

Sous la direction de
Jean-Pierre Béland
Charles-Étienne Daniel

LA PERSONNE TRANSFORMÉE

Nouveaux enjeux
éthiques et juridiques



Collection dirigée par Jean-Pierre Béland et Johane Patenaude

Cette collection traite de questions et d'enjeux éthiques et les inscrit dans un dialogue ouvert, pluraliste et interdisciplinaire. Les textes sont destinés à un large public dans le but de faire avancer la réflexion sur ces enjeux et ainsi contribuer à la prise de décisions socialement acceptables sur chacun des sujets traités.

LA PERSONNE TRANSFORMÉE
Nouveaux enjeux éthiques et juridiques

LA PERSONNE TRANSFORMÉE

Nouveaux enjeux éthiques et juridiques

Sous la direction de
Jean-Pierre Béland et Charles-Étienne Daniel



**Presses de
l'Université Laval**

Financé par le gouvernement du Canada
Funded by the Government of Canada

| **Canada**

Nous remercions le Conseil des arts du Canada de son soutien.

We acknowledge the support of the Canada Council for the Arts.



Conseil des arts
du Canada

Canada Council
for the Arts

Les Presses de l'Université Laval reçoivent chaque année de la Société de développement des entreprises culturelles du Québec une aide financière pour l'ensemble de leur programme de publication.

SODEC
Québec 

Maquette de couverture: Laurie Patry
Mise en pages: **Santo** *graph*

© Presses de l'Université Laval. Tous droits réservés.
Dépôt légal 4^e trimestre 2019

ISBN 978-2-7637-4635-7
PDF 9782763746364

Les Presses de l'Université Laval
www.pulaval.com

Toute reproduction ou diffusion en tout ou en partie de ce livre par quelque moyen que ce soit est interdite sans l'autorisation écrite des Presses de l'Université Laval.

TABLE DES MATIÈRES

REMERCIEMENTS	XI
INTRODUCTION	1
<i>Jean-Pierre Béland et Charles-Étienne Daniel</i>	
CHAPITRE PREMIER	
Définir la personne transformée: Pourquoi? Comment?	5
<i>Georges A. Legault</i>	
Introduction.....	5
1. L'intervention technique et la remise en question de la définition de la personne	6
2. Pourquoi définir la personne?	12
3. Comment définir la « personne »?	16
3.1 L'approche définitionnelle substantive ou attributive.....	16
3.2 Le choix des critères et leur opérationnalisation	20
Conclusion.....	24
CHAPITRE 2	
La subjectivité à l'épreuve de la technologie	29
<i>Célia Vaz-Cerniglia et Margherita Merucci</i>	
Introduction	29
1. La subjectivité en psychologie	32
2. Illustrations	38
Conclusion.....	41

CHAPITRE 3

Personne transformée, médecine régénérative et humanisme 43

Christian Byk

Introduction : dynamique des contradictions et (nouvelles) finalités sociales 43

1. La transformation biologique de la personne : l'émergence d'une nouvelle réalité aux finalités ambiguës..... 44

A) La légitimité de la finalité médicale et scientifique..... 45

1) De la médecine réparatrice à la médecine régénérative 45

2) De l'acquisition de nouvelles connaissances (génétique, embryogenèse, neurosciences) à la tentation d'améliorer des capacités humaines..... 46

B) Les finalités sociales et l'acceptabilité de la médecine régénérative 47

1) La quête d'identité..... 48

2) La réalisation d'un nouveau projet collectif est-elle encore possible? 49

2. Le corps régénéré et le renouvellement de la perception sociale de la personne humaine 50

A) La personne humaine en mutation..... 50

1) Les craintes que soulève l'idée d'une transformation de l'homme sont-elles justifiées et dans quelle mesure?..... 51

2) L'appropriation sociale des technologies 52

B) « Il faut qu'une porte soit ouverte ou fermée » 54

Conclusion : l'homme médiateur d'un monde en mouvement..... 55

CHAPITRE 4

Le transsexualisme : transformation fondamentale de la personne 57

Pascale Boucaud

1. Une transformation imprévue par la loi 58

2. L'assouplissement du grand principe d'indisponibilité de l'état des personnes..... 60

3. Un troisième sexe? 61

4. Revirement de jurisprudence : le droit au respect de la vie privée l'emporte sur le principe de l'indisponibilité de l'état des personnes 62

5. Exigence primordiale de l'irréversibilité de la transformation physique 64

6. Assouplissement de cette exigence au nom du respect de l'intégrité physique des personnes..... 65

7. Évolution ultime du droit français : abandon de l'exigence de traitements induisant des modifications difficilement réversibles 69

8. Les questions en suspens: incidences du changement de sexe.....	69
Conclusion.....	72

CHAPITRE 5

Personne transformée par l'intelligence artificielle et droit de la responsabilité.....	75
---	----

Marjolaine Monot-Fouletier

1. De l'adaptabilité du droit commun de la responsabilité.....	77
A) L'hypothétique recours à la responsabilité pour faute personnelle	78
1) La faute humaine.....	78
2) La faute du système d'intelligence artificielle « individualisé » : responsabiliser le robot	80
B) Les incertitudes liées au recours à la responsabilité du fait de la garde	81
2. Les perspectives ouvertes par des régimes spéciaux de responsabilité	83
A) La déclinaison des régimes spécifiques de responsabilité existants	83
1) Le cas particulier du « produit défectueux »	83
2) La prise en compte du risque	85
B) Un nécessaire changement de paradigme: de la responsabilité à la solidarité	86

CHAPITRE 6

De la signification éthique des limites.....	91
--	----

Mathieu Guillermin et Mathieu Lefort

Introduction.....	91
1. Vers une transgression radicale des limites?	94
2. Reconnaître les limites des pouvoirs technologiques.....	100
3. Signification éthique des limites.....	106
Conclusion: Une éthique de la limite en quête de performance globale....	111

CHAPITRE 7

Quelle(s) forme(s) de gouvernance juridique pour l'encadrement de l'IA et de la robotique interactive?.....	113
--	-----

Charles-Étienne Daniel

Introduction.....	113
Partie I – L'innovation dans le domaine de l'IA comme moteur économique et les défis de son encadrement	116

Section 1. L'importance de l'innovation dans le développement économique	116
Section 2. Les défis de l'encadrement du développement économique et technologique de l'IA	119
Partie II – Le recours au droit comme mode de gouvernance pour encadrer le développement de l'IA et de la robotique	124
Section 1. Une méthode d'analyse pour examiner la capacité du droit comme mode de gouvernance à faire face au développement de l'IA et de la robotique	125
Section 2. Le droit dur par rapport aux enjeux de santé, de sécurité et de vie privée soulevés par l'IA et la robotique	128
2.1 Le droit réglementaire	128
2.2 La responsabilité civile	132
Section 3. Le recours aux autres formes du droit pour encadrer le développement de l'IA et de la robotique	136
3.1 Le droit souple et la normalisation technique	136
3.2 Le droit mou et les enjeux de nature éthique, légale et sociale (ELS)	139
Conclusion.....	144
 CHAPITRE 8	
La personne transformée : comment peut-on mettre une limite?	147
<i>Jean-Pierre Béliand</i>	
1. Implant de mémoire dans le domaine des BCI en « neurotechnologie »	151
2. Systèmes d'aide à la décision (intelligence algorithmique)	158
3. Les implants RFID	171
4. Le robot d'assistance aux personnes âgées.....	182
Conclusion.....	194
 Présentation des auteurs	 197

REMERCIEMENTS

L

a réalisation de ce travail a nécessité la collaboration de plusieurs personnes et organismes, que nous tenons à remercier.

Il importe d'abord de remercier tous les auteurs collaborateurs pour leur engagement et leur responsabilité. Nous avons aimé le sérieux avec lequel ils ont répondu à l'appel, leur enthousiasme lié à ce projet de création d'un ouvrage collectif et la qualité de leurs contributions.

Nous remercions tout particulièrement le professeur Frank Violet de nous avoir réunis lors de la tenue du colloque international sur *La personne transformée* qui s'est tenu le 15 juin 2018, à la Faculté de droit, sciences économiques et sociales de l'Université catholique de Lyon.

Nous remercions les Instituts de recherche en santé du Canada (IRSC) pour leur contribution financière à nos travaux.

Introduction

Jean-Pierre Béland et Charles-Étienne Daniel

Relativement aux récentes percées biomédicales (implants, puces, nanorobots) et aux avancées de l'intelligence artificielle (avec le développement d'algorithmes performants capables de suggérer des diagnostics) et de la robotique (robot d'assistance médicale), la personne humaine se retrouve de plus en plus confrontée à l'ampleur des effets que de telles technologies peuvent entraîner sur elle. Tant l'intégration de différents dispositifs technologiques au sein de la personne humaine que leur utilisation dans les sociétés caractérisées par l'immixtion croissante du numérique induisent de profonds bouleversements à l'égard de son identité. Compte tenu de ces différentes conséquences, il n'est pas surprenant que la personne humaine transformée par l'incorporation d'implants et par ses interactions avec l'intelligence artificielle dans divers secteurs (santé, finance, transport, défense, justice) suscite de nombreux questionnements. Nous trouvons parmi ces questionnements certaines réflexions relativement à l'adéquation du recours au droit ou à l'éthique pour encadrer ou réprimer la transformation de la personne humaine par le développement technologique.

C'est précisément sur la base de tels questionnements que le colloque international *La personne transformée*, organisé par le professeur Frank Violet, s'est tenu le 15 juin 2018 à la Faculté de droit, sciences économiques et sociales de l'Université catholique de Lyon. Les réflexions issues de ce colloque constituent donc le fondement de cet ouvrage, et celui-ci propose par ses contributions d'approfondir d'une manière inédite les enjeux les plus intéressants pour pénétrer plus avant le sujet du droit et de l'éthique de la transformation de la personne humaine.

La première partie, « Qu'est-ce qu'une personne? », regroupe deux textes qui nous invitent respectivement à réfléchir de prime abord sur l'importance de la notion de personne et sur son rôle dans le discours normatif, tout en nous questionnant par la suite sur notre identité comme personne humaine par le biais de la subjectivité. Le premier texte du philosophe Georges A. Legault, intitulé « Définir la personne transformée: pourquoi? comment? », ouvre d'entrée de jeu la réflexion en cherchant à clarifier l'opération de la définition de ce qu'est une personne humaine si l'on veut comprendre les débats actuels autour de cette définition. Il montre ainsi que la notion de personne n'a pas de définition en soi, mais qu'elle prend forme dans différents discours normatifs (notamment de nature morale, juridique ou éthique). Ces discours conçoivent différemment la personne qui subit l'effet du développement des transformations de l'humain selon deux trajectoires, celle de la robotisation de l'humain par des implants technologiques et celle de l'humanisation du robot. Le second texte de Célia Vaz-Cerniglia et de Margherita Merucci Ciliberto, intitulé « La subjectivité à l'épreuve de la technologie », poursuit la réflexion du point de vue psychologique, en posant la question suivante: « Qu'est-ce qui motive la fascination pour les robots? » Par le biais de cette question, les auteures remettent en cause la capacité des robots à favoriser le développement de l'identité subjective de la personne humaine dans une réelle rencontre avec l'autre, dans une incomplétude similaire, sa solitude, ses limites. Elles cherchent à établir, d'une part, les conditions de possibilité (finitude, apprentissage, souffrance, mort) du développement de la personne comme personne humaine dans son rapport avec l'autre personne humaine. D'autre part, elles se demandent jusqu'à quel point l'interaction humain-robot (IHR) défavorise ce développement de l'identité personnelle humaine.

La seconde partie, « Quel diagnostic de crise en éthique et en droit? », regroupe des textes significatifs qui déterminent le problème crucial d'adaptation de l'éthique et du droit à la situation présente de la transformation de la personne, et qui permettent de proposer un traitement. Si, pour eux, le développement de ces nouvelles technologies de transformation humaine confronte l'univers éthique et l'univers juridique dans lesquels s'inscrit la notion de personne humaine, il force ces univers à entrevoir une redéfinition possible d'eux-mêmes. Le premier texte, intitulé « Personne transformée, médecine régénérative et humanisme » de Christian Byk, juge à la cour d'appel de Paris et président du Comité intergouvernemental de bioéthique de l'UNESCO (depuis 2017), aborde davantage l'incidence de la médecine régénérative et de la personne transformée sur le primat de la notion de la personne et le droit à la liberté. Pour l'auteur, un tel primat implique un appel à une nouvelle éthique humaniste. Si cette éthique s'incarne dans les droits de l'homme à exercer une liberté de choix, comment le droit pourrait-il alors limiter les impacts de la

médecine régénérative sur la personne humaine? Une telle logique exclut, par son discours normatif, la considération qu'une personne demeure immuable. Ce faisant, l'auteur souligne l'importance de la recherche d'un nouvel humanisme plus souple permettant de respecter la personne humaine qui se transforme par les nouvelles technologies. La contribution de Pascale Boucaud, «Le transsexualisme: transformation fondamentale de la personne», donne un exemple de problème d'adaptation juridique à prévoir pour la transformation technologique de l'humain. Si le transsexualisme a déjà un impact *réel* sur l'identité de la personne du point de vue juridique, qu'en sera-t-il donc pour la personne transformée dans le sens de la robotisation de l'humain? Le dialogue des jurisprudences en France et en Europe ouvre sur un questionnement au sujet des effets du changement de sexe qui nous laisse entrevoir toute la difficulté de l'adaptation du droit à la situation de la personne transformée. La contribution de Marjolaine Monot-Fouletier, intitulée «Personne transformée par l'intelligence artificielle et droit à la responsabilité» aborde davantage l'effet de la transformation technologique de la personne sur le droit à la responsabilité. Elle nous invite à considérer l'impact de la personne transformée par l'intelligence artificielle (IA) sur l'univers du droit de la responsabilité civile. Le point de départ de sa réflexion consiste en l'éventualité qu'une personne soit transformée en son corps ou assistée par une IA déterminante, qui pourra agir sur elle ou à travers elle sans l'intervention d'une décision humaine. «Dans ce cas, la personne engage-t-elle encore – et sous quelle forme – sa responsabilité, cette responsabilité qui est au cœur des relations humaines, au cœur du contrat social?», se demande-t-elle. Elle nous invite ainsi à poursuivre notre réflexion sur l'IA et le robot qui décident à la place de l'humain et à considérer l'incidence d'une telle décision dans l'univers juridique sous l'angle d'une responsabilité civile applicable à la machine. En sachant qu'il n'existe pas de forme unique et universelle d'intelligence artificielle, l'auteure souligne qu'il serait utopique et peu pertinent de poser un régime juridique de responsabilité unique pour des réalités plurielles. De fait, elle s'interroge également sur l'adaptabilité du droit à la responsabilité au contexte d'action des nouvelles technologies et sur la possibilité d'élaborer un régime spécial de responsabilité.

La troisième partie de l'ouvrage, «Quelles pistes de solution?», vise à donner quelques exemples significatifs de réponses à la demande de redéfinition du droit et de l'éthique invoquée dans la seconde partie, en vue d'assurer le développement responsable des technologies induisant une transformation de la personne. Le texte de Mathieu Guillermin et de Mathieu Lefort, intitulé «De la signification éthique des limites», pose la question suivante: «Quelles sont les voies de développement les plus intéressantes, celles à éviter?» Le propos de l'auteur montre l'importance de réaliser une réflexion éthique

approfondie reconnaissant les limites techniques de l'IA et des robots. Il établit différents types de limites (autonomie des machines, singularité technologique et transhumanisme...) comme étant à la base de l'idée de transgression de la limite (intelligence humaine obsolète). Ensuite, il met en question la limite au-delà de la transgression. Susceptible d'être ainsi affectée et vulnérable par ces technologies de la transgression, la personne humaine doit voir son rôle d'évaluateur réaffirmé et renforcé, dans sa capacité à accepter ou à refuser en conscience les résultats des biotechnologies et des procédés algorithmiques des IA. Le texte de Charles-Étienne Daniel, intitulé «Quelle(s) forme(s) de gouvernance juridique pour l'encadrement de la robotique et de l'IA?», constitue un exemple de redéfinition du cadre méthodologique de l'univers juridique par une approche plurielle en trois catégories (droit dur, droit souple et droit mou). Par l'illustration de différentes initiatives et instruments normatifs récemment adoptés en matière d'IA et de robotique, l'auteur montre comment les trois différentes formes de droit redéfinissent l'univers juridique traditionnellement mobilisé dans l'encadrement de ces technologies. Le texte de Jean-Pierre Béland, intitulé «La personne transformée: comment peut-on mettre une limite?», constitue un exemple plus précis de réponse à la question de la redéfinition de l'éthique pratique en cherchant à clarifier par quel processus simplifié d'évaluation on peut arriver à une décision d'interdire, de freiner ou encore de favoriser la transformation humaine, et quel type d'évaluation choisir (l'évaluation morale, l'évaluation juridique ou l'évaluation éthique). Il retient alors quatre exemples de produits qui sont en voie de commercialisation ou en développement pour la transformation humaine (l'implant de mémoire, les systèmes d'aide à la décision ou IA, les implants RFID et les robots d'assistance aux personnes âgées) dont il est question pour illustrer les trois types d'évaluation.

