ET DU NUMÉRIQUE**



**OBSERVATOIRE INTERNATIONAL
SUR LES IMPACTS SOCIÉTAUX
DE L'IA ET DU NUMÉRIQUE**

L'OBVIA est un réseau de recherche ouvert qui contribue aux réflexions entourant l'utilisation responsable de l'IA et du numérique en favorisant les discussions et le maillage entre les communautés de pratique, la société civile et les décideurs publics. Soutenu financièrement par les Fonds de recherche du Québec (FRQ), il fédère les expertises de plus de 260 chercheur-es couvrant les secteurs des sciences humaines et sociales, des sciences et génie et de la santé. Il bénéficie également de l'appui de plus de 125 partenaires issus de centres de recherche et d'organisations publiques et privées du Québec, du Canada et de l'étranger.

Son objectif est de mettre en lumière les enjeux cruciaux soulevés par les développements et l'utilisation de l'IA et du numérique et d'identifier des solutions plausibles aux problèmes et opportunités qu'ils soulèvent.

L'OBVIA réalise des travaux de recherche et des mandats d'intervention en partenariat en vue de développer de nouvelles connaissances, de les diffuser et d'aider les organisations québécoises à maximiser les impacts positifs du numérique et de l'IA dans nos sociétés. Ces travaux sont priorisés en fonction des orientations scientifiques de l'OBVIA et visent à répondre à la fois aux objectifs des partenaires et des équipes de recherche impliquées. Ils peuvent prendre plusieurs formes, comme des études sur le terrain, des synthèses et analyses de la littérature, la conception d'ateliers ainsi que le soutien dans l'élaboration, l'adoption et la mise en œuvre de bonnes pratiques en matière d'enjeux touchants à l'IA et au numérique.

**Observatoire international sur les
impacts sociétaux de l'intelligence artificielle et du numérique**

Pavillon Charles-De Koninck, local 2489
1030, avenue des Sciences-Humaines
Université Laval
Québec (Québec) G1V 0A6

observatoireia@ulaval.ca
observatoire-ia.ulaval.ca

ISBN : 978-2-925138-17-4

Liste des abréviations

ACMTS	Agence canadienne des médicaments et des technologies de la santé
ANS	Agence nationale de santé (France)
BCH	Bureau de la certification et de l'homologation
C.c.Q.	Code civil du Québec
CDPJ	Commission des droits de la personne et des droits de la jeunesse
CHUM	Centre hospitalier de l'Université de Montréal
CNN	Conseil canadien des normes
DIM	Direction des instruments médicaux
DRI	Direction des ressources informationnelles
DSQ	Dossier santé Québec
ECG	Électrocardiographie
EEG	Électroencéphalographie
ETS	Évaluation des technologies de santé
FDA	Food and Drug Administration
GDHP	Global Digital Health Partnership
GPUs	Graphics Processing Units (Unités de traitement graphique)
IA	Intelligence artificielle
INR	Institut national de recherche
ISO	Organisation internationale de normalisation
LRDPDE	Loi sur la protection des renseignements personnels et des documents électroniques
MHRA	Medicines and Healthcare Products Regulatory Agency
NASA	National Aeronautics and Space Administration
NHS	National Health System (Royaume Uni)
NLP	Natural Language Processing (Traitement automatique de la langue)
OCDE	Organisation de coopération et de développement économique
OMS	Organisation mondiale de la santé
QALYs	Quality-adjusted life years (Année de vie ajustée pour la qualité)
SIA	Systèmes d'intelligence artificielle
TGV	Trousse globale de vérification
TRL	Technology Readiness Level (Niveau de maturité technologique)
UE	Union Européenne
VIH	Virus de l'immunodéficience humaine

Table des matières

Introduction	5
Démarche	6
1. Principaux enjeux éthiques liés à l'exploitation des données massives	7
1.1 Enjeux liés au consentement à la collecte et au partage des données personnelles.....	8
1.2 Enjeux liés à la confidentialité et à la protection de la vie privée des personnes	9
1.3 Enjeux de fiabilité, de qualité et de représentativité des données	10
1.4 Enjeux liés à l'archivage et la conservation des données.....	11
2. Principaux enjeux éthiques liés au développement des systèmes d'IA en santé	12
2.1 Utilisation non-éthique des SIA.....	12
2.2 L'explicabilité des SIA	13
2.3 La responsabilité décisionnelle	14
2.4 Les biais et la discrimination algorithmique.....	14
2.5 Les enjeux liés au développement d'agents moraux artificiels	15
2.6 Les enjeux existentiels de l'IA	15
2.7 Les enjeux liés au futur du travail et à la transformation des emplois en santé	16
2.8 L'accès variable aux technologies de l'IA et l'accroissement des injustices sociales par l'IA	16
2.9 Les enjeux liés à commercialisation des technologies de l'IA en santé.....	17
2.10 Les impacts environnementaux de l'IA	17
3. Principales lignes directrices en éthique de l'IA	18
3.1 La Déclaration de Montréal pour le développement responsable de l'intelligence artificielle	18
3.2 Les Principes directeurs de l'OMS sur l'éthique et la gouvernance de l'IA en santé	20
Pour en savoir plus	21
Références	22

INTRODUCTION

Les applications d'intelligence artificielle en santé offrent de nombreux bénéfices potentiels, mais génèrent aussi des risques et soulèvent plusieurs enjeux éthiques. En effet, l'évolution rapide des technologies d'intelligence artificielle, l'incertitude quant à leurs impacts sur les individus et sur nos sociétés, ainsi que le manque de réglementation les encadrant, forcent une réflexion profonde sur les enjeux éthiques que le développement de systèmes d'intelligence artificielle (SIA) soulève, ainsi que sur notre positionnement face à leur utilisation.

La présente fiche propose une revue des différents enjeux éthiques liés au développement et à l'utilisation des technologies d'intelligence artificielle dans le milieu de la santé, en trois parties. D'abord, nous aborderons les enjeux éthiques liés à l'exploitation de données massives nécessaires à l'entraînement des algorithmes de l'IA. Ensuite, nous présenterons les principaux enjeux éthiques liés au développement et à l'utilisation des SIA en santé, en abordant la façon dont ces systèmes impactent nos vies ainsi que l'environnement physique et social dans lequel nous vivons. Nous présenterons finalement les principales initiatives nationales et internationales en matière d'éthique de l'IA et de la gestion des données, fruits et reflets d'une réflexion globale sur ces sujets. Ces initiatives ont notamment proposé des lignes directrices et principes normatifs servant de guides pour le développement de technologies de l'IA éthiques et responsables.