Dans le contexte du développement responsable des nouvelles technologies pour l'augmentation/réalisation humaine, le lecteur pourra ainsi mieux comprendre la spécificité de l'évaluation éthique des impacts de ce développement des nouvelles technologies sur la personne. Les différentes contributions à cet ouvrage offrent ainsi de mieux cerner i) l'importance de la notion de personne et son rôle dans le discours normatif, ii) la demande de redéfinition de l'éthique humaniste et du droit à la responsabilité par rapport aux transformations que subit la personne humaine, et iii) les manières dont le droit et l'éthique peuvent effectivement se redéfinir en réponse aux changements engendrés par le développement technologique.

CHAPITRE PREMIER

Définir la personne transformée : Pourquoi ? Comment ?

Georges A. Legault

INTRODUCTION

La transformation génétique des organismes vivants a ouvert le débat éthique sur le clonage végétal, animal et humain. Tant que la transformation du vivant concernait les végétaux, les enjeux éthiques portaient essentiellement sur les impacts sur la santé, la sécurité et l'environnement. Mais le clonage des animaux, comme celui de Dolly en 1996 et de deux macaques¹ en 2018, soulève à nouveau la possibilité de cloner l'humain et, dès lors, nous force à nous prononcer sur le statut du clone. Si la transformation du vivant par l'intervention humaine nous oblige à repenser la définition de l'humain, celle de la machine qui conduit au robot animé par l'intelligence artificielle soulève aussi un débat analogue. En effet, le 25 octobre 2017, l'Arabie saoudite octroyait au robot Sophia la nationalité saoudienne². Deux questions se posent à partir de ces exemples : « Un clone humain est-il encore une personne humaine ? » « Le robot Sophia peut-il être considéré comme une personne humaine ? » Pour répondre à cette question, il nous faut au

-
1. <<https://www.quebecscience.qc.ca/normand-baillargeon/doit-on-encore-sinquieter-du-clonage-humain>>.
 2. <https://www.lemonde.fr/idees/article/2017/11/04/sophia-robot-saoudienne-et-citoyenne_5210094_3232.html>.

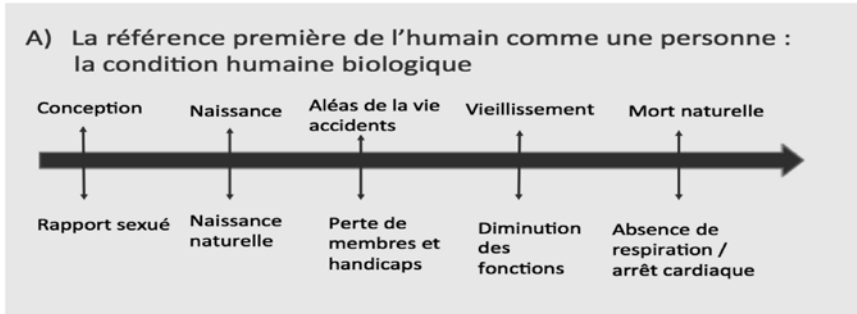
préalable définir ce qu'est une personne humaine. Toutefois, cette opération ne peut réussir tant qu'une question d'arrière-plan n'est pas élucidée. En effet, pour bien définir la personne, il faut au préalable clarifier l'opération même de la définition. Autrement dit, il faut répondre aux questions suivantes : « qu'est-ce que définir? », « comment le faire? » et « pourquoi en avons-nous besoin? ».

Le présent chapitre vise à éclairer ces questions d'arrière-plan, qui sont essentielles lorsque l'on veut comprendre les débats autour de la définition de la personne. Dans un premier temps, nous proposons d'examiner comment l'intervention technique est venue à remettre en question, au fur et à mesure de son développement, notre mode usuel de définition de la personne basé sur notre expérience biologique de la vie. Ce bref parcours des transformations de l'humain sert à dégager une première trajectoire, la robotisation de l'humain, à laquelle on pourra confronter celle de l'humanisation du robot. Si ces deux trajectoires imposent chacune à leur façon la nécessité de redéfinir la personne, il faut comprendre pourquoi il devient nécessaire de définir la personne en cherchant à clarifier le rôle de la définition. En effet, définir la personne n'est pas neutre ; si nous avons besoin de le faire, c'est parce que cette définition est centrale aux différents discours normatifs. Dans la seconde partie, nous analyserons le rôle de la notion de personne dans les discours normatifs que sont la morale, le droit et l'éthique. Enfin, dans la troisième partie, nous aborderons la question du comment définir et nous dégagerons certains problèmes liés à l'opérationnalisation de définitions dans les contextes sociaux et culturels.

1. L'INTERVENTION TECHNIQUE ET LA REMISE EN QUESTION DE LA DÉFINITION DE LA PERSONNE

Il y a 70 ans, la définition de la personne ne faisait pas problème, puisque nos références pour la définir provenaient de l'expérience de notre condition humaine. Notre expérience de la personne humaine comme être biologique va de sa conception jusqu'à sa mort. Ainsi, la personne est conçue par des parents après un rapport sexué ; c'est l'union du sperme et de l'ovule qui constitue le début de la personne. Viennent ensuite la naissance et la croissance jusqu'à l'âge adulte, la maturité et le vieillissement. Évidemment, tout au long de notre parcours de vie, il peut y avoir des aléas incontrôlables. Dès la naissance, on peut naître handicapé, comme il peut y avoir des accidents nous privant de certains membres. Avec le vieillissement, l'usure du temps se manifeste tant dans le corps que dans l'esprit. Enfin, l'aventure humaine se termine par l'arrêt du cœur et de la respiration en « rendant son dernier souffle ». Tant que rien ne vient perturber cette expérience biologique, elle constitue la matrice « naturelle » de notre conception de l'humain. Le schéma ci-dessous illustre la

conception de la personne à partir de cette expérience biologique. Or c'est cette conception qui va se modifier avec les interventions techniques.



En 1960, la première pilule contraceptive est mise en vente. Grâce à cette invention technique pharmacologique, il est possible d'empêcher la grossesse. Cela modifie profondément la nature du rapport sexué, car la fonction procréation et la fonction plaisir deviennent deux finalités différentes. La contraception qui résultait du hasard de la vie après un rapport sexué devient alors objet de contrôle : c'est la planification des naissances. En 1978, la fécondation in vitro, communément appelée « le bébé-éprouvette », est rendue possible grâce aux développements des techniques de fécondation hors du corps humain. À l'époque, plusieurs se questionnaient sur le statut de ce bébé et de son équivalence avec la fécondation naturelle. Enfin, en 1996, Dolly est le premier mammifère à être cloné. Contrairement à l'insémination artificielle, le clonage ne fait appel qu'à un seul noyau cellulaire techniquement déposé dans un ovule énucléé. L'ovule est uniquement un réceptacle du noyau cellulaire. Il n'y a que la reproduction de l'information génétique du donneur. Ces trois exemples nous permettent de constater que la remise en question de la définition de la personne apparaît chaque fois que l'intervention humaine par le biais de la technique intervient dans notre processus biologique naturel. En intervenant déjà à la conception, l'humain prend une distance des contraintes de son état biologique naturel. De ce fait, il devient responsable de l'effet de son intervention sur la condition humaine et de tout effet qui pourrait modifier le statut de personne humaine.

Lors du trajet de la vie, de la conception à la mort, il survient plusieurs événements qui modifient notre condition biologique. Certes, nous sommes tous vulnérables aux diverses maladies, mais il arrive des accidents qui nous privent de certains membres. L'histoire centenaire de l'Association des amputés de guerre est révélatrice non seulement de l'importance du développement technologique pour les amputés, mais aussi des luttes contre les préjugés associés

à avoir un « membre artificiel »³. La lutte de l'Association pour aider les amputés de guerre s'est élargie en 1962 aux civils. C'est en 1957 que la première prothèse articulée avec genou et cheville articulée par contrôle hydraulique a été approuvée au Canada. Depuis lors, les progrès ont été remarquables et il existe sur le marché un large éventail de prothèses articulées.

L'histoire du coureur Oscar Pistorius⁴ montre comment l'intégration de l'artificiel au naturel soulève des problèmes de définition et de classification. Pistorius a été amputé des deux jambes à l'âge de 11 mois, mais cela ne l'a pas empêché de devenir un coureur grâce à ses prothèses composées de deux lames fines de carbone. Or il veut participer aux Jeux olympiques pour personnes valides et non pas être limité aux jeux paralympiques. La question est simple : peut-on admettre, suivant les règlements, la participation de Pistorius ? Autrement dit, est-il un coureur comme un autre ? Toute la question repose sur la définition que l'on donne à un coureur pouvant participer aux Olympiques. Le critère utilisé est celui de la performance, car il s'agit d'établir si les membres artificiels donnent un avantage au coureur comparativement aux autres coureurs. Malgré un premier refus essuyé en 2008, après un rapport montrant l'avantage suscité par des membres artificiels, d'autres études en ont démontré l'équivalence, ce qui lui a permis de participer à d'autres Jeux olympiques. L'histoire de Pistorius montre clairement comment l'amélioration des capacités humaines est au cœur du problème de définition entre naturel et artificiel.

Le développement technologique vient aussi au secours des personnes malades ou vieillissantes qui perdent différentes capacités motrices et visuelles. Aujourd'hui, la greffe d'organes est une pratique reconnue qui fait partie de l'offre des services en santé. La première greffe de rein réussie chez l'homme a été réalisée en 1959. Au Canada, en soixante ans, on a réussi la greffe du cœur, du foie, du pancréas, du cœur-poumon et, plus récemment, la première greffe du visage. Qui aurait pu penser que l'on puisse arriver à changer chez l'humain des « pièces défectueuses » comme on le fait pour d'autres machines ? Tant qu'il s'agit de transplantation d'organes humains, on demeure dans la sphère biologique, et même si le fait de posséder un organe d'un autre peut susciter certaines appréhensions, il n'en demeure pas moins que c'est du biologique qui est incorporé en nous. Mais qu'arrive-t-il si l'organe n'est plus biologique ? Par exemple, on peut greffer des mâchoires en titane ou encore des cœurs mécaniques. Les transplantations d'organes mécaniques soulèvent le questionnement sur l'intégrité de la personne humaine. En effet, selon l'importance de l'organe biologique substitué par un objet mécanique, on se demandera si c'est encore une personne humaine. Évidemment, certains voient déjà se pointer à

3. <<https://www.amputesdeguerre.ca/a-propos/historique>>, voir 1957.

4. <https://fr.wikipedia.org/wiki/Oscar_Pistorius>.

l'horizon les transplantations dans le cerveau. Un second enjeu, outre l'intégrité, que soulève l'incorporation de la mécanique dans le corps humain, est similaire à celui des prothèses et renvoie à l'amélioration des capacités humaines grâce à l'incorporation d'organes mécaniques plus performants que les organes naturels. Nous sommes dans la trajectoire de la robotisation de l'humain.

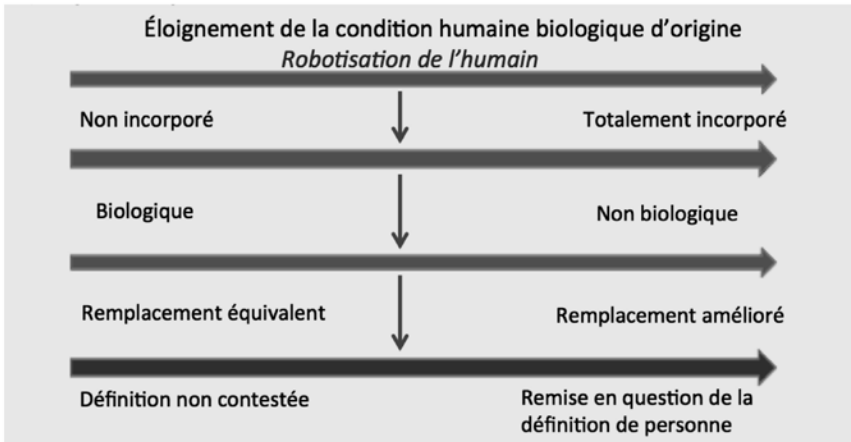
Il ne faut pas oublier que c'est la redéfinition de la mort qui a rendu possibles les transplantations d'organes humains. Tant que la mort était définie par l'arrêt de la respiration ou l'arrêt du cœur, il était impossible de recueillir les organes pour les transplanter. Or, le respirateur artificiel a figuré au cœur de la redéfinition de mort. En effet, comme le précise Isabelle Cuchet :

Jusqu'en 1968, les choses étaient simples : un individu était déclaré mort quand il ne respirait plus et que son cœur ne battait plus. Une fois constaté l'arrêt de ces deux fonctions vitales (respiratoires et sanguines), le décès était prononcé. Mais avec les progrès des techniques de réanimation (respirateur artificiel, massage cardiaque), qui permettent de relancer le cœur et de maintenir un individu en vie avec un cœur et des poumons défaillants, le moment de la mort a été précisé : aujourd'hui, c'est l'état du cerveau qui sert d'ultime critère légal de la mort de l'individu, quand bien même le cœur et les poumons marchent toujours...⁵

Alors que la mort se constatait par les signes biologiques de la respiration et du pouls, aujourd'hui la mort n'est déclarée qu'après une batterie de tests comme l'angiographie cérébrale et les électroencéphalogrammes.

Le développement technologique sans précédent à partir des années 1950 est intervenu de plus en plus dans le parcours de nos vies biologiques. De la naissance jusqu'à la mort, ce développement nous éloigne de la définition de la personne comme être biologique gouverné par les lois de la nature. Ce bref parcours d'interventions technologiques dans la vie biologique permet de dégager trois facteurs qui incitent à la redéfinition de la personne. Le tableau ci-dessous illustre les facteurs en cause.

5. I. Cuchet (juillet 2017), « Quand est-on déclaré mort? », *Science et Vie*, <<https://www.science-et-vie.com/corps-et-sante/quand-est-on-declare-mort-9072>>.



Le premier facteur à considérer dans la transformation de l'humain par la technique est celui de son incorporation. Par exemple, une prothèse qui n'est pas intégrée au corps humain ne pose pas de problème. En effet, porter des lunettes ou des lentilles amovibles ne remet pas en question notre statut. Ceci peut s'expliquer à la lumière des catégories philosophiques de la substance et de l'accident. Chez Aristote (384-322 av. J.-C.), la « substance » constitue l'identité de la chose, alors que l'« accident » renvoie à ce qui est à la surface et qui change avec le temps. Ainsi, lorsqu'on compare des photos d'un enfant avec celles de la personne à sa maturité ou au stade de sa vieillesse, on se rend compte que beaucoup de choses ont changé. Cependant, malgré la difficulté à reconnaître la personne par les signes externes, il n'en demeure pas moins que nous sommes convaincus que celle-ci est restée la même malgré le passage du temps. Cela nous aide à comprendre pourquoi une prothèse qui est incorporée dans le corps humain soulève la question de l'intégrité de la personne. Autrement dit, plus la technique est incorporée dans l'humain, plus on se questionne sur son identité, parce que la transformation pourrait atteindre la substance de l'humain, et non pas seulement un de ses accidents.

Le second facteur à prendre en compte est celui des matériaux utilisés dans la transformation. La perception des transplantations varie selon la nature de la greffe. Lorsqu'il s'agit de transplanter, par exemple, un cœur humain, on ne voit plus en principe de problèmes, même si certains greffés ont de la difficulté à vivre avec une partie du corps d'un autre. La pénurie d'organes humains pour les besoins de transplantation a suscité beaucoup de recherches liées à la modification génétique d'organismes pour favoriser la xéno greffe. « Le porc l'avenir de l'homme » est le titre accrocheur d'un article portant sur les

recherches actuelles menées en Belgique sur les xénogreffes⁶. Le titre nous montre que même si nous demeurons dans une greffe biologique, le bris des barrières animales peut faire problème. Lorsque la greffe est un organe mécanique, le problème se pose cette fois en matière de bris des barrières biologiques et mécaniques. Jusqu'où peut-on remplacer les organes humains par des organes mécaniques? Cette question comporte deux volets, celui de la quantité avec le nombre d'organes mécaniques incorporés dans un même individu, et celui de la qualité, notamment l'importance dans la vie biologique de l'organe remplacé.