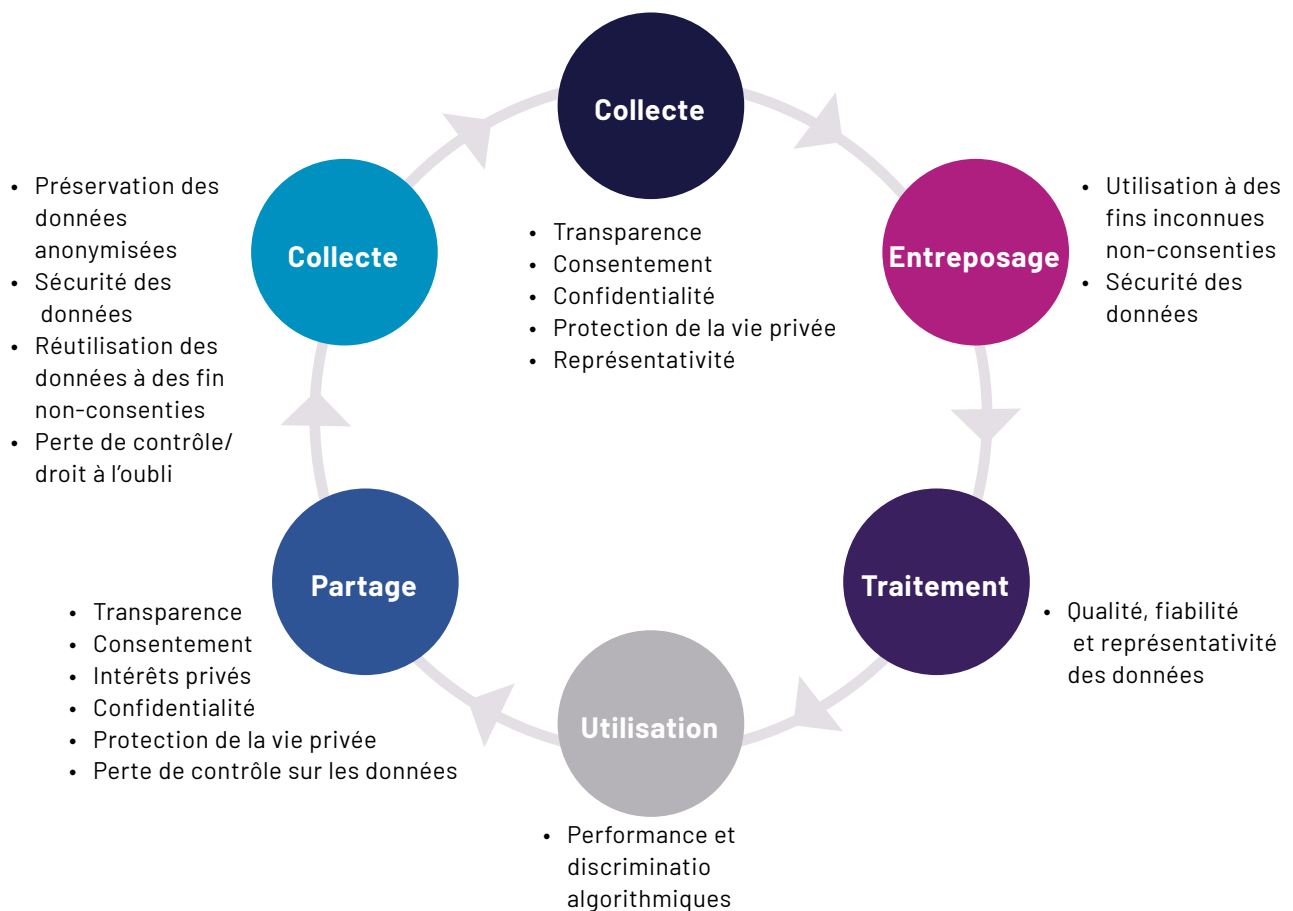
DÉMARCHE

Pour construire cette fiche, nous avons exploré la littérature pertinente sur l'éthique de l'IA et la gouvernance éthique des données dans le domaine de la santé dans le but de répondre à la question : « À quels enjeux éthiques devons-nous porter attention dans l'intégration de l'IA en santé ? ». Nous ne prétendons pas que cette revue des enjeux soit exhaustive. Les technologies d'IA et leurs applications en santé évoluant rapidement, les enjeux éthiques sont aussi amenés à évoluer et à se diversifier avec le temps. Pour construire cette revue, nous avons exploré la littérature scientifique et la littérature grise (rapports gouvernementaux et d'organismes reconnus) sur l'éthique de l'IA et sur l'éthique des données, en ayant comme objectif général de fournir aux lecteurs un portrait d'ensemble des enjeux touchant ces technologies, et pour en exposer les principes reconnus et les principales lignes directrices en la matière. Nous débutons la fiche en traitant des enjeux éthiques liés à l'exploitation des données massives. Ces enjeux comprennent notamment le consentement à la collecte et au partage des données personnelles, la confidentialité et la protection de la vie privée des personnes, la qualité et la fiabilité des données, ainsi que les risques liés à l'archivage et la conservation des données. Nous poursuivons notre aperçu avec une deuxième section englobant les enjeux liés au développement des SIA en santé. Cette deuxième section traite de l'utilisation non-éthique des SIA, de leur manque d'explicabilité, de l'enjeu de la responsabilité décisionnelle, des biais et de la discrimination algorithmique, du développement d'agents moraux artificiels, des enjeux existentiels de l'intelligence artificielle, du futur du travail et de la transformation des emplois, de l'accès variable aux technologies de l'IA et de l'accroissement des injustices sociales lié à l'utilisation des SIA, de la commercialisation des technologies de l'IA et, enfin, des impacts environnementaux de celles-ci. Finalement, nous bouclons la fiche avec une troisième section portant sur les principales lignes directrices en éthique de l'IA, offrant un aperçu sur deux initiatives reconnues : la Déclaration de Montréal pour le développement responsable de l'intelligence artificielle et les principes directeurs de l'OMS sur l'IA en santé.

1 PRINCIPAUX ENJEUX ÉTHIQUES LIÉS À L'EXPLOITATION DES DONNÉES MASSIVES

Le développement et le déploiement des technologies d'intelligence artificielle, dont notamment celles mobilisant l'apprentissage automatique, nécessitent la collecte et l'exploitation d'une énorme quantité de données, souvent à juste titre nommées « données massives » ou « *big data* ». Ces données sont nécessaires pour entraîner les algorithmes utilisés par les différents systèmes d'intelligence artificielle (SIA). De ce fait, les données sont donc devenues des ressources précieuses et convoitées par différents acteurs publics et privés, poursuivant des objectifs divers. De nombreux enjeux éthiques liés à l'IA en santé sont donc liés à la nature même de ces données, à la façon dont elles sont collectées, entreposées, utilisées, partagées, etc. Nous présentons ci-dessous les principaux enjeux éthiques associés à la gouvernance des données, résumés dans la Figure 1 ci-dessous selon les étapes du cycle de vie des données.

Figure 1. Enjeux éthiques et cycle de vie des données



1.1 Enjeux liés au consentement à la collecte et au partage des données personnelles

Plusieurs des enjeux éthiques liés au déploiement de l'IA dans le domaine de la santé trouvent leur source dans la façon dont les données sur les individus sont collectées. D'abord, ce qu'on entend par « données personnelles » dans le domaine de l'IA en santé est souvent beaucoup plus large que l'idée que la population générale s'en fait intuitivement. Au-delà des données biomédicales et/ou cliniques comprises dans le dossier médical (les renseignements de santé), les SIA en santé peuvent aussi exploiter une panoplie de données de la vie quotidienne des individus, telles que des données d'entraînement physique, de nutrition, de sommeil et d'habitudes de vie qui proviennent bien souvent d'objets connectés de la vie quotidienne. Téléphones intelligents, montres et autres appareils d'entraînement physique connectés, pour ne nommer que ceux-là, collectent des données de santé sur les individus en temps réel, données qui sont ensuite converties en données « médicales » et analysées pour en dériver des informations sur la santé globale des personnes (OMS, 2021). Comme les sources de données se multiplient, les acteurs impliqués dans la collecte de ces données se diversifient aussi. Ceux-ci peuvent inclure des organisations publiques du milieu de la santé, mais aussi les entreprises privées qui développent les technologies utilisées par le système de la santé ou encore les objets connectés précédemment mentionnés (Someh et al., 2019). Les différents moyens de collecte des données personnelles, ainsi que leur partage entre acteurs publics et privés, sont rarement compris et anticipés par les individus, tout comme la quantité de données qui sont collectées par ceux-ci. En effet, les différentes organisations développant et utilisant les technologies d'IA déployées en santé collectent souvent beaucoup plus de données que nécessaires, parfois anticipant des utilisations commerciales ultérieures comme la vente de données, qui sont inconnues des individus auprès de qui ces données sont initialement collectées (Hand, 2018).

Bien que l'obtention du consentement à la collecte et au partage des données personnelles dans le cadre du système de santé québécois soit de mieux en mieux balisé et encadré (voir la Fiche 3, « IA en santé : aspects juridiques », section 3a), le consentement à la collecte et au partage des données personnelles par les applications de santé et objets connectés est dans de nombreux cas obtenu de façon trop rapide, voire parfois fallacieuse (Grundy et al., 2019). Dans les faits, l'utilisateur est souvent amené à cocher aveuglément une case dans une application ou sur un site web, sans avoir nécessairement bien compris les implications de ce consentement et sans s'être fait expliquer, dans des mots clairs et accessibles, les tenants du « contrat » qu'il signe en cochant cette case. Et lorsque l'utilisateur s'attarde à lire les formulaires de consentement et contrats de confidentialité des données personnelles des entreprises privées commercialisant des objets connectés ou applications d'IA, il se perd souvent dans le jargon juridique inaccessible et opaque de ces très longs documents, qui échappent à la compréhension du commun des mortels (Andrejevic, 2014; Hand, 2018).

Ces pratiques et le manque de transparence avec lequel elles sont communiquées remettent en question la notion même de consentement libre et éclairé et entravent l'autonomie des individus. Valeur fondamentale en santé, l'autonomie désigne la capacité d'un individu de décider pour lui-même, et ce, en connaissance de cause, ainsi que d'agir à partir de sa propre volonté et sans contrainte (Calvo et al., 2020; Université de Montréal, 2018). Le consentement libre et éclairé a donc pour objectif de favoriser l'autonomie des personnes, en les protégeant eux-mêmes, et par extension en protégeant leurs données, contre des actes qui seraient faits contre leur gré ou à leur insu. Le consentement libre et éclairé requiert, d'une part, une compréhension adéquate de la part de l'individu concerné de ce à quoi il consent, et d'autre part, une absence de contrainte à consentir (Porsdam Mann, Savulescu et Sahakian, 2016), ce qui est largement entravé par les pratiques dépeintes plus haut.