Le troisième facteur est lié à l'amélioration de l'humain issue des diverses incorporations. La technique en génétique nous permet d'intervenir en modifiant le processus biologique. Cela a soulevé le débat éthique autour de la thérapie germinale et la thérapie somatique. La principale différence entre la thérapie somatique ou la thérapie germinale ne se situe pas sur le plan de l'intervention génétique elle-même, mais plutôt sur celui de l'incidence de cette modification génétique sur la personne seulement (somatique) ou sur sa descendance (germinale). Si la thérapie germinale est interdite en France⁷, c'est justement, en reprenant les catégories d'Aristote, parce que son incidence sur la descendance est considérée comme un facteur relevant de la substance de la personne, alors que l'effet sur le seul individu est de l'ordre de l'accident. Ainsi, tant que la modification consiste à corriger une maladie pour l'individu, cela est admissible parce que la transformation est équivalente à un traitement de maladie. Cependant, dès que la modification donne un avantage à la descendance, cela soulève des réticences. Le même questionnement se pose pour toute prothèse ou tout implant qui pourrait offrir à son utilisateur une performance améliorée, comparativement à la moyenne des performances humaines.

L'éloignement de notre conception biologique de la personne par l'intervention technologique met en question la définition de personne, à des degrés divers, selon la portée de l'intervention en regard des trois facteurs précisés. Plus l'intervention technologique incorpore totalement un produit non biologique améliorant les performances du porteur comparativement aux autres, plus la question de savoir s'il s'agit encore d'une personne se posera. Les transformations de notre expérience biologique ne sont pas les seules à faire pression sur la définition de la personne. Les développements récents de la robotique qui ont pour effet de sortir les robots d'endroits confinés pour interagir avec les personnes conduisent à intégrer dans le robot des fonctions analogues à celles de l'humain. Grâce à des capteurs et à la reconnaissance faciale, le robot

6. <<https://www.lecho.be/culture/general/le-porc-est-l-avenir-de-l-homme/9983279.html>>.

7. Garantir les conditions de développement et la sécurité sanitaire des produits de thérapie génique et cellulaire, <https://www.senat.fr/rap/r95-053/r95-053_mono.html#toc16>.

intègre des fonctions des sens de la vue et du toucher et grâce à l'apprentissage profond, le robot peut apprendre de ses erreurs et corriger son programme de base en conséquence. Plus le robot reproduit les fonctions humaines, plus la question de sa reconnaissance comme une personne se pose. Ainsi, les deux trajectoires de la robotisation de l'humain et de l'humanisation du robot nous conduisent à reconnaître que la définition classique de la personne ne va pas de soi. C'est pourquoi la redéfinition de la personne est à l'ordre du jour. Cependant, on peut se demander pourquoi il est si important de redéfinir la personne et de clarifier ainsi l'ambiguïté sur le statut de personne. Pour répondre à cette question, il faut situer le contexte global du rôle des définitions dans la conduite humaine en société.

2. POURQUOI DÉFINIR LA PERSONNE ?

Définir, selon le dictionnaire Larousse, c'est : « Énoncer la nature, les qualités, les propriétés essentielles de l'être ou de la chose que le mot désigne ; donner sa définition : *Définir le mot "table"* »⁸. La définition a donc pour but de donner l'identité de ce qui est défini et ainsi de le distinguer des autres choses qui ont leur identité propre. Autrement dit, une table n'est pas une chaise, même si parfois on peut s'asseoir dessus. Encore une fois, on retrouve l'idée de « substance », puisque la définition renvoie à « propriétés essentielles », et que cette expression est utilisée comme équivalent chez Aristote. La définition sert à distinguer une chose d'une autre, mais on peut se demander s'il s'agit de son unique fonction. Les propos de Donald Trump, président des États-Unis, permettent d'illustrer en quoi une définition n'est pas une opération neutre. « We have people coming into the country, or trying to come in – and we're stopping a lot of them – but we're taking people out of the country. You wouldn't believe how bad these people are. These aren't people. **These are animals** »⁹. En qualifiant les immigrants clandestins d'animaux et non d'hommes, le président laisse donc entendre qu'il est justifié de traiter les immigrants comme il le fait. La distinction entre « personne » et « non-personne » a donc une importance fondamentale dans nos façons d'agir envers elles.

Dans toutes les cultures, à travers des formes différentes, il existe des légendes, des textes, des paroles dont la finalité est celle de guider les rapports interpersonnels afin d'assurer le meilleur vivre-ensemble. Dans nos sociétés d'aujourd'hui, on trouve trois types de textes ou de paroles guidant nos

8. <<https://www.larousse.fr/dictionnaires/francais/définir/22691>>

9. <<https://www.vox.com/first-person/2018/5/18/17369044/trump-ms-13-gang-animals-immigrants>>.

conduites humaines : la morale, le droit et l'éthique. Parce que ces textes visent à guider la conduite, malgré des modalités très différentes, ils sont regroupés sous le vocable de langages normatifs¹⁰. Puisqu'il s'agit de guider les conduites, la distinction entre ce qui est une personne et ce qui n'en est pas une est fondamentale. Cependant, le rôle de la définition est différent d'un langage à l'autre, compte tenu de la portée normative de chacun d'eux.

En morale, religieuse ou philosophique, la définition de la personne est fondamentale, puisqu'elle permet de définir, en premier lieu, la communauté morale. En effet, la morale s'adresse à tous ceux qui sont considérés comme des personnes. Certains sont surpris de constater que, malgré l'importance de la morale philosophique dans la société grecque, l'esclavage était de mise pour tout peuple vaincu. Ainsi on pouvait les traiter comme des propriétés. La communauté morale sera plus ou moins étendue selon la définition proposée et son application dans un contexte particulier. Plus la définition de personne sera restreinte à un groupe particulier, moins les membres du groupe auront d'obligations envers les autres.

La définition de la personne est aussi importante en morale, parce qu'elle définit l'idéal de la personne que l'on cherche à actualiser en respectant la morale. Toute morale se construit autour d'une distinction entre « ce qui doit être » et « ce qui est ». Dans toute morale, on reconnaît que l'être humain n'est pas parfait, puisqu'il n'agit pas spontanément en fonction de ce qui est le meilleur pour assurer le vivre-ensemble en société. On reconnaît donc en morale l'écart entre ce que la personne devrait faire si elle agissait totalement en conformité avec sa nature et ce qu'elle fait réellement. La représentation que l'on se fait de l'être humain dans les théories philosophiques ou théologiques est donc la clef de voûte du langage moral. En effet, le langage moral est hiérarchisé. Le point de départ est la représentation de la personne issue de la position philosophique ou théologique. De cette représentation, on déduit toutes les obligations morales envers les autres afin d'assurer le vivre-ensemble. Pour que la morale soit effective, il faut donc que les personnes développent leur capacité de raisonnement moral et les bonnes habitudes morales pour bien agir en société. Certaines morales religieuses soumettront les manquements à la morale à des peines subséquentes après la mort.

Le langage normatif du droit ressemble partiellement au langage moral, parce qu'il impose des obligations de faire pour assurer le vivre-ensemble. Contrairement à la morale, le droit ne repose pas sur une représentation philosophique de la personne. En effet, le droit génère les distinctions entre

10. C. Thibierge (2009), « Rapport de synthèse », dans *Le droit souple*. Actes du colloque « Le droit souple », Association Henri Capitant, Dalloz, Paris, p. 151.

personnes et non-personnes en fonction de la règle qu'il veut établir. Autrement dit, les définitions juridiques se font en cohérence avec l'obligation que l'on veut décréter. Par exemple, le droit reconnaît le statut de personne à des organismes qui sont alors désignés comme des personnes morales. Du point de vue juridique, ces organismes sont des personnes juridiques à qui l'on attribue des droits et des obligations. Grâce à cette reconnaissance du statut de la personne morale, ces organismes peuvent être poursuivis au même titre qu'un individu. Le statut de l'embryon est un autre exemple. Le droit ne cherche pas à statuer sur le statut ontologique du fœtus pour savoir s'il s'agit d'une personne dès la conception. Pour le droit, la personnalité juridique d'un enfant est reconnue dans la section portant sur les successions. En effet, en droit, il faut trancher la question à savoir si un fœtus peut hériter. L'article 617 du Code civil précise : « Peuvent succéder les personnes physiques qui existent au moment de l'ouverture de la succession, y compris l'absent présumé vivant à cette époque et l'enfant conçu, mais non encore né, s'il naît vivant et viable. » Autrement dit, tant l'enfant n'est pas né vivant et viable, il n'a aucune existence juridique. De plus, le droit se distingue des langages normatifs de la morale et de l'éthique par sa force exécutoire. L'efficacité du langage juridique repose évidemment sur la volonté des citoyens de se conformer, mais la plupart du temps sur sa capacité de contraindre les personnes à exécuter leurs obligations ou à fournir une compensation à défaut d'exécution.

Le langage normatif de l'éthique, contrairement à la morale et au droit, ne guide pas la conduite par le biais d'obligations, mais par des valeurs partagées. Bien qu'il existe plusieurs façons de concevoir la valeur, elle représente, selon notre approche en éthique appliquée, essentiellement un idéal que nous cherchons à rendre actuel dans nos façons de vivre¹¹. Cette approche rejoint celle de Paul Ricœur, qui définissait ainsi l'éthique : « Cette visée éthique est *visée de la "vie bonne" avec et pour autrui dans des institutions justes*¹² » En effet, la notion de « vie bonne » provient d'Aristote, et elle concerne les choix de vie qui façonnent le vivre-ensemble, tant du point de vue personnel que du point de vue collectif. Tout choix a sa visée, et c'est pourquoi les choix de vie sont guidés par des jugements de valeur qui déterminent le meilleur pour nous et les autres. Comme dans le langage moral, celui de l'éthique reconnaît l'écart entre l'idéal de la vie bonne avec et pour autrui dans les institutions que nous mettons en place et la réalité de ce que nous arrivons à construire ensemble.

En éthique, on ne cherche pas, comme en morale et en droit, à définir la personne et ses caractéristiques. C'est par la valeur de « dignité humaine » que

11. G.A. Legault (1999), *Professionalisme et délibération éthique*. Québec, Presses de l'Université du Québec, p. 290.

12. J.-M. Gaté (2001), « Entretien avec Paul Ricœur », *Le Philosophoire*, 3, p. 9-21.

se fait le lien entre la définition de la personne et la valeur. En effet, si l'on reprend les mots du président Trump sur les immigrants, nous voyons que pour guider son comportement envers eux, il juge que les migrants clandestins ne doivent pas être traités avec la dignité que l'on accorde aux personnes. Le jugement de valeur sur eux est celui porté envers les animaux méchants. Ce jugement de valeur permet alors de justifier des traitements indignes de l'humain.

Les langages normatifs de la morale et du droit guident la conduite en imposant une obligation à suivre. Ces langages incitent donc les personnes à se conformer à ces obligations. Autrement dit, dans le choix d'agir, la personne se demande si l'action envisagée correspond ou non à l'obligation juridique ou à l'obligation morale. Dans le champ des valeurs, la décision d'agir repose sur un autre mode de raisonnement pratique que celui de la conformité à la règle. En effet, les jugements de valeur au cœur des décisions d'agir se posent sur les impacts prévisibles de ce que nous envisageons de faire. Toute action que nous décidons de mettre en œuvre a des conséquences positives et négatives sur des personnes, des institutions ou l'environnement qui nous entourent. L'évaluation des impacts sur les personnes par le biais de la valeur de « dignité humaine » cherche à établir dans quelle mesure les impacts augmentent ou diminuent cette valeur. Puisque toute décision augmente et diminue des valeurs, la décision finale devra pondérer les jugements de valeur et établir quel jugement de valeur a une priorité sur l'autre. Pour assurer la qualité de nos relations avec les autres, il est important que les raisons justifiant la priorisation des jugements de valeur soient précisées à toutes les parties prenantes. Dans certaines décisions, on peut prioriser la valeur de la sécurité sur celle de la dignité, mais ce choix a besoin d'être justifié, et les raisons partagées pour assurer son acceptabilité.

La réponse à la question de départ « Pourquoi définir la personne? » s'éclaire lorsqu'on étudie le rôle joué par cette notion dans les langages normatifs qui structurent nos façons de vivre ensemble. Définir la personne n'est pas neutre, car cette définition permet de distinguer deux groupes, ceux qui sont considérés comme des personnes et qui méritent d'être traités en tant que telles, et ceux qui ne font pas partie des personnes et qui méritent, de ce fait, un traitement différent. Compte tenu de l'importance de la définition de la personne dans les discours normatifs, il faut porter une attention particulière aux différentes approches proposées pour définir la personne.

3. COMMENT DÉFINIR LA « PERSONNE » ?

3.1. L'approche définitionnelle substantive ou attributive

Dans la tradition de la philosophie morale, avant le changement de paradigme apporté par le philosophe Emmanuel Kant (1724-1804), la représentation de l'être humain était la clef de voûte de tout système philosophique, puisque les obligations morales ne pouvaient découler qu'à partir de l'exigence de réaliser la représentation de l'humanité comme définie par la philosophie ou la théologie. Comme nous l'avons esquissé rapidement, la conception que l'on se fait de l'humain en philosophie classique a deux volets : celui de représenter l'humain réel (ce qu'il est), et celui de fixer l'humain idéal, ce qu'il peut devenir (devoir-être). Autrement dit, la représentation que l'on se fait de l'essence de l'humain n'est pas fixe comme le serait celle d'un objet. L'humain est en devenir, et la quête de la sagesse consiste à devenir, au cours de notre vie, cet humain idéal. On retrouve ici l'exigence fondamentale de la morale classique : tu dois devenir l'humain que tu as le potentiel d'être.

Puisque toute la morale repose sur la représentation de l'humain, la question de la validité de cette représentation va se poser. En effet, comment peut-on affirmer que la définition de l'humain véhiculée dans la représentation philosophique est la bonne ? Chercher à définir la personne à partir de caractéristiques inhérentes à notre nature humaine et connaissables par nous, c'est la recherche d'une définition substantive. La définition substantive est créée à partir de caractéristiques jugées essentielles à l'être humain, qui permettent de distinguer la personne humaine d'autre chose. De plus, ces caractéristiques ne sont pas fixes, mais en mouvement ; autrement dit, la définition renvoie à une capacité de devenir, c'est-à-dire un humain potentiel. La définition substantive prétend décrire la structure du réel, puisque la représentation proposée est jugée vraie. Les philosophes classiques pensaient pouvoir atteindre une représentation de l'humain qui correspond à la structure du réel. Pour justifier la véracité de la représentation, les philosophes proposent différentes théories explicitant la capacité de connaître chez l'humain. C'est la branche de la philosophie nommée épistémologie, qui en grec signifie la parole (*λόγος*) de la connaissance (*ἐπιστήμη*). La définition substantive est centrale dans la philosophie, mais elle a été remise en question par le philosophe Emmanuel Kant, qui a démontré l'impossibilité pour l'humain d'atteindre la connaissance du réel tel qu'il est en lui-même. Pour lui, la représentation que nous nous faisons du réel est une construction à partir de notre expérience humaine. Notre connaissance du monde ne peut pas dépasser les limites de notre expérience humaine.

L'appel à la définition substantive est très présent dans les débats touchant la représentation de l'humain. Les débats liés à l'acceptabilité de l'avortement illustrent bien cet enjeu. Dans sa catéchèse sur la Place Saint-Pierre, le pape François a soutenu que le commandement « Tu ne tueras point » s'appliquait à l'avortement. Il a même comparé le fait d'avorter à l'utilisation de tueurs à gage pour régler un problème¹³. Une telle interprétation repose évidemment sur la reconnaissance du fœtus comme une personne. Pour soutenir ce statut, on réfère souvent à l'argument suivant :

Le phénomène du vivant humain est caractérisé par le développement et la croissance. Le sujet en développement est un sujet qui croît, qui devient ce qu'il est déjà au départ. Il contient, à l'origine, ce qu'il va actualiser au cours de son devenir. C'est donc le même sujet qui devient de la conception à la naissance. On peut donc affirmer sans se tromper : le vivant devient ce qu'il est. En ce sens, on peut donc dire que le zygote est un individu humain. Il est en puissance d'être à terme ce qu'il est déjà au début, au moment de la fécondation¹⁴.

Ainsi, la reconnaissance du statut de personne repose sur le potentiel génétique du fœtus. En tant qu'être humain en devenir, il est déjà une personne. Dans l'Église catholique, la définition substantive est vraie, puisqu'elle est fondée sur la parole de Dieu et, dès lors, sur la foi des fidèles.

La définition substantive veut ainsi ancrer les caractéristiques de la personne dans la réalité. Le but poursuivi ici est de rendre « objective » la définition et ses implications morales par la suite. Autrement dit, la définition substantive, puisqu'elle reflète la réalité, devient un fondement universel pour les obligations morales. Or, les débats épistémologiques autour de la définition de la personne traversent l'histoire de la philosophie et ils sont encore présents dans les débats moraux contemporains comme ceux soulevés par le transhumanisme. Le dictionnaire Linternaute.com définit ainsi le transhumanisme : « Courant culturel et intellectuel répandu dans le monde entier. L'une des volontés du mouvement est de promouvoir l'amélioration de la vie via l'utilisation de la technologie, éliminer le vieillissement et augmenter les capacités humaines, que ce soit au niveau intellectuel, psychologique ou physique¹⁵. » Pour justifier l'ensemble de sa position et rejeter la position humaniste, le transhumaniste Kurzweil déclare : « Then perhaps our basic disagreement is over the nature of being human. To me, the essence of being human is not our limitations – although we do have many – it's our ability to reach beyond our limitations. We didn't stay on the

13. <<https://www.la-croix.com/Religion/Catholicisme/Pape/Avortement-tueur-gages-quadit-pape-Francois-2018-10-10-1200975037>>.

14. <<http://www.lapresse.ca/debats/le-cercle-la-presse/actualites/201210/09/48-1323-lem-bryon-est-il-un-individu-humain.php>>.

15. <<https://www.linternaute.fr/dictionnaire/fr/definition/transhumanisme>>.

ground. We didn't even stay on the planet. And we are already not settling for the limitations of our biology¹⁶. » Bien que l'auteur souligne comment la divergence sur la définition de la personne humaine est au cœur de la différence entre humanisme et transhumaniste, il prétend dans son œuvre que c'est celle qu'il propose qui correspond à l'expérience humaine et qui devrait être retenue.

Fonder une obligation morale sur une représentation vraie de la personne humaine constitue la quête de l'approche de la définition substantive de la personne. Cependant, ces tentatives conduisent inévitablement à des discours de sourds entre les différentes positions substantives. S'ensuit alors l'impasse dans le débat public et la confrontation rendant stérile toute visée d'une approche morale partagée¹⁷.

L'approche retenue en droit n'est pas celle de la définition substantive, mais la définition attributive. Cette dernière ne vise pas à donner une définition objective et incontestable, mais plutôt une définition qui est en cohérence avec l'énonciation de la norme. Alors que la morale donne un statut particulier à la définition, la posant comme fondement des obligations, le droit intègre la définition de la personne dans la solution globale du problème qu'il veut résoudre par la loi. Dans l'exemple de l'avortement, la question n'est pas pour le droit de savoir si le fœtus est objectivement une personne avant de statuer sur la légitimité de l'avortement ou sur la possibilité d'un fœtus d'avoir un héritage. L'énonciation de la règle de droit vise à résoudre des conflits potentiels entre les personnes. Il ne faut pas oublier que la non-reconnaissance du fœtus comme personne, mais uniquement de l'enfant né vivant et viable, est une solution aux problèmes de l'héritage. En effet, si un fœtus pouvait hériter de l'un de ses parents qui décèdent avant lui, il aurait des droits sur l'héritage. Mais s'il meurt avant de naître vivant et viable, qui héritera du fœtus? Dans ce contexte, il est plus simple de reconnaître que seul l'enfant né vivant et viable aurait le droit d'hériter. Cependant, en affirmant que seul l'enfant né vivant et viable est une personne ayant des droits juridiques, cela suppose que le fœtus n'est pas protégé par le droit. Pour le pape François, la « suppression de la vie humaine dans le sein maternel au nom de la sauvegarde d'autres droits¹⁸ » est inacceptable. Cette prise de position est conséquente avec la définition substantielle de la personne soutenue par l'Église catholique. Mais énoncer une règle de droit, c'est toujours balancer les droits des uns et les droits des autres.

16. R. Kurzweil (2005), *The Singularity Is Near: When Humans Transcend Biology*, New York, Viking Penguin.

17. J. Patenaude, G.A. Legault, M. Parent et J.-P. Béland (2011), « Moral arguments in the debate over nanotechnologies: Are we talking past each other? », *NanoEthics*, vol. 5, n° 3, p. 285-293.

18. <<https://www.la-croix.com/Religion/Catholicisme/Pape/Avortement-tueur-gages-quadit-pape-Francois-2018-10-10-1200975037>>.

Peut-on imposer à une mère d'accoucher d'un enfant qu'elle ne désire pas au nom du droit du fœtus? Peut-on permettre à une mère d'avorter jusqu'au dernier jour avant l'accouchement, alors que le fœtus peut être vivant et viable? Au Québec, l'avortement est légal, gratuit et peut se faire en tout temps, car il n'y a aucune loi qui fixe un délai.

En droit, la définition attributive est développée de façon cohérente avec la solution normative envisagée. La création d'une norme juridique est une solution pour régler un problème de vivre-ensemble. La définition est une opération fondamentale, mais qui est intrinsèque au développement de la solution juridique au problème vécu. C'est pourquoi le droit, au lieu de chercher à définir la substance de l'humain, attribue la définition de la personne selon la solution envisagée.

Le discours éthique ressemble à celui du droit, puisqu'il n'a pas recours à une définition substantive de la personne. Toutefois, il s'en distingue parce que ce ne sont pas les caractéristiques de la personne qui sont attribuées, mais la valeur de la « dignité humaine ». L'évaluation figure au cœur de l'approche éthique¹⁹. En effet, celle-ci ne se concentre pas sur la conformité d'une action avec une obligation, mais sur la prise de décision d'agir dans un contexte donné. C'est ce qui permet de comprendre le rôle du jugement de valeur dans le processus. Ainsi, dans le cas de l'avortement, il ne s'agit pas de savoir si l'avortement est permis ou non, mais plutôt de savoir comment se prend la décision de passer par un avortement pour résoudre un problème. La question de l'avortement se pose uniquement dans le cas d'une grossesse non désirée. La femme se pose la question suivante : que faire? avorter ou non? Pour prendre une décision délibérée et éclairée²⁰, elle analysera tous les impacts de sa décision sur le fœtus, elle-même, le père biologique, ses proches, etc. Autrement dit, toutes les personnes qui pourront subir l'impact positif ou négatif de sa décision. Une fois les impacts établis, elle procédera à la formulation des jugements de valeur pour chacun d'eux. C'est en considérant l'effet sur le fœtus qu'elle évaluera le degré de « dignité humaine » qu'elle lui attribue. Plus elle évaluera que le fœtus a de la dignité, plus elle sera sensible au développement utérin du fœtus.

La morale, le droit et l'éthique mobilisent donc différentes approches de la définition de la personne. La morale cherche à fonder ses obligations sur une définition substantive objective préalable à l'énoncé des obligations morales.

19. G.A. Legault, S.K. Bédard, C. Bellemare, J.-P. Béland, L. Bernier, P. Dagenais, C.-É. Daniel, H. Gagnon, M. Parent et J. Patenaude (2018), « Revisiting the Fact/Value Dichotomy: A Speech Act Approach to Improve the Integration of Ethics in Health Technology Assessment », *Open Journal of Philosophy*, vol. 8, n° 5.

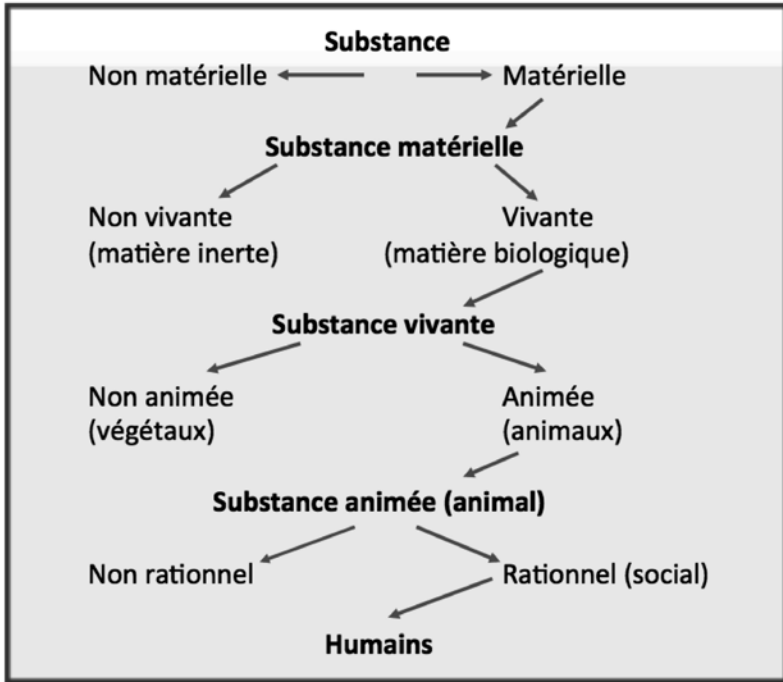
20. G.A. Legault (1999), *Professionnalisme et délibération éthique*, op. cit., p. 290.

Le droit attribue la personnalité juridique en tant qu'une partie intégrante de l'opérationnalisation de la norme qui vise à régler un problème de vivre-ensemble. L'éthique ne définit pas la personne, mais attribue la valeur de la dignité humaine à des parties prenantes qui subissent les effets de la décision. Peu importe l'approche substantive ou attributive de la définition de la personne, l'établissement de critères est l'étape nécessaire pour opérationnaliser la définition dans des contextes précis.

3.2 Le choix des critères et leur opérationnalisation

La classification que le philosophe Porphyre (234-310) a érigée à partir des traités de logique d'Aristote représente la matrice que les Grecs utilisaient pour définir les êtres dans l'univers. Cette classification nous permet de voir comment les critères sont développés lorsqu'on veut faire une définition. Pour eux, cette classification représente les choses telles qu'elles sont, c'est une définition substantive, mais le procédé de définir par critères est aussi valable pour la définition attributive. La définition a pour fonction de distinguer une chose d'une autre grâce à deux critères : le genre commun et la différence spécifique. Par exemple, la définition du mot *table* selon le dictionnaire Antidote est : « Meuble constitué d'une surface plane reposant sur un ou plusieurs pieds ou sur un support quelconque, où l'on prend les repas » et celle d'*armoire* : « Meuble de rangement, fermé par des battants et muni de tablettes. » La table comme l'armoire sont des meubles (genre commun) et la suite de la définition précise la différence spécifique. La définition fournit les caractéristiques dites « essentielles » de l'objet défini. La distinction d'Aristote entre « essence » (ce qui est la structure fondamentale de quelque chose) et l'« accident » (ce qui peut varier sans changer la structure) s'applique aux définitions de table et d'armoire, puisque leurs définitions ne tiennent pas compte de ce qui peut varier comme la grosseur, la forme, la couleur, le composite utilisé, etc.

Compte tenu de l'importance des définitions de la personne dans les approches substantives, une étude de la diversité des définitions permet d'illustrer la complexité du choix des critères pour résoudre des enjeux du vivre-ensemble. C'est pourquoi nous prenons comme point de départ cette classification : l'arbre de Porphyre, comme présenté dans le schéma suivant.



Pour les Grecs, tout être est substance, c'est-à-dire que l'être de la chose perdure dans le temps. Le genre commun à toutes choses est la substance. La classification va donc permettre de définir différentes substances. La substance non matérielle, c'est Dieu, alors que la substance matérielle renvoie à tout ce qui existe dans notre monde humain ; c'est le dénominateur commun à notre univers. La substance matérielle se divise en deux, la matière inerte (non vivante) et la matière vivante (matière biologique). La matière biologique se divise aussi en matière vivante non animée (végétaux) et en matière vivante animée (animaux). Dans la substance animée, on distingue la substance animée non rationnelle (animaux non rationnels) et la substance animée rationnelle, c'est la personne humaine. Puisque, dans la pensée d'Aristote, c'est le rationnel qui permet à l'humain d'être social en tant qu'être politique, on arrive donc à la définition de la personne : animal rationnel et social.

L'arbre de Porphyre représente une classification des êtres qui est au cœur de la culture occidentale. Cette classification n'est pas neutre, puisque la personne humaine est considérée comme la forme supérieure de la substance matérielle, ce qui l'autorise à utiliser comme elle l'entend toutes les substances non rationnelles.

La définition est générale par nature, puisqu'elle s'applique à tous les êtres diversifiés qui font partie de la catégorie. Or le passage du général au particulier

peut poser problème. La personne est un animal rationnel et social, mais si on veut opérationnaliser ce critère, on doit détenir d'autres critères plus précis permettant de distinguer ce qui est rationnel de ce qui ne l'est pas. Ainsi, pour appliquer le commandement «Tu ne tueras point», il faut déterminer quels êtres ne peuvent pas être tués. S'agit-il de tous les êtres vivants ou seulement des animaux rationnels et sociaux? Dans ce dernier cas, est-ce que ce commandement s'applique aux fœtus, aux personnes aliénées, etc.?

Au cours de l'histoire, les humains ont opérationnalisé les critères généraux de la personne dans différents contextes sociaux. L'esclavage a constitué une pratique qui a duré jusqu'au milieu du XIX^e siècle. Comment justifier cette pratique, si toute personne rationnelle et sociale fait partie de la communauté morale? Une première justification pour excommunier certaines personnes repose sur l'appartenance communautaire. Les Grecs ne sont pas les Romains, et si certains Romains sont vaincus, ils deviennent alors des trophées de guerre. Ainsi, ils sont considérés comme d'autres objets, pouvant être objet de commerce et d'exploitation. La découverte de peuples dits non civilisés dans les colonies a conduit à distinguer le statut de personne selon le degré de civilisation. La traite des esclaves, notamment des Noirs, a été centrale dans le développement économique de plusieurs pays. Pas étonnant que l'abolition de l'esclavage aux États-Unis ait été au cœur de la guerre de Sécession sous Abraham Lincoln. Au Canada, colonie britannique, l'abolition de l'esclavage date de 1833, à la suite de l'adoption de la loi anglaise. Enfin, pour l'Église catholique, *Sicut Dudum*²¹ est la première encyclique à condamner l'esclavage. Elle a été écrite par Eugène IV en 1454.

La pratique d'exposition d'enfants en Grèce²² et dans d'autres cultures consiste à prendre une décision de ne pas élever un enfant. Le tout repose sur l'autorité du chef de la maison, mais aussi d'une autorité politique. Ainsi, on pouvait soit ne pas élever l'enfant en l'excluant de la maison, soit par le geste d'exposition. L'exposition consiste à déposer un enfant dans un panier et à l'exposer dans un endroit passant. L'enfant était rarement recueilli par une famille et, lorsqu'il l'était, c'était pour la traite d'êtres humains. Les raisons d'exposer sont multiples, dont l'infirmité, le sexe de l'enfant et les contraintes économiques.

Ces deux exemples démontrent que l'opérationnalisation de la définition de la personne pour établir des balises morales ou légales dépend toujours du contexte social et culturel. L'appel à la morale ou au droit pour interdire

21. <https://fr.wikipedia.org/wiki/Sicut_dudum>.

22. P. Brulé (2009), «L'exposition des enfants en Grèce antique: une forme d'infanticide», *Enfances et Psy*, 3 (n° 44) p. 19-28, <<https://www.cairn.info/revue-enfances-et-psy-2009-3-page-19.htm#no5>>.

l'esclavage ne peut aboutir qu'avec la transformation des pratiques sociales. Interdire l'esclavage ou l'exposition d'enfants a un effet direct sur la vie économique des producteurs, dans le premier cas, et des familles dans le second. Abolir l'esclavage et l'exposition d'enfants oblige à considérer les esclaves et les enfants comme faisant partie de la famille humaine, et à modifier les pratiques sociales en conséquence.

Le philosophe moraliste Peter Singer a soulevé la controverse dans la communauté philosophique en proposant que nos obligations morales ne soient pas réservées exclusivement aux humains, mais aussi aux animaux. Pour élargir la communauté morale aux animaux, Singer prend donc comme point de départ la notion d'animal. Dans une entrevue avec Johann Hari, il précise sa pensée, comme le rapporte le journaliste :

“You shouldn't say animals,” he says in a level tone when I raise the topic, “to distinguish between humans and non-humans. We are all animals.” This objection captures Singer's thoughts in a neat sound bite. He thinks there is nothing special about being human. “Every living thing has preferences, and those preferences need to be taken into account,” he says. “Non-human animals can't be left out of utilitarian equation²³.”

Pour opérationnaliser la morale étendue à tous les animaux, le philosophe ne propose pas une égalité ontologique, c'est-à-dire basée uniquement sur le fait que l'on est un animal humain ou non humain. Il propose une égalité qui prend en considération la différence des rapports entre les souffrances et les préférences humaines et celles des animaux. Ainsi affirme-t-il :

Look: pain and suffering are bad and should be prevented or minimised, regardless of the race, sex or species of the being that suffers. It's a simple fact that a three-year-old human has pretty much the same self-awareness, rationality and capacity to feel pain as an adult ape. So they should be given equal moral consideration²⁴.

Ainsi donc, nous devrions traiter un singe avec la même dignité qu'un enfant de trois ans. Ceci permet d'inclure le singe dans la communauté morale. Cependant, on doit aussi accorder la même dignité au singe et à un enfant de trois ans. C'est ce qui permet au philosophe de soutenir :

Gunning down a load of normal people in a supermarket is worse than shooting the same number of cows in a field – but that's not because of their species. It's because normal humans have a capacity to want to carry on living – and relatives who will grieve for them – to a greater degree than cows. If you were massacring

23. J. Hari (juillet 2004), « Some people are more equal than others », *The Independent*, <<https://www.independent.co.uk/news/people/profiles/peter-singer-some-people-are-more-equal-than-others-551696.html>>.