1.2 Enjeux liés à la confidentialité et à la protection de la vie privée des personnes

Les pratiques de collecte et de partage des données précédemment mentionnées peuvent considérablement entraver la confidentialité et la protection de la vie privée des individus concernés. La protection de la confidentialité correspond à la revendication des individus, des groupes ou des institutions à déterminer eux-mêmes quand, comment et dans quelle mesure les informations les concernant sont communiquées à d'autres (Mai, 2016), et implique que les individus maintiennent un contrôle sur et un choix concernant leurs données personnelles. Il en est de même pour la vie privée qui, elle, réfère à la protection de zones qui échappent au regard d'autrui, ainsi qu'à la limitation et au contrôle de l'accès aux informations qui concernent les individus (Mai, 2016; Richterich, 2018). L'abondance de données collectées de toute part à des fins qui sont parfois difficiles à saisir, ainsi que le partage des données, parfois même à l'insu des individus, entre les diverses organisations de l'écosystème dont les intérêts ne sont pas divulgués, peuvent compromettre l'exercice de ce contrôle et de ces choix (Amann et al., 2020).

Afin de protéger la confidentialité des données personnelles, qui contiennent souvent des informations sensibles et très privées sur les individus, les organisations utilisent des méthodes qui visent à empêcher l'identification des sujets, comme l'anonymisation, la pseudonymisation et la désidentification des données¹. Vues comme des mécanismes de préservation de la confidentialité et de la vie privée, ces techniques ne sont pas à toute épreuve, et n'offrent pas toujours le haut niveau de confidentialité promis. Effectivement, elles ont démontré des failles considérables, méconnues des utilisateurs, qui peuvent mettre à risque la confidentialité de leurs informations et leur vie privée (Basso et al., 2016). À titre d'exemple concernant l'anonymisation, technique qui implique le retrait des données identificatoires des bases de données de façon permanente, il a été démontré qu'il ne suffit que de quatre données anonymisées, provenant de différentes banques de données, pour permettre à un acteur spécifique de ré-identifier l'individu que ces données décrivent avec une exactitude quasi-parfaite (OMS, 2021).

À ces risques s'ajoutent ceux de la protection de la sécurité des données, et des cyberattaques potentielles inhérentes à tout système informatique. L'augmentation de l'utilisation de l'intelligence artificielle en santé et la collecte de données massives rend le milieu plus susceptible d'être visé par des fuites de données et cyberattaques qui peuvent compromettre le fonctionnement des SIA, la performance des algorithmes ainsi que la confidentialité des données sur lesquelles ils sont entraînés, parfois même à l'insu des développeurs (OMS, 2021). En plus d'entraver la sécurité, la vie privée et la dignité des individus touchés, ces incidents potentiels peuvent avoir des répercussions graves sur les personnes dont les données sont concernées, comme notamment le ciblage indésirable de leur profil, des tentatives d'hameçonnage crédibles sur la base de leurs données médicales, le vol de leur identité, ou encore le traitement discriminatoire de leurs demandes auprès de fournisseurs à l'extérieur du milieu médical.

¹ La désidentification est une technique qui empêche les identifiants personnels d'être reliés aux informations concernant les personnes. L'anonymisation est une sous-catégorie de la désidentification qui désigne le processus selon lequel les informations identificatoires, directes et indirectes, sont retirées de façon permanente de sorte à empêcher la ré-identification des individus (Leighton Paisner, 2017). Finalement, selon l'Office québécois de la langue française (2002), pseudonymiser signifie « remplacer des données associées à un utilisateur par un pseudonyme, dans un but de confidentialité » dans une banque de données.

1.3 Enjeux de fiabilité, de qualité et de représentativité des données

Les algorithmes utilisés dans les SIA mobilisant l'apprentissage automatique sont entraînés sur des ensembles de données qui se doivent de rencontrer certains standards en termes de qualité, de fiabilité et de représentativité, sans quoi les algorithmes développés sur ces données peuvent être biaisés et causer d'importants torts aux individus, groupes et communautés concernés. La qualité et la fiabilité des données varient en fonction de la diversité de sources desquelles proviennent les données, le soin porté à la gestion du cycle de vie des données, la quantité de données contenue dans la banque d'entraînement ou encore la représentativité des données de cette même banque (OMS, 2021). D'autant plus important, les données d'entraînement doivent être appropriées pour les fins auxquelles les algorithmes sont créés. Un manque de constance dans l'application des différents critères qui nous permettent d'assurer la qualité, la fiabilité et la représentativité des données utilisées peut grandement affecter la performance des algorithmes en question et influencer la qualité de leurs prédictions et recommandations.

Suite à l'entraînement de l'algorithme vient l'étape de validation, qui permet de vérifier la qualité de la performance de cet algorithme ainsi que le caractère généralisable de cette performance. Cette validation doit être faite à l'aide d'un ensemble de données distinct de celui utilisé pour l'entraînement de l'algorithme. Effectivement, le but de cette étape est d'évaluer la performance de l'algorithme sur une variété de données pour assurer sa performance sur le terrain (Hand, 2018), ce que valider l'algorithme sur le même ensemble de données ne permettrait pas de faire. Prenons l'exemple fictif d'un algorithme entraîné pour détecter une condition médicale à partir de marqueurs biologiques. Si les données utilisées pour entraîner l'algorithme ont été majoritairement collectées auprès de personnes de sexe masculin, il est possible que lorsque déployé en contexte hospitalier, l'algorithme tirera des conclusions inexactes lorsqu'il rencontrera des données concernant des personnes de sexe féminin. Lorsque respectés, ces différents critères de validation nous permettent de limiter les chances qu'une fois commercialisés et déployés, les SIA performant de façon sous-optimale auprès de certaines populations, en arrivent à des conclusions qui renforcent des préjugées et engendrent un traitement discriminatoire à leur égard (voir la Fiche 3 « IA en santé : aspects juridiques » pour un complément d'information sur les aspects juridiques liés à la sécurité des données).



1.4 Enjeux liés à l'archivage et la conservation des données

Les données personnelles sont parfois conservées de façon anonymisée, ce qui ne les qualifie plus comme telles. Une fois converties dans un format qui n'identifie plus les personnes, ces données peuvent être utilisées à des fins autres que celles initialement consenties et ont parfois pour but de servir les objectifs de tierces parties (Reddy et al., 2020). Plusieurs enjeux éthiques sont soulevés par un tel scénario. D'abord, rappelons que les données anonymisées peuvent potentiellement être ré-identifiées (Basso et al., 2016), ce qui devient problématique lorsque les données médicales de patients, informations très sensibles, se retrouvent dans des mains autres que celles de professionnels de la santé.

Aussi, il est possible que des personnes ayant initialement partagé leurs données de façon consentie, comme des patients ou des utilisateurs d'objets connectés et applications de santé, désirent éventuellement retrouver un certain contrôle sur celles-ci, et les voir supprimées. Plusieurs raisons peuvent être évoquées pour justifier le retrait éventuel du consentement à la conservation des données. D'abord, les données qui nous concernent représentent ce qui est parfois appelé notre profil digital (Breidbach et al., 2019). Nos données personnelles représentent donc ce que les différents acteurs du milieu de la santé connaissent à propos de nous. Dans le cas d'individus qui vivent avec des conditions de santé spécifiques, parfois même stigmatisantes, les données personnelles peuvent devenir particulièrement contraignantes (Someh et al., 2019). En effet, imaginons le cas d'une personne aux prises avec des problèmes de dépendance, qui se voit bombardée de publicités l'incitant à consulter les sites internet de centre de désintoxication, ou celui d'une femme ayant avorté qui se voit proposer des items pour femmes enceintes lors de ses achats en ligne, ou encore celui d'une personne en rémission d'un cancer qui demeure non-assurable aux yeux des compagnies d'assurance. Au-delà même du fait que ces réalités difficiles leur soient constamment rappelées et les souffrances qui y sont rattachées exacerbées, il devient difficile pour ces personnes de s'affranchir de ces conditions présentes et passées, et de vivre une vie qui n'est pas contrainte par celles-ci et par les données y étant rattachées. Dans plusieurs cas, les personnes peuvent être définies par une condition médicale qui, alors que parfois temporaire, laisse des traces indélébiles dans les banques de données (Martin, 2015). Ces individus peuvent souhaiter retirer leur consentement à l'archivage de leurs données personnelles pour éviter de vivre avec les entraves à leur liberté et leur autonomie que la conservation et le traitement ultérieur de leurs données sensibles peuvent causer. La perte de contrôle des individus sur leurs données personnelles, ainsi que leur désir de retrouver ce contrôle et de pouvoir demander la suppression des données les concernant est, en partie, ce qui motive les revendications envers le droit à l'effacement des données, ou « droit à l'oubli » (voir la Fiche 3 « IA en santé : enjeux juridiques » pour complément d'information).