24. *Ibid.*

severely brain-damaged humans who had the same understanding of the world as a herd of cows, both would still be very wrong, but it would be equally wrong. It's pain and grief and denial of a living creature's will to carry on that count – not species²⁵.

La suite logique de son argument moral le conduit à soutenir l'acceptabilité morale de l'infanticide: « He believes that severely disabled or defective animals who cannot live except in terrible pain can legitimately be killed. Why, he asks implicitly, should this not be the case for human babies²⁶? »

L'approche morale de Peter Singer, considéré comme le père du mouvement de la reconnaissance des droits des animaux, propose de considérer l'animal comme un être capable de souffrir et d'avoir des préférences; c'est la catégorie centrale de la morale qui s'applique à tous les animaux. Ainsi, pour opérationnaliser cette morale, une gradation doit être faite non plus entre humain/non-humain, le critère retenu est le degré de conscience permettant de mesurer les préférences et la souffrance vécue. La proposition de Singer a été et est encore très contestée. Cela s'explique aisément, car accepter son approche morale impliquerait de modifier beaucoup de nos pratiques humaines envers les animaux (élevages, consommation, etc.) et les personnes humaines ayant des handicaps à la naissance ou encore supportant des souffrances interminables.

CONCLUSION

Définir la personne n'est pas neutre, car cette catégorisation est la base de tous les discours normatifs : la morale, le droit et l'éthique. Chaque discours mobilise un type de définition : la définition substantive en morale et la définition attributive en droit et en éthique. Chaque définition se développe en distinguant le genre commun et la différence spécifique auquel appartient ce qui est défini. Ce qui les distingue essentiellement, c'est que la définition substantive cherche à établir la véritable définition de la personne, alors que la définition attributive donne les caractéristiques de la personne qui sont cohérentes avec la solution normative proposée pour résoudre le problème du vivre-ensemble. C'est à l'étape de l'opérationnalisation de la définition dans le contexte social et culturel que se manifestent les débats autour de l'acceptabilité ou non de telle ou telle approche morale, juridique ou éthique.

L'impossibilité de fonder une définition substantive sur une connaissance universelle de la personne nous oblige à adopter une approche attributive fondée sur la cohérence entre la définition, la solution normative et le problème

25. *Ibid.*

26. *Ibid.*

de vivre-ensemble que posent les transformations des pratiques. Dans une telle approche, le point de départ est le diagnostic qui détermine le problème de vivre-ensemble. La robotisation de l'humain et l'humanisation du robot que promet le développement technologique ont plusieurs effets sur le vivre-ensemble. Le diagnostic s'établit à partir des pratiques actuelles et la projection de leurs développements. Deux développements majeurs se pointent à l'horizon et sont en rupture complète avec notre expérience de la personne humaine. Le premier est le clonage de l'humain. On a déjà cloné des macaques, alors si la tendance se maintient, ce sera le tour de l'homme. Quels problèmes le clonage pose-t-il dans le vivre-ensemble? Évidemment, celui de savoir si le clone appartient à la personne qui l'a fait cloner à partir de son code génétique ou non. Autrement dit, le clone est-il une personne humaine au même titre que les autres? Si le clone n'est pas considéré comme une personne à part entière, mais une fabrication, alors il pourrait servir de banque d'organes compatibles pour son propriétaire. Par contre, s'il est reconnu comme une personne humaine, qui en aura la responsabilité légale? C'est parce que le clone ouvre une situation inédite qui brise nos catégories «naturelles» de penser l'humain que la détermination de son statut doit être précisée. On compare souvent le clonage à des jumeaux monozygotes, c'est-à-dire qu'ils proviennent de la division de l'œuf initial et non de deux œufs différents. Toutefois, contrairement aux jumeaux monozygotes, le clone n'est pas le résultat d'une fécondation d'ovule par le sperme, mais plutôt de la manipulation génétique permettant la reproduction du même. Comme le souligne Mark Hunyadi :

Si nous sommes tous effectivement dépendants d'un patrimoine génétique que nous n'avons pas choisi, le clone se trouve cependant dans cette situation inédite que *quelqu'un d'autre l'a choisi pour lui*. Si, comme nous venons de le voir, cela n'entame pas la capacité formelle d'exercer son autonomie, cela le met néanmoins dans cette position ontologique inédite d'être en tant qu'être humain biologique l'expression du désir *déterminé* de quelqu'un d'autre. Le point saillant est ici précisément que ce désir soit *déterminé*: ce n'est pas alors *un* enfant que l'on attend, mais *tel* enfant, un enfant que l'on veut voir reproduit au maximum de similarité de quelqu'un d'autre – le plus souvent, sans doute, de soi-même²⁷.

Les questions éthiques à résoudre pour assurer le vivre-ensemble avec les clones sont d'abord celles de savoir si nous voulons intégrer dans nos modes de vie des banques d'organes vivantes pour favoriser les transplantations d'organes comme assurance pour le bien-être futur, ou bien les considérer comme des personnes humaines et alors redéfinir la parentalité.

27. M. Hunyadi (2004), «Je est un clone». *Laval théologique et philosophique*, vol. 60, p. 16.

L'autre technologie de rupture est celle de l'intelligence artificielle intégrée dans des robots dits « humanoïdes ». L'évolution de l'intelligence artificielle permet non seulement aux robots de ressembler physiquement à l'être humain, mais aussi de réaliser plusieurs activités semblables à l'être humain, dont la sensibilité par les capteurs, la capacité à se déplacer, la capacité à aller chercher de l'énergie lorsqu'il est en manque, l'expression d'émotions par mimiques faciales, la capacité à apprendre comment corriger son comportement après des essais et des erreurs, la capacité à parler et à entrer en conversation avec les humains. La création de tels robots humanoïdes est un terrain de jeux pour plusieurs ingénieurs qui doivent relever des défis techniques sans précédent. Or tant que le robot était confiné dans un espace clos comme les robots industriels, il y avait peu d'interactions entre les humains et les robots. Se pointe à l'horizon la présence de robots parmi les humains pour accomplir différentes tâches collaboratives dans les usines, les maisons, les centres de santé. En modifiant nos pratiques sociales, notamment en remplaçant des humains dans les services aux personnes, plusieurs effets se feront sentir sur la main-d'œuvre, sur les relations psychologiques entre soignant et soigné, sur la qualité de soins, etc. Le problème de la responsabilité légale de la voiture autonome se pose déjà dans les débats juridiques, et la question va se compliquer avec la diversité des robots interactifs envisagés. Advenant des dommages aux personnes, qui devra en assumer les coûts ? De plus, compte tenu de la complexité du robot, il y aura beaucoup de débats sur la preuve de la responsabilité. La transformation des pratiques de soins avec des robots aura des effets sur les organisations de santé, sur les pratiques professionnelles, sur les soignés. Un jour ou l'autre, l'attribution de la personnalité juridique aux robots se posera concrètement comme solution envisageable au problème de vivre-ensemble. Cependant, la rupture fondamentale apparaîtra le jour où le robot atteindra un certain stade d'autonomie dans ses décisions. Un jour se posera la question de savoir si nous allons traiter les robots qui semblent posséder les mêmes capacités que l'être biologique animal ou humain comme partie de notre communauté morale, juridique ou éthique.

Le développement technologique, depuis 70 ans, a provoqué des transformations dans notre conception de la personne, caractérisée par notre expérience biologique de la naissance jusqu'à la mort. Ces transformations touchent toutes les étapes du cycle de notre vie et nous les avons acceptées parce qu'elles nous apportaient plus d'autonomie et de mieux-être. Mais le développement technologique comporte différents impacts sociaux et culturels qui modifient nos façons de faire et de vivre-ensemble qui ne sont pas nécessairement pris en considération lors de leur acceptation sociale. À chaque étape du développement technologique, nous devons choisir les moyens d'intégrer ces technologies dans nos façons de vivre. À chacune de ces décisions, on doit

décider du degré de « dignité » que nous accordons aux personnes, aux animaux, aux robots, aux institutions et à l'environnement qui nous entourent. Notre responsabilité éthique n'est pas d'agir conformément à notre nature comme le proposaient les anciens, mais de décider quel humain nous voulons être avec les autres dans nos institutions.

CHAPITRE 2

La subjectivité à l'épreuve de la technologie

Célia Vaz-Cerniglia et Margherita Merucci

INTRODUCTION

En 1947, Georges Bernanos publia *La France contre les robots*. Il mettait en garde ses contemporains contre certaines dérives de la société moderne, présentée comme la civilisation des machines. Cette civilisation est accusée d'être « une conspiration universelle contre toutes espèces de vie intérieure » (Guillebaud, 2018¹). En particulier, il affirme que : « Le danger n'est pas dans la multiplication des machines, mais dans le nombre sans cesse croissant d'hommes habitués, dès leur enfance, à ne désirer que ce que les machines peuvent donner. » (Bernanos, 1947, p. 172)

Soixante-dix ans plus tard, nous sommes réunis pour réfléchir aux transformations dont la personne humaine sera l'objet. Les questions qui émergent se centrent sur la recherche des attributs irréductibles permettant de préserver la singularité de l'être humain. Notamment, savoir si la personne humaine considérée comme un être de désir peut continuer à être libre de désirer dans un contexte sociétal qui accorde de plus en plus de place aux machines. Le risque de croire que les besoins seront inévitablement inculqués par des algorithmes, des interfaces homme-machine, est présent et exacerbé par des conceptions critiques de certains psychiatres, psychanalystes tels que Roland Gori, au risque de paraître réactionnaire.

1. J.C. Guillebaud (2018), « La France contre les robots », *La Vie*, p. 6.

Ainsi, depuis quelques années, la technologie qui passe par l'intelligence artificielle se développe pour faciliter la vie quotidienne du sujet en palliant ses déficiences, ses incapacités, en souhaitant améliorer son niveau d'autonomie ou encore augmenter ses performances.

Quand il s'agit de créer un environnement favorable au maintien ou à la restauration d'un niveau d'autonomie chez le sujet, ce dernier est amené à retrouver son équilibre psychique, cognitif et physique interne. Dès lors, il nous apparaît surprenant que l'on puisse évoquer l'expression d'homme augmenté. Cependant, nous nous interrogeons sur la demande qui est faite quant au développement de la technologie pour augmenter les performances humaines. En effet, d'où émane-t-elle, à qui s'adresse-t-elle et pour quelle finalité? En d'autres termes, nous mettons en question l'origine du désir de ce développement technologique. S'agit-il d'un désir issu du sujet, provenant de la famille, émergeant de la société, du pouvoir économique? À qui s'adresse ce progrès technologique? Conduit-il véritablement à l'autonomie du sujet, à la construction de son identité? Si oui, avec quelles contraintes?

Nous rappelons que programmer un ou des algorithmes, c'est déjà délimiter, restreindre les imprévus même si ces algorithmes peuvent se reprogrammer en interagissant avec leur milieu par un processus d'accommodation et d'assimilation pour atteindre un certain équilibre fonctionnel, semble-t-il.

Cependant, il s'agit d'algorithmes sans incarnation vivante, dans l'incapacité d'éprouver quelques affects, qui peuvent peut-être se montrer empathiques mais limités à un certain niveau rudimentaire (Tisseron, 2011²).

En outre, pour que les algorithmes d'intelligence artificielle des machines parviennent à avoir une identité humaine, il serait nécessaire qu'ils acquièrent une subjectivité et une singularité afin d'être reconnus et identifiés comme sujets. Or, nous savons que pour être sujet parmi d'autres, un processus d'affiliation psychique à un groupe d'appartenance doit se mettre en place. Pour être sujet, nous sommes amenés à vivre l'absence, le manque, la souffrance psychique, et nous avons crucialement besoin de l'autre, différent et pourtant semblable pour être reconnu comme sujet singulier.

De plus, entre sujets, nous pouvons partager une expérience à la fois commune et subjective dans une véritable rencontre propice à une écoute plus soucieuse du fonctionnement psychique de chacun.

La rencontre juste, véritable suppose donc de vivre ce moment partagé dans une authentique réciprocité inter-psychique. Cette notion de réciprocité implique aussi que des sentiments similaires soient éveillés de part et d'autre.

2. S. Tisseron et H.-P. Bass (2011), « L'empathie, au cœur du jeu social ». *Le journal des psychologues*, 3, n° 286, p. 20-23. DOI 10.3917/jdp.286.0020.

Or, jusqu'à preuve du contraire, même un androïde ne dispose pas d'une psyché. Certains chercheurs scientifiques, néanmoins, tels que Raymond Kurzweil, dans *The Singularity is Near* (2005), attendent une technologie suffisamment avancée pour rendre les intelligences artificielles véritablement intelligentes, en considérant les humains comme naturellement bénéficiaires, en devenant surhumains et en dépassant l'ère humaine. Leur motivation réside dans le contrôle total de la faim, de la vieillesse, de la mort, tout ce qui appartient à l'être vivant, en somme.

Nous soulignerons que ce n'est que dans le cadre de l'authenticité de la rencontre que nous pouvons éprouver ce qui se joue entre soi et l'autre et que nous pouvons jauger du degré de notre implication émotionnelle sans chercher à connaître parfaitement l'autre. Un espace intime propre à chaque sujet est de mise, espace qu'il convient de respecter.

D'ailleurs, il arrive que le sujet soit craintif, qu'il ait peur, qu'il ait honte dans ce que l'autre va pouvoir décoder, percevoir de ce qu'il a bien voulu donner à voir ou qu'il a intentionnellement ou non caché. En prenant soin de mettre en question le jeu entre le transfert et le contre-transfert entre les deux protagonistes, une relation de confiance se met en place en se construisant selon un parcours complexe empreint de résistance, de contre-résistance, d'alliances, de mésalliances...

Ainsi, qu'est-ce qui motive la fascination pour les robots, les machines, les nouvelles technologies? Est-ce un besoin de trouver une solution à toutes nos incomplétudes, comme le croient fortement les singularistes comme Kurzweil? Est-ce un besoin de ne pas ressentir notre fragilité, caractéristique qui nous constitue en tant que sujets humains, et notre difficulté à nous dessaisir de notre fantasme d'immortalité, à rechercher une satisfaction immédiate, à ne pas souffrir de notre solitude?

Or, l'humain naît dans l'incomplétude. Comme il a déjà été avancé, il a besoin d'étagage pour se construire et il a besoin de l'autre pour construire son identité singulière (Ricœur, 1990). Ceci nous amène encore à mettre en doute la capacité d'un robot ou encore d'un hologramme à matérialiser cet autre.

Pour revenir à Roland Gori³, cet auteur nous met on garde en précisant que «c'est sur la scène des langages plus ou moins infectés par la "religion du marché" que se développent aujourd'hui les luttes pour l'hégémonie culturelle, instaurant de nouvelles hiérarchies de valeurs et de nouvelles manières de civiliser les individus et de fabriquer des subjectivités». Il évoque la religion du marché qui aurait besoin d'être privée de toute subjectivité, car désirant seulement ce qu'elle veut proposer. Comme il l'avance dans son ouvrage,

3. R. Gori (2013), *La dignité de penser*, Babel éditions: Lonrai, p. 11.

il convient de s'interroger sur notre société qui perd progressivement son modèle démocratique classique en se muant progressivement en démocratie totalitaire, qui tue le désir et la liberté à leur naissance même.

Ainsi, nous poursuivons notre questionnement sur ce qui fait la subjectivité d'une personne. En effet, de quelle manière chaque être humain se construit-il comme être unique et singulier ?

Bien que ce sujet soit vaste et controversé, il s'avère intéressant, car il permet de réfléchir au devenir de notre société. Nous nous référerons, dans un premier temps, à la psychologie du développement et à la psychologie clinique afin d'aborder la question de la subjectivité humaine. Nous nous appuierons également sur les apports de la psychologie existentielle et de la théorie des émotions pour cerner le rôle de l'empathie dans la construction des relations humaines. Mais cela signifie-t-il que nous devons nous opposer de façon stérile à la technologie et à ce qu'elle peut apporter ? Au contraire, dans un second temps, nous illustrerons comment les apports de la technologie, qui permettent de dépasser les limites imposées par certains handicaps, peuvent aider des personnes handicapées à acquérir de nouvelles capacités et à devenir des sujets acteurs de leur vie.

1. LA SUBJECTIVITÉ EN PSYCHOLOGIE

Depuis leur origine, la psychologie du développement et la psychologie clinique se sont penchées sur cette question du développement humain dans leur tentative pour définir ce qui fonde la subjectivité et la singularité de la personne dans ses forces et dans ses fragilités.

La psychologie du développement, quel que soit son angle d'approche et dans une perspective fonctionnaliste, met en évidence le processus de transformations successives dans lequel le sujet humain est engagé pendant toute sa vie afin de s'assumer et de s'adapter aux différents défis que la vie et l'environnement lui proposent et auxquels il doit faire face. La lente transformation de l'immaturation à la maturité, de l'indifférenciation à l'acquisition de sa propre identité, de la dépendance absolue à l'autonomie, de la pulsionnalité à la socialisation, révèle que ce processus, comme tout processus, est toujours en renouvellement et accomplissement. Ce processus d'advenir dans le temps dépend de la rencontre de chacun avec son environnement et avec les contraintes, les limites et les libertés que ce dernier lui octroie. De plus, tout ce travail de subjectivation est impensable en dehors du lien avec soi-même et

avec l'autre. D'après Irvin Yalom⁴, ce qui fait notre humanité, c'est la possibilité de développer un « je » en relation avec un « nous », dans une optique systémique de co-construction et d'interdépendance réciproques. Toute vie humaine « s'articule autour de forces, de motivations et des peurs, conscientes et inconscientes », ainsi qu'autour de relations réelles et/ou fantasmées.

La psychologie clinique, de son côté, et ce, quel que soit son angle d'approche, nous alerte sur la souffrance et sur ses effets sur la psyché humaine. Peu importe l'angle de lecture choisi et la théorie de référence, comme il a été dit précédemment, c'est toujours la vulnérabilité du sujet en conflit avec soi et avec l'autre qui est à l'origine du mal-être. Elle peut être à l'origine de la pathologie lorsque le processus développemental partant de la symbiose, de la mutualité entre le très jeune enfant et sa mère, n'a pu contribuer à établir le narcissisme primaire chez cet enfant. Nous sommes là en deçà des pulsions si chères à S. Freud. Ce processus, s'il se développe harmonieusement dans une rencontre empathique, peut favoriser la séparation et la différenciation propices à l'installation d'un narcissisme secondaire, suffisamment fort pour supporter la solitude, la capacité à être seul, concept si cher à D. Winnicott. Le conflit et la confrontation avec sa propre survie tout d'abord, sa propre vulnérabilité à certains moments de sa vie figurent donc au cœur du développement de chacun. En effet, les confrontations avec ses propres angoisses, ses besoins et ses pulsions, ainsi qu'avec son entourage familial d'abord et social ensuite, sont le moteur de son développement et aussi sa pierre d'achoppement. Tout le travail du sujet consiste à pouvoir dépasser les conflits internes et externes vécus dans la relation à l'autre pour les transformer. En outre, aux besoins intrinsèques à l'individu comme l'édification de son unité de base grâce à la préoccupation maternelle, à son vécu chaotique ou harmonieux et à sa participation au monde plus ou moins performante, s'ajoutent d'autres besoins fondamentaux qui définissent son existence et qu'Irvin Yalom (2008) définit comme des « enjeux ultimes ». Nous partageons la vision de cet auteur lorsqu'il affirme : « Parmi ces situations *limites* ou *frontière*, figurent certaines expériences comme la confrontation à sa propre mort, à sa liberté, à l'isolement ou au besoin de donner du sens à sa propre vie » (*ibid.*, p. 16). Ces quatre situations constituent les enjeux fondamentaux du processus de subjectivation de l'être humain.

L'être humain a des besoins certes, mais est-ce que nous n'avons pas tendance à penser pour les autres, à créer des besoins qui ne sont pas au service de la vie et du sujet ? Autrement dit, est-ce que nous prenons le temps de nous interroger sur ce qui sert au déploiement des fondamentaux du sujet à savoir

4. I. Yalom (2008), *Thérapie existentielle*, Paris : Galaade Édition (1^{re} édition : *Existential Psychotherapy*, Basic Books, 1980), p. 13.

la mortalité, la liberté, la créativité, la recherche du lien avec l'autre, la recherche de sens? Ces fondamentaux seront abordés un peu plus loin.

Est-ce que nous n'avons pas tendance à répondre à des finalités de productivité, de consommation, de création d'emplois, de performance, pour repousser plus loin encore cette sensation d'incomplétude, comme si nous avions besoin de régresser plutôt que de progresser?

Accéder au mieux-être passe-t-il donc nécessairement par la création de nouveaux remèdes, de nouvelles technologies? Si oui, jusqu'à quelle mesure? Quelles en seraient les limites? Quelles en seraient les frontières? Comment respecter la vulnérabilité et l'identité subjective si elles ne peuvent s'exprimer?

L'identité subjective étant un sentiment qui renvoie à l'acte de désirer, ce qui renvoie à notre propre moi, d'éprouver nos fantasmes, de pouvoir penser librement, ce qui n'exclut pas d'être aussi traversé par des angoisses de mort, de persécution, de séparation, de perte, etc.

Nous partons du postulat que l'homme aspire à améliorer sa condition de vie, à accéder à un mieux-être en recherchant une complétude à l'instar du jeune enfant qui a besoin de sa mère pour vivre une symbiose totale. Il tend ainsi à oublier le sens de la vie avec les autres ainsi que sa finitude, car il souhaite vivre dans une fusion totale. Son appartenance à l'espèce vivante est négligée au profit de la performance individuelle, confondant un véritable travail de groupe qui respecte la subjectivité et l'intimité de chacun d'un travail avec une juxtaposition d'individualités.

Nous entendons ici par travail le travail psychique, la transmission, la construction de la qualité du lien psychique social, lien qui permet de maintenir l'autonomie du sujet. Nous faisons l'hypothèse que l'homme clive son monde psychique afin de se réaliser dans des projets et oublier un temps que son temps est compté, altérant de ce fait sa subjectivité. Temps et espace sont deux dimensions indispensables qui favorisent notre inscription dans notre propre histoire et celle de l'humanité.

Nous pouvons considérer que développer les technologies robotiques, inclure les intelligences artificielles dans notre environnement quotidien facilite un espace de création dont la finalité servirait peut-être à oublier notre mortalité. Cependant, face à un accident, à un événement traumatique survenant chez un sujet ou un groupe de sujets, nous estimons que la capacité de créativité est altérée et que les nouvelles technologies viennent temporairement combler ce qui fait trou, trauma, ce qui manque, ce qui est absent chez le sujet.

Faudra-t-il pour autant en conclure que les fondamentaux, tels que la confrontation à la mort, la souffrance psychique, la recherche et le maintien du lien inter et intra subjectif et le sens de notre existence, soient de plus en

plus négligés, différés, voire oubliés dans ce contexte de développement des nouvelles technologies robotiques sans laisser un espace propice à la construction de sa propre subjectivité et créativité?

Il nous importe ici de poursuivre notre réflexion sur ces fondamentaux de la subjectivité comme des enjeux du développement de la personne dans un tel contexte. Nous pourrions ensuite évoquer des situations particulières dans lesquelles certains apports technologiques sont au service du sujet humain.

La mort est probablement l'enjeu le plus « facile » à définir, car il constitue une évidence : pour tout homme, il n'y a aucune possibilité d'échapper à son destin ultime. La mort reste une expérience inconnue et inconnaissable. Associée chez l'enfant à un départ, à une absence, elle est souvent refoulée dès que le sujet, en grandissant, en comprend l'irréversibilité. La question de la recherche d'un sens à une vie si brève y est strictement liée.

L'isolement⁵ renvoie chacun à ses propres éprouvés. Chaque être humain éprouve, avec plus ou moins d'angoisse, le sentiment de se sentir à part, hors place au moins une fois dans sa vie. Le fossé qui le sépare des autres, ou de soi-même, avec le sentiment de ne pas se reconnaître, s'accompagne d'un sens de solitude difficile à vivre. Certains essayent d'oublier la perte du sentiment de familiarité avec le monde en se maintenant hyperconnectés dans un monde de connexions et d'objets. Toutefois, d'après Irvin Yalom (Yalom, p. 498), il est inévitable de se confronter à la solitude pour nous engager pleinement dans le lien à autrui.

De façon paradoxale, la confrontation avec la liberté est aussi source d'angoisse. Toutefois, cette confrontation constitue un enjeu fondamental du développement. Ce dernier concept a toujours eu une valeur positive, et la liberté a toujours été considérée comme le moteur de l'évolution et de l'avancée des hommes et des sociétés. Cependant, il n'y a pas de liberté sans restriction, et, comme le disait Freud, tout enfant a besoin pour grandir d'un mur sur lequel faire rebondir son ballon. Il ne s'agit donc pas de vivre dans un monde de toute-puissance et d'absence de règles, mais de se construire en lien avec autrui et d'équilibrer ses propres désirs, son vouloir et sa responsabilité avec les limites imposées par la réalité. Nous faisons bien sûr référence aux trois

5. Yalom reconnaît trois types d'isolement. L'isolement « interpersonnel » (c'est le sentiment de solitude éprouvé par le sujet face aux autres), l'isolement « intrapersonnel » (le sujet se coupe d'une partie de lui-même et on assiste à une fragmentation du *self*[soi]). Pour finir, il évoque l'« isolement existentiel », qui « renvoie à un abîme infranchissable entre soi et un autre. Il renvoie à un isolement encore plus fondamental, la séparation entre le sujet et le monde. [...] l'isolement existentiel est une vallée de solitude dont il est possible de s'approcher de maintes façons. » (*ibid.*, p. 486-487) « Une confrontation à la mort et à la liberté conduit inévitablement l'être humain dans cette vallée. » (*ibid.*, p. 489)

instances psychiques freudiennes, le moi régulant la force pulsionnelle pour s'adapter aux contingences et aux exigences extérieures. Cela comporte d'assumer ses propres agissements et de revendiquer la paternité de ses propres actions et de leurs conséquences :

Être conscient de sa responsabilité, c'est être conscient d'être le créateur de son self, de son destin, de ses situations de vie difficiles, de ses émotions et le cas échéant de sa souffrance. L'être humain porte donc la responsabilité de ce qu'il fait et de ce qu'il a décidé d'ignorer⁶.

Désir, volonté, liberté sont des enjeux qui s'entrecroisent et se façonnent réciproquement. Grâce et à travers la confrontation avec chacun d'eux, chaque être humain façonne sa propre singularité et l'inscrit dans le temps. Selon Rollo May, le désir précède la volonté et nous relie au futur :

Le désir représente une reconnaissance de la forme que nous voulons donner à l'avenir. C'est une aptitude à accéder profondément à l'intérieur de nous-mêmes et à prendre en compte une aspiration à changer l'avenir. De plus le « désir » apporte à la « volonté » la chaleur, le contentement, l'imagination, le jeu d'enfant, la vivacité, la richesse. La « volonté » donne au « désir » l'autodirection, la maturité. La « volonté » protège le désir, lui permet de continuer sans courir des risques trop grands. Mais sans le « désir », la volonté perd son sang vital, sa viabilité, et tend à expirer dans la contradiction interne. Si vous n'avez que de la « volonté » sans « désir », vous avez le néo-puritan victorien, desséché. Si vous n'avez que le « désir » et point de « volonté », vous avez la personne compulsive, sans liberté, infantile, qui, en tant qu'adulte-demeuré-en-bas-âge, risque de devenir l'homme-robot⁷.

La façon de grandir et d'évoluer par rapport à ces enjeux construit la subjectivité et la singularité de chacun, car l'expérience qui est vécue construit la dignité humaine, celle qui provient « de la pensée, de la capacité de penser. Cette capacité de penser est subordonnée à la parole, à la parole sans laquelle il n'y a davantage de singularité que de démocratie. » (Gori, 2013⁸)

Avoir peur et ne pas se confronter à ces enjeux, en se noyant dans un monde virtuel notamment, peut amener l'être humain à ne désirer que ce que la machine désire pour lui et de se sentir satisfait de ce qu'elle lui propose. L'être humain peut ainsi perdre sa capacité de penser, de se penser dans le monde, et en définitive perdre sa liberté.

Jusque-là, en reprenant les apports de la psychologie existentielle avec Irvin Yalom surtout, mais aussi psychodynamique avec d'autres auteurs comme Sigmund Freud, Donald Winnicott, nous avons tenté de décrire comment

6. *Ibid.*, p. 303.

7. *Ibid.*, p. 414.

8. R. Gori (2013), *La dignité de penser, op. cit.*, p. 12.

l'individu confronté à lui-même construit sa propre subjectivité pour devenir sujet à part entière. Nous rappelons, comme le préconisait Donald Winnicott avec sa phrase célèbre, qu'un bébé sans sa mère, ça n'existe pas, que l'être humain est aussi un être social qui se construit en relation avec les autres. Chacun de nous est façonné par les liens qu'il tisse consciemment et inconsciemment au cours de sa vie sans véritablement comprendre comment notre désir d'autrui et nos désirs en général nous mettent en contact avec nos semblables.

Pour désirer, il faut pouvoir ressentir ses émotions et ses affects avec force, et comme le dit le neurologue Antonio Damasio, « la conscience et l'émotion ne sont pas séparables⁹ ».

Henri Wallon (1941), psychologue développemental de son côté, affirme que le social se construit à partir des émotions, ces « systèmes d'attitudes qui, pour chacune, répondent à une certaine espèce de situation¹⁰ ». Il s'agit là de la forme la plus archaïque de communication. Cet auteur précise que c'est souvent l'émotion qui donne le ton au réel. Il ajoute que le réel peut certainement acquérir le pouvoir de déclencher les émotions. Les émotions sont à la base de l'éprouvé empathique, qui seul permet de les reconnaître dans ce que l'autre nous montre à voir. Jacqueline Nadel en donne la définition suivante :

Par essence contagieuse, l'émotion établit directement des liens intuitifs, sur un système d'accord entre celui qui ressent l'émotion, la donne à voir, et celui qui la voit et la réfère à ses propres états. Pour communiquer il faut être capable de la double opération qui consiste à décoder de façon stable chez les autres des signaux qu'on émet soi-même. (Camélio, 2006¹¹)

Ressentir nos émotions nous met en communication et en lien avec l'autre et ressentir le désir de l'autre ainsi que le nôtre nous confronte à l'acte de décider et de choisir. Nous sommes ici au cœur de notre problématique au sujet du développement de la subjectivité dans le contexte du développement des robots. L'homme risque d'oublier le sens de ce qui définit la vie avec les autres et la participation à un projet qui est collectif (la valeur de la « polis ») et la représentation collective d'une vision d'humanité. Comme dit Georges Bernanos (1947¹²) : « Le danger n'est pas dans la multiplication des machines, mais dans

-
9. M.C. Camélio (2006), « L'enfant polyhandicapé et les aléas de la communication empathique et émotionnelle », recherches cliniques et sociologiques sur le couple et la famille, *Dialogue* 174 (4), p. 29.
 10. H. Wallon (1941), *L'évolution psychologique de l'enfant*. Paris: Armand Colin, rééd. 1968, p. 114.
 11. M.C. Camélio (2006), *op. cit.*, p. 29.
 12. G. Bernanos (1947), *La France contre les robots*, p. 172.

le nombre sans cesse croissant d'hommes habitués, dès leur enfance, à ne désirer que ce que les machines peuvent donner. »

2. ILLUSTRATIONS

Nous constatons, depuis des décennies dans le domaine du handicap, que la qualité de vie des personnes handicapées a énormément évolué de concert avec l'évolution de la technologie. Notamment, l'enfant tétraplégique peut aujourd'hui jouer du piano grâce à un lecteur de mouvements oculaires qui, connecté à un ordinateur, permet d'actionner les touches du piano. Cette possibilité nouvelle ouvre des perspectives d'expression de soi en agissant sur son environnement et en renforçant sa position d'acteur.

Si le bébé aveugle, grâce à un casque à ultrasons, peut « voir » les obstacles dans une pièce et donc les éviter avec assurance (même principe que le fonctionnement radar des chauves-souris), il commencera à marcher avec quelques mois d'avance par rapport à d'autres enfants aveugles non bénéficiaires du même outil, outil qu'il abandonnera assez rapidement dès qu'il aura construit d'autres modalités fonctionnelles pour se repérer dans l'espace grâce au processus neurologique de vicariance.

Il existe aussi des logiciels de communication qui permettent de « parler » pour les personnes qui ne peuvent utiliser la parole à cause de pathologies motrices (IMC, troubles dysarthrique). Ainsi, une image ou un symbole devient parole permettant à l'enfant de sortir de son silence forcé.

Bien sûr, les innovations vont bien plus loin. Cependant, ce qui est commun à tous ces outils technologiques, ce sont les interrogations qu'ils suscitent. En effet, sont-ils des outils qui permettent de dépasser des manques, impliquant que, pour les dépasser, il est inévitable de les prendre en compte ? Ces sujets handicapés ont-ils pris le temps de les accepter et de les intégrer dans leur identité ?

Certes, l'outil devient une compensation qui facilite la vie. Le sujet en est dépendant pour s'exprimer en partie, mais cet outil n'est pas intégré à l'identité même. Il faudra alors travailler la dépendance à cet objet non humain, produit de l'humain, et qui aide à mieux vivre au quotidien¹³.