2 PRINCIPAUX ENJEUX ÉTHIQUES LIÉS AU DÉVELOPPEMENT DES SYSTÈMES D'IA EN SANTÉ

Après avoir discuté des enjeux éthiques spécifiques à l'exploitation des données massives, nous vous présentons ci-après les principaux enjeux éthiques liés au développement et à l'implantation des systèmes d'IA dans le milieu de la santé.

2.1 Utilisation non-éthique des SIA

L'intelligence artificielle peut être utilisée afin d'améliorer plusieurs aspects du système de santé. Face aux nombreuses possibilités offertes par ces avancées technologiques, il est de la responsabilité des décideurs d'évaluer le caractère éthique des différentes applications, puisque toutes ne sont pas nécessairement souhaitables. De façon générale, celles qui contreviennent aux principes éthiques établis par différentes initiatives nationales et internationales et qui causent préjudice de façon claire, voire intentionnelle, ne sont pas souhaitables et devraient soit subir des modifications profondes permettant d'éviter les effets néfastes anticipés, ou encore être retirées du marché (OMS, 2021, voir aussi la section sur les Principales lignes directrices en éthiques de l'IA).

Malgré tout, le caractère non-éthique de certains de ces systèmes n'est pas toujours évident à déceler. Effectivement, enfouis dans les conditions d'utilisation et politiques de vie privée sont parfois dissimulées les intentions des développeurs ou propriétaires de ces technologies qui peuvent porter préjudices aux utilisateurs de façons insoupçonnées (Andrejevic, 2014; Hand, 2018). À l'insu de l'utilisateur, une montre intelligente qui enregistre la localisation de ses entraînements physiques peut devenir un outil de surveillance de ses déplacements qui permet d'orienter les publicités qu'il reçoit, et une application lui permettant de tenir un registre quotidien de ses activités et de ses émotions peut devenir un outil normatif qui décourage certaines habitudes de vie et en encourage d'autres, pouvant parfois servir les agendas cachés de tierces parties. Cette subtile incitation comportementale (en anglais nommée « *nudging* ») oriente les choix individuels en fonction de valeurs et d'intérêts spécifiques, parfois perçus comme supérieurs et positifs (ex. arrêt du tabagisme, perte de poids, exercice physique), mais aussi parfois motivés par des intérêts commerciaux (ex. achat de produits aidant l'arrêt du tabagisme, de suppléments alimentaires assistant la perte de poids, ou d'un abonnement payant à une application de conditionnement physique). Cette pratique entrave du fait l'autonomie et l'agentivité morale² des personnes qui ont le droit et le choix d'adhérer à un mode de vie qui leur convient, qu'il s'aligne ou non avec les recommandations des professionnels de la santé, de la santé publique ou des organisations privées, recommandations qui, d'ailleurs, peuvent changer suivant des avancées scientifiques et technologiques (Birhane, 2020).

² L'agentivité morale est une capacité des humains en matière d'éthique, qui correspond à la capacité d'action, de raisonnement, de jugement et de prise de décision éthiques, par opposition à seulement subir des conséquences morales des situations vécues (Coeckelberg, 2020).

Aussi, la validité scientifique et l'efficacité d'un SIA ne sont pas suffisantes pour justifier le déploiement d'une technologie et la qualifier de « souhaitable ». C'est pourquoi l'évaluation des aspects éthiques de ces technologies est si importante. Entre autres, un des facteurs importants à considérer est la balance des coûts et des bénéfices reliés au déploiement d'une solution d'IA en santé. Malgré les bienfaits potentiels pouvant résulter de l'utilisation d'un SIA, la probabilité et l'ampleur des effets négatifs y étant associés doivent être évaluées et peuvent, selon l'analyse des aspects éthiques effectuée, venir décourager ou même exclure l'utilisation de cette technologie (OSM, 2021). En effet, il serait difficile de justifier l'utilisation d'un SIA promettant des économies de ressources considérables pour le milieu de la santé s'il compromettait la qualité des soins octroyés aux patients, ou s'il entraînait de grands risques de discrimination ou de bris de confidentialité, par exemple. L'ensemble des bénéfices et des coûts potentiels d'une technologie doivent être explorés et attentivement évalués avant de pouvoir justifier le déploiement d'une technologie qui pourrait, malgré ses bénéfices, engendrer des torts non-prévus. Bien qu'il soit difficile de prévoir et contrôler, il n'empêche que les organisations derrière ces technologies doivent, de façon proactive, s'assurer de mettre en place des processus transparents qui permettent d'explorer les enjeux éthiques liés à leur technologie, de mitiger les risques, de monitorer la performance de l'application en contexte réel d'utilisation, de recueillir la rétroaction des différentes parties prenantes suite à son utilisation et de réagir rapidement en cas qu'un tort imprévu survienne et que des individus soient lésés.

2.2 L'explicabilité des SIA

À l'heure actuelle, les systèmes de décision algorithmique en santé servent majoritairement d'outil de support à la décision humaine et n'ont pas une pleine autonomie décisionnelle (Amann et al., 2020). Ces systèmes ne prennent donc pas de décision par eux-mêmes, mais formulent des prédictions et des recommandations. Cependant, même si le jugement final revient au professionnel de la santé (par exemple le diagnostic d'un patient), la grande complexité du fonctionnement de certains SIA fait qu'il est parfois difficile, ou même impossible, pour ce dernier de comprendre le raisonnement qui permet à l'algorithme d'arriver à une recommandation précise. C'est ce qu'on appelle le phénomène de la « boîte noire », symbole de l'opacité du fonctionnement de certains SIA, notamment ceux mobilisant l'apprentissage automatique (Amann et al., 2020). L'absence ou la faible explicabilité des SIA peut contribuer à diminuer l'interprétabilité, l'acceptabilité et l'adoption de la technologie par les professionnels et patients, et occasionne plusieurs enjeux. Imaginons le cas où un clinicien serait en désaccord avec la recommandation d'un SIA. Dans un tel cas, l'impossibilité d'explicitement le raisonnement du SIA afin de le rendre compréhensible pour le clinicien l'empêche de déterminer si la recommandation du système est valable et fiable dans un contexte précis, et donc s'il peut avoir confiance en la recommandation émise par la technologie. Le patient pourrait lui aussi demander à comprendre les tenants d'une recommandation faite par un SIA. Qui plus est, des recommandations potentiellement erronées engendrées par les algorithmes qui servent de base pour la prise de décision clinique pourraient possiblement causer de graves torts aux patients qui, souvent, ne sont pas mis au courant que des décisions concernant leur santé ont été prises à l'aide de SIA (OSM, 2021).

En effet, bien que ces systèmes soient développés avec l'intention d'améliorer la santé des individus et l'organisation des services de santé, ils sont sujets aux erreurs de ceux qui les développent et les approuvent, ainsi qu'à des problèmes techniques, et peuvent parfois offrir de mauvaises recommandations dans l'allocation des ressources, la planification ou encore les soins et les traitements fournis aux patients. Ces recommandations ne sont que rarement contestées et contestables, vu le manque de transparence des systèmes d'IA et peuvent avoir des répercussions importantes sur la qualité des soins octroyés, ainsi que la sécurité et la santé des individus touchés. Il est souhaitable que les entreprises qui développent des SIA augmentent l'explicabilité et l'interprétabilité de ces systèmes, du moins qualitativement.

2.3 La responsabilité décisionnelle

La prise de décisions cliniques assistées par un SIA issu d'algorithmes d'apprentissage automatique, de par leur opacité, complexifie alors la notion de responsabilité, que nous explorons ci-après.

Poursuivons avec cet exemple dans lequel une erreur commise par un clinicien appuyé par un SIA cause des torts sérieux à un patient. Dans un tel cas, qui du clinicien ou du SIA est ultimement responsable des conséquences de la décision prise ? Si le clinicien n'a ni une bonne compréhension ni un réel contrôle sur ce système expert et sur ses recommandations, peut-il être tenu responsable, moralement et/ou légalement, d'une erreur commise sur la base de celui-ci ? Nous pourrions lui reprocher de s'être fié aveuglément à un mauvais système. Cependant, de nombreux SIA possèdent une capacité d'analyse et une performance qui sont parfois supérieures à celle du clinicien et qui justifient leur utilisation en milieu clinique. Cette performance supérieure des systèmes d'IA est possible grâce à l'analyse de millions de données issues de jugements cliniques et équivaut donc, selon certains, à l'expertise de plusieurs milliers de cliniciens. Dans le cas qui nous occupe, l'erreur est-elle donc la faute de l'organisme et des personnes qui ont approuvé l'utilisation de ce système au niveau clinique (par exemple Santé Canada), de l'organisation de santé qui a implanté cette technologie, ou encore de l'entreprise qui a développé le SIA en question ? Selon les recommandations de l'OMS (2021), bien que les cliniciens devraient raisonnablement pouvoir se fier aux recommandations des SIA, celles-ci ne devraient jamais supplanter le jugement des cliniciens. Le partage de la responsabilité entre les différents acteurs impliqués dans le développement et la commercialisation de systèmes d'IA, de même que ceux impliqués dans leur déploiement dans un hôpital, est complexe. Effectivement, la multitude d'acteurs impliqués diffuse la notion de responsabilité qui finit parfois par n'être portée par personne, ce qui peut causer des torts aux individus touchés par des erreurs médicales impliquant des SIA. Les juristes s'intéressent particulièrement à la notion de la responsabilité légale liées aux SIA (voir fiche 3 « IA en santé : aspects juridiques » pour plus d'informations).

2.4 Les biais et la discrimination algorithmique

Comme mentionné plus haut, les algorithmes d'apprentissage automatique apprennent en fonction des données à partir desquelles ils sont entraînés, mais ils apprennent aussi en fonction des décisions humaines prises lors de leur développement. Effectivement, ces décisions orientent le développement d'un SIA et déterminent la raison d'être d'un projet, les professionnels qui le réaliseront, les paramètres des algorithmes, les données utilisées pour l'entraîner et valider leur performance, mais aussi, auprès de quelles populations et au sein de quels milieux le SIA sera déployé (Cuttillo et al., 2020). De façon involontaire, les individus qui prennent ces décisions teintent souvent inconsciemment les étapes de développement du SIA de leurs perspectives, de leurs croyances et de leurs biais individuels. Les algorithmes sont donc formés en fonction de ces paramètres sociaux intangibles transmis par leurs concepteurs, entre autres. Ils peuvent donc reproduire et amplifier leurs biais et performer au désavantage de certains groupes non-considérés, par exemple en discriminant des individus issus de communautés marginalisées (Birhane, 2021) (voir fiche 3, « IA en santé : aspects juridiques » pour un complément d'informations).

2.5 Les enjeux liés au développement d'agents moraux artificiels

Qu'est-ce qu'un agent moral artificiel ? Pour reprendre une notion abordée plus haut, un agent moral artificiel est un système d'intelligence artificielle capable « d'action morale, de raisonnement, de jugement et de prise de décision » (Coeckelberg, 2020, traduction libre : 204). De tels systèmes autonomes avancés ne sont pas encore réellement implantés dans le système de santé, mais commencent à faire leur apparition dans certaines industries (pensons par exemple aux agents conversationnels « *chatbots* » basés sur l'apprentissage automatique, ou encore aux voitures autonomes, qui sont de bons exemples d'agents moraux artificiels en développement). Néanmoins, le rythme de développement des technologies et le grand intérêt général pour l'innovation en intelligence artificielle nous poussent à croire qu'un jour ces systèmes avancés puissent être développés et utilisés dans le milieu de la santé. En effet, les agents moraux artificiels pourraient trouver plusieurs utilités dans le domaine de la santé, notamment celle de diagnostiquer les patients de façon autonome (OMS, 2021). Cela dit, la délégation complète de l'expertise médicale du professionnel de la santé dans les « mains » d'un tel agent aurait de nombreuses implications. Par exemple, la participation du patient dans la prise de décision concernant sa propre santé pourrait en être affectée. Cette participation (on y réfère souvent par la notion de partenariat de soin) permet au patient d'exercer son autonomie, de défendre ses intérêts et de consentir à des soins qui respectent son projet de vie et ses valeurs (Amann et al., 2020). La complexité de fonctionnement de ces systèmes hautement spécialisés, qui notons-le à nouveau, échappe parfois aux développeurs eux-mêmes, pourrait potentiellement entraver la qualité du consentement fourni par le patient. De son côté, le professionnel de la santé pourrait perdre le contrôle d'une partie importante de ce qui construit son identité professionnelle, mais aussi le contact avec une dimension essentielle du soin : la relation avec son patient (OMS, 2021).

Une autre implication liée à la délégation d'une agentivité morale aux SIA est à soulever. Effectivement, de quelle « morale » parlons-nous ici ? Cette question relève des valeurs morales et visions du monde qui seront véhiculées de façon implicite par ces systèmes, et des intérêts et intentions des personnes derrière le développement de ces systèmes. Comme mentionné plus haut, les personnes qui développent ces SIA sont porteurs de biais, de perspectives, de valeurs et d'intérêts individuels, et ils représentent aussi bien souvent des intérêts privés et commerciaux. Ce faisant, ils contribuent à concevoir des systèmes autour d'objectifs et d'intentions qui, de façon consciente ou non, ne sont pas neutres. En laissant le développement de ces systèmes opaques non-réglémenté et non-encadré au niveau éthique, rien ne nous assure que les intérêts et les valeurs promus par ceux-ci soient alignés avec les valeurs du patient, qui sont par ailleurs en soi variables, ou mêmes celles de la société (Amann et al., 2020; OMS, 2021).

2.6 Les enjeux existentiels de l'IA

La dimension existentielle de l'intelligence artificielle aborde les choix fondamentaux que nous faisons en tant que société et la place que nous désirons donner aux technologies d'intelligence artificielle dans nos vies et dans notre société. Cette dimension nous ramène à la question de ce que signifie vivre une vie satisfaisante et digne, et à prendre des décisions sociétales qui sont arrimées avec cette définition et des valeurs morales susceptibles de varier à travers les cultures et à travers le temps (Birhane, 2021). Prenons l'exemple d'un robot compagnon muni d'un agent conversationnel déployé en hébergement pour personnes âgées dans le but de briser leur isolement (voir la Fiche 1 « IA en santé : Définitions et usages »). Les intentions derrière un tel projet sont louables et les retombées potentiellement positives. N'empêche, ce scénario pose plusieurs questions éthiques. Qu'est-ce que cette décision révèle-t-elle à propos de nous, en tant que société, de notre relation avec les personnes âgées, et avec la notion du vieillissement en général ? Est-il désirable de remplacer le contact humain par celui d'un robot, et est-ce réellement ce que nous souhaitons pour nos aînés ? Les réponses à ces questions peuvent d'ailleurs varier, selon les cultures, et évoluer dans le temps.

Certains éthiciens explorent aussi d'autres questionnements existentiels liés au développement de l'IA et se questionnent notamment sur les enjeux éthiques liés au transhumanisme, c'est-à-dire à l'augmentation des capacités humaines par la science et les technologies, dont notamment l'IA et la robotique.

2.7 Les enjeux liés au futur du travail et à la transformation des emplois en santé

De nombreux questionnements éthiques sont liés aux bouleversements du marché du travail générés par l'intégration de l'intelligence artificielle dans le milieu de la santé. Beaucoup d'incertitude plane sur la question, tant sur la nature des changements potentiels aux tâches et aux emplois à travers les différents secteurs de la santé, que sur les impacts de ces changements sur les travailleurs de la santé et sur leur identité professionnelle. Ainsi, plusieurs acteurs se questionnent sur la désirabilité et sur l'acceptabilité sociale et organisationnelle de ces changements.