La machine permet ainsi de déplacer les limites en les poussant toujours plus loin sans vraiment les penser, sans peser le pour et le contre, sans entrevoir

13. Nous pensons à des situations plus proprement médicales comme l'implantation des valves, par exemple, où quelque chose d'externe et de mécanique est inséré dans le corps de l'individu. Ce dernier peut oublier leur existence, mais elles permettent au corps de fonctionner plus ou moins normalement.

et anticiper les modifications du comportement humain. Nous revenons sur notre interrogation relative à la fascination pour les robots : les machines seraient-elles liées au besoin de trouver une solution à toutes nos incomplétudes, en oubliant que l'humain naît dans l'incomplétude ? Nous insistons sur le fait que l'humain a besoin de l'autre pour se construire une identité singulière, comme nous l'avons vu. Pour répondre à ce besoin, certains chercheurs et entrepreneurs construisent un robot afin de pallier la solitude et pouvoir interagir avec un robot. À titre d'illustration, nous songeons à la petite Marie, qui est une enfant autiste. Elle ne supporte pas que ses parents lui demandent d'arrêter une activité pour aller manger, passer aux toilettes, se coucher, se brosser les dents, ou encore sortir avec eux. Mais, si le petit robot acheté par ses parents lui rappelle qu'il est l'heure – « c'est l'heure » –, Marie, obéissante, arrête toute activité et range ses affaires et s'implique dans la nouvelle tâche. Un vrai soulagement pour sa maman, qui évite ainsi d'avoir une relation empreinte de tensions et de crises de son enfant. Cette relation parent-enfant médiatisée par un robot soulève des interrogations sur le plan des identifications, des incorporations et des introjections.

Ces exemples ne révèlent malheureusement rien des quatre enjeux fondamentaux de la subjectivation que nous avons évoqués plus haut. Nous sommes aussi loin des présupposés de Henri Wallon ou de Donald Winnicott ou encore de John Bowlby sur la construction du sujet humain, construction médiatisée irrévocablement par un autre sujet suffisamment contenant, maintenant qui le reconnaît. Dans le dernier exemple illustratif, nous sommes en présence d'un être obéissant certes, capable d'adopter un comportement social adapté aux exigences du milieu familial, mais restreint à des mesures éducatives et sous couvert d'injonctions très directives.

D'autres situations nous interpellent également lorsque des firmes construisent des robots pour remplir les fonctions de réconfort, de « présence ». Nous ne pouvons nous empêcher de penser au dispositif expérimental de John Harlow avec les bébés macaques et les deux poupées en fil de fer, l'une recouverte d'une couverture douce et l'autre distribuant du lait, dispositif créé pour démontrer le besoin vital de proximité affective. Seulement à long terme, une interaction authentique et mutuelle s'avère indispensable pour construire son unité de base, son intégrité, ce que ne montre pas ce dispositif expérimental, et nous doutons que l'interaction homme-machine puisse le faire.

Concevoir un robot pour servir d'aide auxiliaire peut se montrer utile lorsque la personne peut difficilement se mouvoir. Cette fonction d'aidant en quelque sorte nous invite à concevoir l'utilité des machines pour alléger et faciliter les tâches de pénibilité, mais en aucun cas ces machines ne permettent d'entretenir une relation authentique. Nous pouvons même envisager que cette

aide technologique permet de dégager du temps pour s'investir dans des tâches plus nourrissantes sans doute.

Le robot émotionnel peut être représenté par un hologramme, véritable fée du logis pour combattre le sentiment de solitude si insupportable pour certains d'entre nous. Une relation de dépendance se tisse entre l'homme et son hologramme, comblant artificiellement tout manque, donnant l'illusion de complétude, tout comme le sujet qui est dépendant d'un objet addictif. Ce dernier revêt ainsi tous les attributs de l'idéal qui satisfait tous nos besoins.

Enfin, nous terminerons sur le rôle de l'avatar pour aborder la problématique du double dans certains jeux vidéo. Certains jeunes et moins jeunes ne parviennent plus à se différencier de l'avatar, oblitérant toute distance, toute différenciation entre le Je et l'autre, et n'étant motivés que par la recherche de satisfactions immédiates. Nous sommes là dans la pathologie. Nous précisons que des recherches en psychologie clinique, en psychologie du développement et en sciences cognitives, notamment, se multiplient pour tenter de comprendre les processus mis en jeu selon les dispositifs individuels ou groupaux médiatisés par le jeu vidéo, pour tenter de répondre à certaines demandes issues du monde de l'apprentissage et de la rééducation pour acquérir des comportements sociaux, pour renforcer son estime de soi par exemple.

Aussi, dans cette interface homme-machine, nous ne pouvons nous empêcher de songer à la création d'un robot psychothérapeute. Évidemment, cela a été pensé et conçu par l'intermédiaire d'un logiciel créé par la USC (University of Southern California). Depuis déjà deux ans, cet outil analyse les comportements des êtres humains simplement en discutant avec eux.

- « Bonjour, comment allez-vous aujourd'hui? », demande Ellie, le psychothérapeute artificiel.

Pendant que la personne répond à la machine, les moindres mouvements ou le ton de la voix sont examinés de façon approfondie. Le logiciel s'intéresse davantage à la manière dont la personne répond aux questions qu'aux réponses elles-mêmes. Nous cernons déjà les limites de cette prouesse technologique, qui n'est finalement que réduite à des algorithmes.

Toujours au regard des enjeux fondamentaux qui fondent la subjectivité du sujet, nous découvrons une autre pratique humaine qui consiste à nier la mort d'un être cher. Celui-ci continue à vivre dans le monde virtuel, une sorte de « i.cloud », un espace extérieur à son psychisme, à sa mémoire épisodique qui permet de prolonger artificiellement le lien à cet autre mort, mais auquel on interdit de partir. Tout travail de distanciation et d'intégration psychique et émotionnelle de cette perte réelle est empêché, car il suffit d'ouvrir cette interface pour revoir et réentendre cette image désincarnée.

En corollaire, nous voyons comment notre liberté perd en espace d'extension et d'expression finalement si nous entretenons l'illusion d'être dotés d'un pouvoir qui se limite à brancher et débrancher l'objet technologique.

Enfin, quel sens pouvons-nous donner à notre vie avec un robot, même si celui-ci est notre robot, personnalisé pour répondre à tous les critères facilitant une identification projective, et donc enfermant encore tout espace de liberté à ce robot, et empêchant par conséquent toute relation authentique pour respecter notre singularité et notre subjectivité ?

Pour aller plus loin dans la réflexion, nous vous invitons à accorder de l'importance au rêve, aux scènes oniriques qui nous sont propres, toujours surprenantes mais qui ne cessent de nous parler. L'imprévisibilité est ainsi de mise, et il nous revient de le savoir, même si nous ne désirons pas en savoir davantage, car nous parlons de notre mort, bien entendu.

CONCLUSION

Enfin, nous songeons aux risques d'évoluer vers une pauvreté des interactions sociales, notamment si le sujet ne se construit que dans un environnement virtuel, un monde prôné par l'intelligence artificielle. Nous pensons aux adolescents, aux adultes qui s'engouffrent à long terme dans ces jeux au point de se confondre avec leur avatar faisant l'économie de l'altérité et de l'incertitude, de l'inconnu, autrement dit, de ce qui nous attend dans notre vie réelle.

Nous pensons au risque d'appauvrir la complexité émotionnelle du sujet s'il reste assujéti à ses mécanismes de défense psychique tels que le clivage notamment, le monde serait ainsi catégorisé en deux, la partie qui nous donne entièrement satisfaction et celle que nous rejetons, car trop complexe et incertaine. Le registre des affects serait insuffisamment développé, entraînant un appauvrissement des sentiments et des réponses, et appauvrissant toute interaction avec autrui, mais aussi avec soi-même. Ainsi, si le sujet s'en remet de suite à un objet qui pallie toute présence au point de s'y substituer, la vie émotionnelle régressera à une relation fusionnelle qui serait vécue comme seule fiable pour maintenir une continuité d'exister. Et si cet objet venait à disparaître ? Un autre, identique et non similaire, viendrait-il de nouveau combler un vide, pourtant nécessaire pour se sentir vivre, en éprouvant le manque, la douleur, la souffrance de la perte du lien à l'autre, en somme qui permet de ne pas oublier le sens de notre existence ?

En outre, tenter de dénier toute souffrance liée à la perte notamment (d'un membre du corps, d'un membre de la famille, de son travail, se sentir abandonné, ne plus se sentir sujet à la suite d'un événement traumatique, etc.)

ne laisserait qu'un espace créatif personnel réduit, voire inexistant, supprimant de ce fait la capacité de concevoir des projets, de créer des alternatives pour continuer à avancer selon sa propre trajectoire et à son propre rythme au détriment de la créativité imposée par l'extérieur.

Ce défaut d'expression humaine pourrait conduire à la difficulté grandissante de ne plus être capable de penser par soi-même, confondant les connaissances, les outils avec le développement de sa réflexivité plus autonome, bien qu'elle soit dépendante de ses environnements (choisir de régresser, de rebondir, etc., selon la période de vie, sa situation).

Or, être humain, c'est pouvoir s'inscrire dans le lien incarné et dans la culture qui comporte maintenant les nouvelles technologies et qui contribuent à construire l'identité à la fois intra et intersubjective. Nous ne pouvons que faire avec et nous interroger sur la finalité du développement de toutes ces technologies qui sont au service de qui et de quoi finalement ? Est-ce que le sujet reste encore au centre des préoccupations des instances qui nous gouvernent, ou est-ce qu'il n'est qu'un objet parmi d'autres pour consolider un fantasme de tout opérationnaliser, rentabiliser excluant toute singularité, toute aspérité qui échapperait au contrôle et entraînerait une perte d'argent ? C'est l'un des risques majeurs qui inquiète Rolland Gori et tant d'autres.

CHAPITRE 3

Personne transformée, médecine régénérative et humanisme

Christian Byk

« Un humanisme bien ordonné ne commence pas
par soi-même. »

Claude Lévi-Strauss, *L'origine des manières de table*,
Paris, Plon, 1968, p. 422.

« Pour être homme, (il faut) refuser d'être Dieu. »

Albert Camus, *L'homme révolté*, coll. « Folio »,
Paris, Gallimard, 1985.

INTRODUCTION : DYNAMIQUE DES CONTRADICTIONS ET (NOUVELLES) FINALITÉS SOCIALES

Dans cette trilogie, quand l'idée d'humanisme rassure, les deux premières notions (personne transformée et médecine régénérative) ouvrent des perspectives encore mal connues. Par ailleurs, le primat de la personne et les droits de l'homme, que l'humanisme implique, constituent des concepts fondamentaux du droit, alors que l'expression médecine régénérative renvoie à des développements récents de la science médicale susceptibles de les transformer.

Nous avons donc à nous interroger sur la relation entre des progrès médicaux en évolution et des notions juridiques fortement ancrées dans notre système de droit, autrement dit à entrer dans une dialectique désormais classique entre une science qui évolue et un droit qui résiste.

Il y a bien sûr de l'ironie dans cette présentation, car la question n'est pas de savoir pourquoi médecine et droit interagiraient dans des dynamiques de forces opposées, mais quelles sont les finalités respectives de ces deux savoirs et pratiques dans une société humaine. Autrement dit, il nous faut interroger les racines morales et politiques du droit, en l'espèce l'humanisme dont nos sociétés occidentales se réclament.

C'est à l'aune de cette recherche qu'il faut lire et interpréter la question de la transformation de la personne et celle du caractère régénératif de la médecine (I). Quant aux droits de l'homme et au rôle que joue la notion de personne humaine, l'attachement qu'y portent aujourd'hui nos sociétés ne doit faire oublier ni leur origine récente, ni leurs limites par rapport aux questions que nous avons à résoudre collectivement (II).

1. LA TRANSFORMATION BIOLOGIQUE DE LA PERSONNE : L'ÉMERGENCE D'UNE NOUVELLE RÉALITÉ AUX FINALITÉS AMBIGUËS

On pourrait penser que la notion de médecine régénérative est plus claire ou, du moins plus limitée, que la notion de personne transformée mais, à vrai dire, l'une et l'autre nourrissent des ambiguïtés ou obligent à investir de nouveaux territoires et nous mènent hors des frontières de nos certitudes.

En effet, l'homme est un « animal » en évolution, tant à travers ses caractères génétiques et physiologiques que par son comportement, ce qui rend complexe toute recherche tendant à définir son identité¹. Toutefois, la connaissance que nous avons acquise de ceux-ci² nous permet d'envisager une contribution active à cette évolution. Mais, après avoir tué Dieu, au sens où nous nous serions emparés, grâce aux nouvelles technologies, d'un pouvoir démiurgique sur l'homme, allons-nous aussi tuer l'homme dans la mesure où l'usage de ce savoir et les finalités auxquelles il nous conduit nous feraient entrer dans une rupture radicale avec la nature humaine, l'humanisme³ ?

-
1. C. De Guibert (dir.) (2016), « Existe-t-il un seuil de l'humain? », *Revue Tétralogiques*, n° 21, Centre interdisciplinaire d'analyse des processus humains et sociaux.
 2. C. Stoll (2008), *Histoire des idées sur l'évolution de l'homme*. Thèse, Université Louis Pasteur, <http://scd-theses.u-strasbg.fr/1551/01/STOLL_Claude_2008.pdf>.
 3. E. Brochier (2004), « La fin de l'évolution humaine dans le grand récit transhumaniste : penser la nature avec Ray Kurzweil, Giordano Bruno et Saint Thomas d'Aquin », *La revue thomiste*,

A priori, la finalité médicale et scientifique est pour l'homme une finalité éminemment humaniste, puisqu'elle vise à nous restituer ce que nous sommes en améliorant les connaissances sur l'homme et en les utilisant afin de combattre les maladies. Mais la médecine, y compris la médecine hippocratique universaliste, est aussi conditionnée par le contexte culturel et social dans lequel elle s'exerce. Le devenir de celle-ci est donc lié d'une certaine manière au projet social et à son évolution. Faut-il alors se méfier de cet ancrage de la pratique médicale dans une société donnée caractérisée par des valeurs, une culture, un système économique ? Faut-il considérer que la dynamique qui est aujourd'hui celle du développement de la médecine et de la science n'est pas exempte d'ambiguïté et de risque pour la sauvegarde de l'idée d'humanisme⁴ ?

A) La légitimité de la finalité médicale et scientifique

Dans une civilisation guidée par la raison et le respect de la personne et du bien commun, la médecine se présente comme une contribution aux objectifs qui doivent permettre de donner une effectivité à ces valeurs.

Le passage d'une médecine réparatrice à une médecine régénérative n'est-il alors qu'une question de degré dans l'effectivité de la finalité sociale confiée à l'art médical ? Dans quelle mesure, au contraire, l'acquisition des connaissances médicales nous ferait-elle entrer dans une ère plus radicale, celle de l'homme amélioré ?

1) De la médecine réparatrice à la médecine régénérative

La raison a fait son chemin. Nous ne considérons plus que la médecine, lorsqu'elle répare les accidents de la vie (maladies, suites d'un accident), s'oppose à la volonté divine. Nous voyons même d'un œil positif le fait qu'elle éradique, au mépris de la biodiversité, certaines maladies, cause d'épidémies effrayantes.

Face à la souffrance, nous acceptons et encourageons sa compassion à soulager nos douleurs, voire à faire disparaître, si possible, les inégalités qui affectent notre santé. Alors, si on doit faire en sorte que la médecine apporte sa contribution à combattre les inégalités sociales en matière de santé, pourquoi ne devrait-elle pas également le faire s'agissant des inégalités d'ordre génétique ou physique ?

Tome CIV, <<https://www.revuetomiste.fr/la-revue-thomiste/nos-publications/4944-article-cxvii-2-273-313>>.

4. M. Foucault (Entretien en juin 1966), « L'homme est-il mort ? ». *Dits et écrits, I, 1954-1975*, <<http://lesilencequiparle.unblog.fr/2009/02/24/lhomme-est-il-mort-michel-foucault/>>.

La médecine régénérative, n'est-ce pas simplement user des nouvelles capacités médicales pour redonner une peau aux grands brûlés (grâce à des transplants de peau fabriqués par des techniques de croissance artificielle), reconstruire des os ayant subi des fractures délicates (en insérant des matériaux qui servent d'échafaudage à cette reconstruction), faire voir l'aveugle ou donner à l'enfant l'immunité biologique qui lui permet de vivre comme les autres enfants? Ainsi, la médecine régénérative ne serait qu'une médecine qui restituerait à la personne sa pleine dimension humaine, puisqu'elle lui permettrait d'être pleinement en relation avec les autres.

Mais la médecine régénérative, qui consiste à réparer une lésion ou un organe malade en remplaçant les parties endommagées par un nouveau tissu cellulaire créé pour l'occasion, repose sur la thérapie cellulaire, consistant à cultiver des cellules avant de les transplanter, et sur l'utilisation de cellules souches, c'est-à-dire non différenciées, à l'image de celles qui composent le très jeune embryon, ou dédifférenciées⁵. Ces cellules sont dites pluripotentes, car elles peuvent ensuite être transformées en différents types cellulaires. En agissant ainsi, ne nous avançons-nous pas dans une nouvelle étape de nos capacités «déliuriques», celle de la production du vivant, et qui plus est en nous servant du vivant humain⁶?