Si les individus ne sont pas remplacés dans leur emploi, plusieurs seront certainement appelés à travailler en proximité avec des SIA. L'intégration de ces systèmes dans le milieu de la santé nous pousse donc à adresser la question de la littératie numérique du personnel médical, administratif et de soutien en matière d'IA (OMS, 2021), sans oublier celle des patients et de leurs proches. La technologie derrière ces systèmes est complexe et le fossé qui sépare le champ d'expertise des différents travailleurs de la santé et les connaissances nécessaires pour pouvoir bien intégrer les SIA dans leur quotidien est parfois très large (Ahmad et al., 2018). Effectivement, à l'heure actuelle, les professionnels de la santé, le personnel administratif, tout comme les patients en connaissent très peu sur l'intelligence artificielle, ses utilisations et son fonctionnement, ce qui pose de nombreux enjeux. Le manque de connaissances et de compréhension pourrait notamment créer une dépendance humaine envers la technologie et affecter la capacité de jugement professionnel. Finalement, l'introduction des SIA dans le quotidien des travailleurs de la santé pourrait porter atteinte à l'autonomie et la dignité de ces derniers qui voient leur expertise être dévalorisée par l'introduction de SIA qui viennent relever ces individus d'une partie ou même éventuellement de l'entièreté de leurs fonctions (OMS, 2021).

2.8 L'accès variable aux technologies de l'IA et l'accroissement des injustices sociales par l'IA

Si l'intelligence artificielle et ses différentes applications ne semblent connaître aucune frontière, l'accès aux technologies d'IA est loin d'être universel. Effectivement, l'accès aux technologies d'intelligence artificielle est souvent réservé aux individus et organisations issus de pays riches ou de milieux sociaux économiques élevés (en anglais, on nomme cet enjeu le *digital divide*). Ces iniquités d'accès et d'utilisation à travers le monde renforcent des disparités sociales existantes et affectent considérablement la qualité et la représentativité des données à notre disposition pour créer, développer et déployer des SIA qui respectent les principes d'équité, de diversité et d'inclusion qui sont au cœur de l'IA responsable (OSM, 2021). Ces écarts sont aussi présents au niveau local, notamment entre les différents groupes et communautés d'un pays. Historiquement, les groupes marginalisés vivent une précarité qui rend difficile l'accès à certaines commodités essentielles ; encore plus aux technologies d'intelligence artificielle. L'écart d'accès et d'utilisation aux échelles locale et mondiale affectent à leur tour la qualité des technologies d'intelligence artificielle qui sont développées (OSM, 2021).

Effectivement, la difficulté d'accès à ces technologies par certains groupes signifie qu'au mieux très peu de données les concernant sont collectées ; ils sont donc sous, voire non-représentés, par ces technologies qui, par manque d'accès à leurs données, ne peuvent être adaptées à leur réalité. Ce phénomène renforce les disparités sociales existantes et participe à les marginaliser davantage (Birhane, 2021), tel que discuté précédemment dans la section 2.4 sur la discrimination algorithmique. Les SIA peuvent donc parfois être créés par et pour certains groupes privilégiés, excluant une majorité d'individus marginalisés par leur genre, leur sexe, leur âge, leur origine ethnique ou leur condition médicale, par exemple. Parfois déployés à tort dans des milieux issus de cultures et de réalités différentes de ceux ayant été pris en compte dans leur développement, les SIA peuvent avoir des effets néfastes considérables et palpables sur la vie de ces individus (Birhane, 2021).

2.9 Les enjeux liés à commercialisation des technologies de l'IA en santé

Les SIA déployés dans le système de la santé émanent principalement de la recherche universitaire et de la recherche menée dans les entreprises privées. Comme mentionné plus haut, certaines entreprises privées opèrent avec un manque de transparence concernant leurs pratiques de gestion et protection des données personnelles des utilisateurs de leur technologie. L'absence de mesures incitatives à la transparence permet à ces entreprises d'opérer sous un voile cachant des informations à propos de leur technologie et des pratiques qui « franchissent parfois des limites éthiques » (OSM, 2021).

Aussi, alors que certaines entreprises divulguent le moins d'information possible à propos de leurs activités, elles en collectent toujours plus à propos de celles de leurs utilisateurs (Richter, 2018). Cette asymétrie est problématique. Effectivement, si l'accumulation de données utilisateurs augmente le pouvoir des entreprises, celui des utilisateurs est considérablement réduit par le manque d'information disponible à l'égard des entreprises et réduit les possibilités d'action contre elles, autant de la part des utilisateurs, que des autorités pertinentes (Someh et al., 2019 ; Mittelstadt et Floridi, 2016).

De plus, le pouvoir des grandes entreprises privées de technologies sert parfois de moyen de pression efficace auprès des autorités et des gouvernements. Cette pression est notamment exercée dans le but de modifier le cours de travaux, le contenu de politiques ou l'application de lois de sorte que ces derniers s'alignent avec les intérêts commerciaux des entreprises privées qui, eux, ne sont pas toujours alignés avec ceux de la santé publique et de la société (OMS, 2021).

2.10 Les impacts environnementaux de l'IA

Quoi qu'encre peu étudiés, les impacts environnementaux de l'IA sont réels. Malgré les nombreux bénéfices qu'ils peuvent avoir dans le milieu de la santé, les SIA s'accompagnent d'une dépense énergétique qui peut, paradoxalement, contribuer à la dégradation de la santé de notre planète ; dégradation qui a des coûts visibles et connus sur la santé des populations. Effectivement, la quantité d'énergie utilisées pour entraîner les algorithmes d'apprentissage automatique, au cœur de nombreuses applications de l'intelligence artificielle aujourd'hui, ainsi que les infrastructures nécessaires pour supporter le développement des algorithmes et l'entreposage des données, par exemple, est non-négligeable et alourdit considérablement la dette énergétique des technologies d'intelligence artificielle (OSM, 2021).

3 PRINCIPALES LIGNES DIRECTRICES EN ÉTHIQUE DE L'IA

En réaction aux nombreux enjeux éthiques soulevés par le développement et le déploiement de l'IA dans les différentes sphères de l'économie et de la société, différents organismes nationaux et internationaux ont proposé dans les dernières années des lignes directrices afin de pouvoir évaluer la dimension éthique de ces technologies et favoriser le développement d'une IA responsable. Nous vous présentons ci-après deux de ces initiatives : la *Déclaration de Montréal pour le développement responsable de l'intelligence artificielle* et les *Principes directeurs de l'OMS sur l'éthique et la gouvernance de l'IA en santé*.

3.1 La Déclaration de Montréal pour le développement responsable de l'intelligence artificielle

Portée par l'Université de Montréal depuis sa création en 2018, la *Déclaration de Montréal pour le développement responsable de l'intelligence artificielle* est une charte de principes et de valeurs éthiques qui permet d'encadrer le développement des SIA et d'orienter la direction de la transition numérique vers le bien-être de tous. Fruit d'un processus délibératif inclusif, la Déclaration de Montréal est le résultat de réflexions issues de toutes les sphères de l'écosystème qui mettent en dialogue les parties prenantes touchées par le développement de l'intelligence artificielle dans notre société (Université de Montréal, 2018). Le Tableau 1 ci-dessous présente les différents principes et sous-principes mis de l'avant par la Déclaration de Montréal.

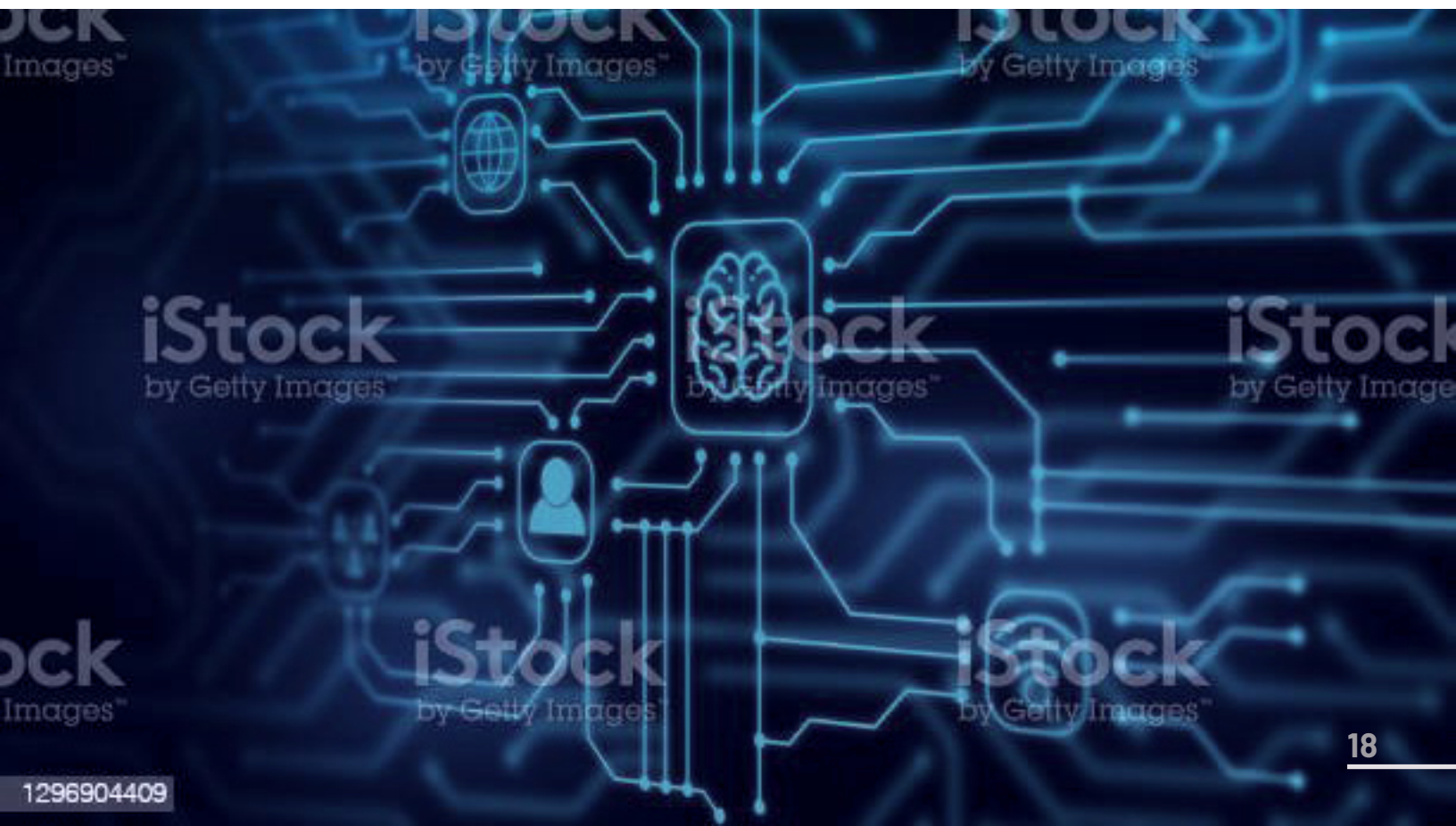


Tableau 1. La déclaration de Montréal pour un développement responsable de l'intelligence artificielle³

1. Principe de bien-être

Le développement et l'utilisation des SIA doivent permettre d'accroître le bien-être de tous les êtres sensibles.

2. Principe de respect de l'autonomie

Les SIA doivent être développés et utilisés dans le respect de l'autonomie des personnes et dans le but d'accroître le contrôle des individus sur leur vie et leur environnement.

3. Principe de protection de l'intimité et de la vie privée

La vie privée et l'intimité doivent être protégées de l'intrusion de SIA et de systèmes d'acquisition et d'archivage des données personnelles (SAAD)

4. Principe de la solidarité

Le développement de SIA doit être compatible avec le maintien de lien de solidarité entre les personnes et les générations.

5. Principe de participation démocratique

Les SIA doivent satisfaire les critères d'intelligibilité, de justifiabilité, d'accessibilité, et doivent pouvoir être soumis à un examen, un débat et un contrôle démocratique.

6. Principe d'équité

Le développement et l'utilisation des SIA doivent contribuer à la réalisation d'une société juste et équitable.

7. Principe d'inclusion de la diversité

Le développement et l'utilisation de SIA doivent être compatibles avec le maintien de la diversité sociale et culturelle et ne doivent pas restreindre l'éventail des choix de vie et des expériences personnelles.

8. Principe de prudence

Toutes les personnes impliquées dans le développement des SIA doivent faire preuve de prudence en anticipant autant que possible les conséquences néfastes de l'utilisation des SIA et en prenant des mesures appropriées pour les éviter.

9. Principe de responsabilité

Le développement et l'utilisation des SIA ne doivent pas contribuer à une déresponsabilisation des êtres humains quand une décision doit être prise.

10. Principe de développement soutenable

Le développement et l'utilisation des SIA doivent se réaliser de manière à assurer une soutenabilité écologique forte de la planète.

Depuis la parution du document en 2018, la discussion autour des principes normatifs qu'il énonce a évolué. Effectivement, l'engouement généré par cette première étape essentielle laisse aujourd'hui place à l'impatience de voir les principes traduits en pratiques concrètes pour le monde organisationnel (Mittelstadt, 2019). Que veulent dire ces principes pour les organisations qui mettent à profit l'intelligence artificielle à des fins commerciales ? Le défi est de taille. En effet, en plus du fossé qui existe souvent entre les principes et la réalité organisationnelle, il existe des scénarios pratiques qui font émerger des tensions entre les différents principes des chartes comme la Déclaration de Montréal (par exemple entre les principes d'équité et de développement soutenable), ce qui complique leur opérationnalisation (Krijger, 2021). Chose certaine : amener la Déclaration de Montréal à la prochaine étape, celle de l'implantation en pratiques organisationnelles concrètes qui auront des retombées positives sur la façon dont les technologies d'IA sont créées, développées et déployées nécessitera une collaboration étroite entre tous les acteurs impliqués, dans une perspective interdisciplinaire.

³ Référence : <https://www.declarationmontreal-iaresponsable.com>

3.2 Les Principes directeurs de l'OMS sur l'éthique et la gouvernance de l'IA en santé

L'Organisation Mondiale de la Santé a récemment publié un guide normatif sur l'éthique et la gouvernance de l'intelligence artificielle en santé. Ce document est le fruit de deux ans de travail réalisé par un groupe d'experts constitué de vingt individus issus de disciplines pertinentes. Il contient l'analyse du groupe d'experts quant aux défis et opportunités de l'intelligence artificielle, ainsi que des recommandations de principes, politiques et pratiques jugés nécessaires pour favoriser l'utilisation responsable de l'intelligence artificielle en santé. Le Tableau 2 présente les six principes directeurs de l'OMS sur l'IA en santé, ainsi que leurs explications.

Tableau 2. Présentation des principes directeurs de l'OMS sur l'IA en santé⁴

1. Protéger l'autonomie de l'être humain

Dans le contexte des soins de santé, cela signifie que les humains doivent rester maîtres des systèmes de soins de santé et des décisions médicales ; la vie privée et la confidentialité doivent être protégées, et les patients doivent donner un consentement éclairé valable grâce à des cadres juridiques appropriés pour la protection des données.

2. Promouvoir le bien-être et la sécurité des personnes ainsi que l'intérêt public

Les concepteurs de technologies d'IA doivent satisfaire aux exigences réglementaires en matière de sécurité, de précision et d'efficacité pour des cas d'utilisation ou des indications bien définis. Des mesures de contrôle de la qualité dans la pratique et d'amélioration de la qualité dans l'utilisation de l'IA doivent être disponibles.

3. Garantir la transparence, la clarté et l'intelligibilité

La transparence exige que des informations suffisantes soient publiées ou documentées avant la conception ou le déploiement d'une technologie d'IA. Ces informations doivent être facilement accessibles et faciliter une consultation et un débat publics significatifs sur la façon dont la technologie est conçue et sur la façon dont elle devrait ou ne devrait pas être utilisée.

4. Encourager la responsabilité et l'obligation de rendre des comptes

Bien que les technologies de l'IA accomplissent des tâches spécifiques, il incombe aux parties prenantes de veiller à ce qu'elles soient utilisées dans des conditions appropriées et par des personnes correctement formées. Il convient de mettre en place des mécanismes efficaces de remise en question et de recours pour les individus et les groupes qui sont lésés par des décisions fondées sur des algorithmes.

5. Assurer l'inclusion et l'équité

L'inclusion exige que l'IA pour la santé soit conçue pour encourager l'utilisation et l'accès équitable le plus large possible, indépendamment de l'âge, du sexe, du genre, du revenu, de la race, de l'origine ethnique, de l'orientation sexuelle, des capacités ou d'autres caractéristiques protégées par les codes des droits de l'homme.

6. Promouvoir une IA réactive et durable

Les concepteurs, les développeurs et les utilisateurs devraient évaluer de manière continue et transparente les applications de l'IA lors de leur utilisation réelle afin de déterminer si l'IA répond de manière adéquate et appropriée aux attentes et aux exigences. Les systèmes d'IA devraient également être conçus de manière à minimiser leurs conséquences environnementales et à accroître leur efficacité énergétique. Les gouvernements et les entreprises devraient prendre en compte les perturbations anticipées sur le lieu de travail, notamment la formation des travailleurs de la santé pour qu'ils s'adaptent à l'utilisation des systèmes d'IA, et les pertes d'emploi potentielles dues à l'utilisation de systèmes automatisés.

⁴ Rapport de l'OMS : [Ethics and Governance of Artificial Intelligence for Health - WHO Guidance](#)

POUR EN SAVOIR PLUS

Vous trouverez ci-bas les principales initiatives et regroupements impliqués dans la réflexion sur l'éthique de l'IA en santé au Québec, au Canada, et à l'international.

Les travaux des axes Éthique, gouvernance et démocratie et *Santé durable* de l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) <https://observatoire-ia.ulaval.ca>

La Déclaration de Montréal pour un développement responsable de l'intelligence artificielle <https://www.declarationmontreal-iaresponsable.com>

« Le guide des principes en innovation et IA responsables en santé et le formulaire d'auto-évaluation », outils du CHUM sur la mise en pratique de la Déclaration de Montréal pour favoriser le développement et l'implantation d'IA en santé responsable : <https://eiaschum.ca/formations/guide-des-principes-d-innovation-et-d-intelligence-artificielle/>

Institut d'éthique de l'IA de Montréal <https://montrealetics.ai/>

Principes généraux pour une IA responsable « Utilisation responsable de l'intelligence artificielle » énoncés par le Gouvernement du Canada <https://www.canada.ca/fr/gouvernement/systeme/gouvernement-numerique/innovations-gouvernementales-numeriques/utilisation-responsable-ai.html>

Guide de l'OMS « *Ethics and Governance of Artificial Intelligence for Health* » <https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200>

Principes de OCDE sur l'intelligence artificielle <https://oecd.ai/fr/ai-principles>

Groupe d'experts de haut niveau pour une IA digne de confiance de la Commission Européenne https://ec.europa.eu/commission/presscorner/detail/fr/IP_19_1893

UK Department of Health and Social Care: A guide to good practice for digital and data-driven health technologies <https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology>

La norme CAN/CIOSC 101:2019 du CIO Strategy Council « Intelligence artificielle : Conception éthique et utilisation de systèmes de décision automatisés » <https://ciostrategycouncil.com/normes/conception-ethique/?lang=fr>

RÉFÉRENCES

Ahmad, M.uhammad A.urangzeb, Carly Eckert, C., Ankur Teredesai, A., & Greg McKelvey, G. Louise A Dennis, Marija Slavkovik, et al. (2018). Interpretable Machine Learning in Healthcare. *IEEE Intelligent Informatics Bulletin*, vol.19, no 1.

Amann, J., Julia, Alessandro Blasimme, A., Effy Vayena, E., Dietmar Frey, D., Vince I & . Madai, V.L. et Q. consortium the Precise (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, vol. 20, no 1, p. 310. <https://doi.org/10.1186/s12911-020-01332-6>

Andrejevic, M. Mark (2014). « Big Data, Big Questions: The Big Data Divide », *International Journal of Communication*, vol. 8, pp.1673–1689. <https://doi.org/10.1080/15358625.2014.900005> ce site web ne fonctionne pas

Basso, Tania., , Roberta Matsunaga, R., Regina Moraes, R., & et Nuno Antunes, N. (2016). « Challenges on Anonymity, Privacy and Big Data », communication présentée dans *2016 Seventh Latin-American Symposium on Dependable Computing (LADC)*, IEEE. <https://doi.org/10.1109/LADC.2016.34>

Beauchamp, T. om L., (2011). « Informed Consent: Its History, Meaning, and Present Challenges », *Cambridge Quarterly of Healthcare Ethics*, vol. 20, pp. 515-523. <https://doi.org/10.1017/S096318011000259>

Birhane, A.beba (2021). Algorithmic injustice: A relational ethics approach,. *Patterns*, vol. 2, no 2, p. 100205. <https://doi.org/10.1016/j.patter.2021.100205>

Breidbach, C.hristophe F., Michael Davern, M., Graeme Shanks, G., & et Ida Asadi-Someh, I. (2019). On the Ethical Implications of Big Data in Service Systems, dans *Handbook of Service Science*, Volume II, Springer, p. 661-674. <https://doi.org/10.1007/978-3-319-98512-1https://doi.org/10.1007/978-3-319-98512-1>

Calvo, R. A., Peters, D., Vold, K., & Ryan, R. M. (2020). Supporting human autonomy in AI systems: A framework for ethical enquiry. In *Ethics of Digital Well-Being* (pp. 31-54). Springer, Cham. https://doi.org/10.1007/978-3-030-50585-1_2

Coeckelbergh, M.ark (2020). *AI Ethics*, The MIT Press. The MIT Press Essential Knowledge Series, 229 p.

Commission Européenne (2016). Règlement général sur la protection des données. RGPD. Chapitre 1, article 4, Récupéré de <https://eurlex.europa.eu/legalcontent/FR/TXT/HTML/?uri=CELEX:32016R0679#d1e1884-1-1>

Cutillo, C.hristine M., Karlie R. Sharma, K. R., Luca Foschini, L., Shinjini Kundu, S., Maxine Mackintosh, M., Kenneth D. Mandl, K.D., and MI in Healthcare Workshop Working Group. et al. (2020). Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency, *Digital Medicine*, vol. 3, no 1, p. 47. <https://doi.org/10.1038/s41746-020-0254-2>

Grundy, Q., Chiu, K., Held, F., Continella, A., Bero, L., & Holz, R. (2019). Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. *BMJ*, 364, I920. doi:10.1136/bmj.I920

Hand, D.avid J. (2018). « Aspects of Data Ethics in a Changing World: Where Are We Now? », *Big Data*, vol. 6, no 3. <https://doi.org/10.1089/big.2018.0083>

Krijger, J.oris (2021). « Enter the metrics : critical theory and organizational operationalization of AI ethics », *AI & Society*, pp. 1-11.

Leighton, P.aisner, Bryan Cave, B. (2017). At a glance: De-identification, anonymisation, and pseudo-anonymisation under the GDP, Boulder (CO). Récupéré le 9 janvier 2022 de <https://www.bclplaw.com/en-US/insights/at-a-glance-de-identification-anonymization-and-pseudonymization-1.html>

Mai, Jens-Erik (2016). « Big Data Privacy: The datafication of personal information », *The Information Society*, vol. 32, no 3, p. 192-199. <http://dx.doi.org/10.1080/01972243.2016.1153010>

Martin, Kirsten E. (2015). « Ethical Issues in the Big Data Industry », *MIS Quarterly Executive*, vol. 14, no 2, pp. 67-85.

Mittelstadt, B.rent (2019). Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, vol. 1, no 11, p. 501-507. <https://doi.org/10.1038/s42256-019-0114-4>

Mittelstadt, B.rent D., & , andet Luciano Floridi, F. (2016). « The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts », *Science and Engineering Ethics*, vol. 22, no 2, pp. 303-341. <https://doi.org/10.1007/s11948-015-9652-2>.

Office de la langue française (2002). Fiche terminologique, https://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?ld_Fiche=8363289

Porsdam Mann, S. Mann, Sebastian,, Julian Savulescu, J., &et Barbara J. Sahakian, B.J. (2016). « Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue », *Philosophical Transactions Royal Society, A* 374:20160130. <https://doi.org/10.1098/rsta.2016.0130>

Reddy, S.andeeep, Sonia Allan, S., Simon Coghlan, S., &and Paul Cooper, P. (2020). « A governance model for the application of AI in health care », *Journal of the American Medical Informatics Association*, vol. 27, no 3, pp. 491-497. doi: 10.1093/jamia/ocz192 et al??

Richterich, A.nnika (2018). *The Big Data Agenda: Data Ethics and Critical Data Studies*. London: *University of Westminster Press*. Récupéré de <https://library.open.org/bitstream/handle/20.500.12657/30155/649695.pdf?sequence=1>

Someh, I.da, Michael Davern, M., Christophe F. Breidbach, C.F., e &t Graeme Shanks, G. (2019). « Ethical Issues in Big Data Analytics: A Stakeholder Perspective », *Communications of the Association for Information Systems*, vol 44, p. 718-747. <https://doi.org/10.17705/1CAIS.04434>

Université de Montréal (2018). Déclaration de Montréal pour un développement responsable de l'IA <https://www.declarationmontreal-iaresponsable.com>

World Health Organization (2021). Ethics and governance of artificial for health: Who guidance. Report. <https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200>



**OBSERVATOIRE INTERNATIONAL
SUR LES IMPACTS SOCIÉTAUX
DE L'IA ET DU NUMÉRIQUE**

Pavillon Charles-De Koninck, local 2489
1030, avenue des Sciences-Humaines
Université Laval
Québec (Québec) G1V 0A6

observatoireia@ulaval.ca
observatoire-ia.ulaval.ca