2) De l'acquisition de nouvelles connaissances (génétique, embryogénèse, neurosciences) à la tentation d'améliorer des capacités humaines

Le premier élément, qui ouvre la voie à cette tentation, est la mise en place d'une véritable chaîne liée au développement des nouvelles technologies médicales, tant sur le plan de la recherche que sur celui de la production industrielle.

En effet, on peut penser que l'essor des biothérapies cellulaires et l'exploitation industrielle des cellules souches mettront en jeu de profondes transformations, qui nécessitent d'étudier dès maintenant la manière et les moyens d'aboutir à des produits cellulaires industrialisés, ainsi que l'automatisation des procédés de production et leur miniaturisation.

5. Voir à titre d'exemple: G. Cossu *et al.* (3 mars 2018), «Lancet commission: Stem cells and regenerative medicine», *Lancet*, vol. 391, n° 10123, p. 883-910.

6. K.S. Amegatsevi (2013), *L'éthique du futur et le défi des technologies du vivant*. Thèse en cotutelle, doctorat en philosophie, Université Laval (Québec), Université Paris-Descartes, <<https://iatranshumanisme.com/wp-content/uploads/2017/01/lc3a9thique-du-futur-et-le-dc3a9fi-des-technologies-du-vivant.pdf>>.

Ce passage de l'artisanat à l'industrie conduira notamment, ainsi que le suggèrent nombre de chercheurs⁷, à disposer d'importantes banques de matériaux cellulaires et tissulaires d'origine humaine.

En outre, l'apport de la biologie de synthèse⁸ ne doit pas être négligé dans ce processus, car si l'ingénierie tissulaire permet de réparer les tissus abîmés, les techniques actuelles ne permettent pas de bien contrôler la forme exacte de leur croissance. Dès lors, grâce à la biologie de synthèse, on pourrait fabriquer des échafaudages « intelligents », capables de mieux guider la croissance du tissu et même d'étendre ces techniques d'ingénierie tissulaire à la (re)construction d'organes!

Les perspectives d'un homme remodelé, reconstruit, ne peuvent dès lors être éludées⁹, et ce, d'autant plus que le vieillissement de la population, tant pour des raisons économiques (coût de la dépendance) que sociales (comment faire fonctionner un système de gouvernement démocratique lorsqu'une part importante de la population, celle qui précisément participe le plus au vote, souffre de maladies neurodégénératives?), invite à redoubler d'efforts pour régénérer les corps ou, du moins, assurer une certaine qualité de vie à l'homme vieillissant¹⁰.

C'est dire combien les choix en matière de médecine régénérative sont liés à des finalités sociales.

B) Les finalités sociales et l'acceptabilité de la médecine régénérative

Du « grand espoir du XX^e siècle »¹¹ à la grande angoisse du XXI^e siècle, la puissance de la technoscience et la mondialisation, qui a succédé à l'échec politique des idéologies, ont inscrit le nouveau millénaire dans l'ère des peurs

-
7. M. Peschanski (2009), *Et si nous rêvions de médecine régénératrice?*, MT, vol. 15, n° 3, p. 171-174; voir également le Rapport CIB-UNESCO sur les métadonnées et la santé, 2017 : « Pour les cellules souches embryonnaires, cela passe probablement par l'établissement d'une banque de cellules internationale et centralisée, dans laquelle les médecins viendraient puiser quelle que soit leur nationalité et leur discipline. »
 8. G. Fioraso (2012), *Les enjeux de la biologie de synthèse*, Office parlementaire d'évaluation des choix scientifiques et technologiques, rapport n° 378, Sénat, <<http://www.senat.fr/rap/r11-378-1/r11-378-11.pdf>>.
 9. A. Fagot-Largeaut (2009), De la chimie de synthèse à la biologie de synthèse, La lettre du Collège de France, n° 26, p. 33-36, <<https://journals.openedition.org/lettre-cdf/158>>; B. Bensaude-Vincent et D. Benoît-Browaeys (2011), *Où va la biologie de synthèse?* Paris, Seuil.
 10. H. Chneiweiss (2012), *L'homme réparé : espoirs, limites et enjeux de la médecine régénératrice*, Paris, Plon, 2012; voir l'introduction : « Mourir jeune, le plus tard possible ».
 11. J. Fourastié (1949), *Le grand espoir du XX^e siècle*, Paris, PUF.

collectives et de l'incertitude sur l'avenir de nos sociétés¹². Dès lors, les finalités sociales errent entre égocentrisme individuel et repli communautaire, rendant ambiguë ou improbable la recherche d'un nouveau projet collectif.

Cela n'est pas sans conséquence sur la question de l'acceptation sociale et l'essor de la médecine régénérative.

1) La quête d'identité

La transformation des liens familiaux¹³ constitue le premier facteur d'incertitude identitaire. Depuis les années 1970, la famille a été profondément bouleversée : fondée sur des normes socialement imposées (« les liens sacrés du mariage »), une hiérarchie stricte (celle du « chef de famille » sur la femme et les enfants), une division des rôles (entre homme et femme, parents et enfants), elle a aujourd'hui perdu ces caractères au profit d'une famille reconnue dans sa diversité, sa complexité, sa dimension égalitaire (entre parents, entre enfants) et les choix qu'elle laisse à ses acteurs pour contribuer à définir leurs liens, notamment en matière de parenté et de filiation. À cet égard, les technologies de la reproduction, largement acceptées aujourd'hui en Europe¹⁴, ont coparticipé à cette maîtrise individuelle du corps, qui sépare procréation et sexualité, libère progressivement la femme de l'emprise masculine et sacralise le désir d'enfant. Les nouveaux droits de l'homme, dont témoignent notamment la *Déclaration universelle sur la bioéthique et les droits de l'homme* (UNESCO) ou la convention d'Oviedo (Conseil de l'Europe), confirment que la famille n'est désormais protégée que comme un cadre d'épanouissement de projets de vie individuels.

La famille est ainsi un bon exemple de la transition de nos sociétés à la fin du XX^e siècle : de la permanence, l'unité, la cohésion du collectif, on est passé à la subjectivation, la diversité et la mise en avant de la pluralité des individus¹⁵. D'où la question de la possibilité de la réalisation d'un nouveau projet collectif.

12. B. Badie *et al.* (2001), *Qui a peur du XXI^e siècle*, Paris, La Découverte, 2006 ; lire également N. Farouki (dir.) (2001), *Les progrès de la peur*, Le Pommier.

13. M. Gauchet (2010), « Mutations dans la famille et ses incidences », *La Revue lacanienne*, 3, n° 8, p. 17-32.

14. Ainsi, même les Irlandais ont, par un référendum du 24 mai 2018, approuvé la légalisation de l'avortement.

15. J.-H. Deschaux (2011), « La famille à l'heure de l'individualisme ». *Revue Projet*, 3, n° 322, p. 24-32, <<https://www.cairn.info/revue-projet-2011-3-page-24.htm>>.

2) *La réalisation d'un nouveau projet collectif est-elle encore possible ?*

Paradoxalement, la seule certitude est aujourd'hui celle de Pierre Dac, lorsqu'il disait qu'il ne faut jamais regarder derrière soi, car c'est le plus sûr moyen de tourner le dos à son avenir¹⁶. Beaucoup d'entre nous, par facilité, voire aveuglement, vivons en effet dans l'immédiateté et l'illusion. Facilité de consommer et de jouir, tentative de nous replier sur des identités communautaires ou religieuses dont nous récitons les dogmes comme une table de multiplication sans nous poser la question du rapport avec un monde qui change.

Si un destin commun suppose un récit pour le construire, une telle narration et les traits essentiels de ce destin doivent toutefois se nourrir des réalités.

Or, l'effectivité d'un projet collectif est tributaire de ce mariage, de cette osmose complexe entre une histoire, ses traditions et un monde environnant dont les dimensions et le temps sont d'une autre nature.

C'est pourquoi ce que nous qualifions de toute-puissance humaine doit nous ouvrir la voie vers une interrogation sur l'homme et la signification de son destin. C'est en ce sens qu'il convient de comprendre le « Dieu est mort » de Friedrich Nietzsche. Cet énoncé affirme, en effet, que la question essentielle n'est plus sur le plan de l'existence de Dieu, mais de la signification de Dieu pour l'existence humaine. Et pour l'appréhender, il faut « faire « désexister » Dieu en tant que soi, en tant qu'« en soi », car « Tant que Dieu est cet en-soi métaphysique qui s'impose à l'homme, la question de la signification ne peut être posée, puisque le sens est imposé¹⁷. »

Mais la mort de Dieu, c'est d'abord « le sentiment que l'espace géographique dans lequel l'homme se situait et prenait conscience de son identité disparaît », car « avec la mort de Dieu, l'homme perd son assurance ontologique [...] Elle est bien exigence que l'homme doit porter pour devenir lui-même créateur de soi¹⁸ ».

Ainsi, l'homme d'aujourd'hui fait-il désormais face aux limites inconnues que la technologie lui ouvre. Il découvre, explore et étend les nouvelles frontières

16. La citation exacte de l'humoriste, en l'espèce le Sahr Rabindrana Duval : « L'avenir de Monsieur est devant lui et l'aura dans le dos à chaque fois qu'il fera demi-tour. »

17. Cette expression apparaît pour la première fois dans le « Gai savoir » (1882) aux aphorismes 108, 125 et 343 ; lire Y. Ledure (2001), « Mort de Dieu et volonté de toute puissance », *Le Portique*, n° 8, <<https://journals.openedition.org/leportique/126>>.

18. *Ibid.*

de son corps¹⁹, ce qui conduit à s'interroger sur le corps régénéré et sur le renouvellement de la perception sociale de la personne humaine.

2. LE CORPS RÉGÉNÉRÉ ET LE RENOUVELLEMENT DE LA PERCEPTION SOCIALE DE LA PERSONNE HUMAINE

Il s'agit de questionner ce qui fonde la personne humaine et, au-delà, sa relation avec le droit²⁰ au vu des nouvelles connaissances scientifiques.

Préliminairement, nous tenons à dire qu'il ne nous paraît pas pertinent d'entrer ici dans la querelle juridique et ontologique sur le rapport entre la personne humaine, censée être hors de la disponibilité du droit au vu d'un « constat objectif », et la personnalité juridique, qui constituerait une notion abstraite et fonctionnelle du droit visant à reconnaître à certains des droits subjectifs²¹.

Il nous semble, en effet, que ce qui est en question ici est notre capacité à transformer la personne humaine, voire à créer d'autres types de personnes, et à s'interroger pour savoir si leur humanité transformée ou créée peut les maintenir ou les introduire dans la catégorie des personnes juridiques. Autrement dit, peut-on envisager de « socialiser » ces transformations de la personne humaine en leur permettant « d'entretenir des relations sociales, d'être un corps socialement situé », en les faisant accéder à un mécanisme, la personnalité juridique, instrument de la mise en œuvre des droits de la personne humaine²² ?

A) La personne humaine en mutation

Il convient ainsi d'entendre la notion de mutation de la personne non plus comme une mutation biologique, mais comme une mutation sociale, car les nouvelles technologies biomédicales agissent aussi sur notre perception sociale et sur les valeurs sociales : elles suscitent des craintes autant que de nouvelles attentes. Dans l'une ou l'autre de ces situations, il y a bien un phénomène d'appropriation sociale, ce qui fait dire à certains auteurs que la

19. C. Byk (2017), « Les nouvelles frontières du corps ». *Revue générale de droit médical*, n° 63, Bordeaux : Les Études hospitalières, p. 33-89.

20. S. Tzitzis (2011), *La Personne, l'Humanisme, le Droit*, Presses de l'Université Laval, Québec.

21. X. Bioy (2012), « Le droit à la personnalité juridique », *Revue des libertés et droits fondamentaux*, chron. n° 12, para. 7, <<http://www.revuedlf.com/droit-fondamentaux/le-droit-a-la-personnalite-juridique/>>.

22. *Id.*, para. 9.

technoscience constitue une nouvelle forme d'idéologie de légitimation de l'exercice du pouvoir²³.

1) **Les craintes que soulève l'idée d'une transformation de l'homme sont-elles justifiées et dans quelle mesure ?**

Rejetons d'abord l'idée que la personne humaine, entendue comme *soma* et *psyché*, aurait été jusqu'à présent singulière et immuable dans ses composantes²⁴, et que seul le jeu démiurgique de l'homme, saisi par la technoscience, serait facteur de mutations profondes.

La connaissance de l'évolution, du rôle des phénomènes physiologiques, génétiques et environnementaux ne saurait, en effet, démentir le parcours de l'évolution qui a conduit à *homo sapiens*²⁵.

La question est de savoir si la maîtrise par l'homme de la génétique humaine et des neurosciences est en capacité d'infléchir ou d'accélérer – et dans quel sens – une évolution qui, dans le temps long, continue à se produire²⁶. Toutefois, nous n'avons guère de recul pour évaluer cette perspective. Aussi, sommes-nous pris en ce XXI^e siècle entre deux logiques contradictoires : la fascination et l'assimilation culturelle, somme toute assez rapide des nouvelles technologies dans leur diversité (PMA, numérique...), et la crainte de l'émergence du « meilleur des mondes ».

Suivant les cultures, suivant les pays, les politiques publiques, lorsqu'elles existent, et les comportements sociaux reflètent la force de ces sentiments opposés.

-
23. J. Habermas (1973), *La technique et la science comme idéologie*, Paris : Gallimard ; lire également pour une « synthèse » des positions sur cette question : A. Dorna (2015), « La technologie est-elle une idéologie qui rêve des androïdes ? », *Cahiers de psychologie politique*, n° 26, <<http://lodel.irevues.inist.fr/cahierspsychologiepolitique/index.php?id=2920>>.
 24. Pour ce qui est de la singularité, voir V. Camos, F. Cézilly et J.-P. Sylvestre (dir.) (2009), *Homme et animal, la question des frontières*, Éditions Quae.
 25. E. Charmetant (2008), « Le propre de l'homme et l'évolution », *Revue Études*, 7, tome 409, p. 148.
 26. J.-D. Vincent (2010), « Hypothèses sur l'avenir de l'homme », *La pensée de Midi*, 1, n° 30, Actes Sud, p. 210 ; D. T. Max (2017), « Au-delà de l'humain, le futur de l'évolution génétique », *National Geographic*, <<https://www.nationalgeographic.fr/magazine/2017/06/au-dela-de-lhumain-le-futur-de-levolution-genetique>>.

2) *L'appropriation sociale des technologies*

L'appropriation des technologies est cependant une caractéristique essentielle du monde industriel, « globalisé »²⁷, car elles sont le moteur des nouveaux équilibres du monde, ceux qui voient s'affronter les États, les puissances émergentes (les BRICS²⁸) et les vieilles puissances, mais aussi ceux qui mettent en avant la mainmise des géants technologiques planétaires, dont les GAFAM²⁹ sont un exemple.

À l'échelon individuel, les nouvelles technologies dans le domaine de la santé, de la procréation, de la communication, des transports nous font entrer dans des modes de vie différents qui se présentent autant comme des facteurs d'émancipation³⁰ que de consommation individuelle³¹. Ce caractère hybride des usages induits des nouvelles technologies participe sans doute au troisième âge de l'individualisme, devenu plus complexe, plus subtil, fruit d'une construction personnalisée³².

Par ailleurs, à un niveau plus global, la dynamique des technologies semble jouer le rôle de facteur de réduction de la diversité culturelle³³.

Certes, le rejet des risques – risque technique et risque pour la démocratie – liés à ces technologies semble peser sur leur acceptabilité sociale³⁴, mais c'est pour autant que ces risques mettent en cause le primat du bien-être

-
27. Académie des technologies (2015), *Quelques réflexions sur la question de l'appropriation des technologies*, EDP Sciences.
 28. Brésil, Russie, Inde, Chine et Afrique du Sud (pour South Africa) ; on lira à titre d'exemple : H. Drouvot et G. Verna (1994), *Les politiques de développement technologique : l'exemple du Brésil*, Éditions de l'IHEAL.
 29. GAFAM, acronyme pour Google, Apple, Facebook, Amazon et Microsoft, les cinq grandes firmes américaines, géantes du web auxquelles on pourrait ajouter les NATU pour Netflix, Airbnb, Tesla et Uber. Voir R. Stross (2009), *Planète Google. Faut-il avoir peur des géants du web ?*; F. Pisani et D. Piotet (2008), *Comment le web change le monde : L'alchimie des multitudes*, Village mondial.
 30. S. Proulx (2015), « Usages participatifs des technologies et désir d'émancipation : une articulation fragile et paradoxale », *Communiquer*, n° 13, p. 67-77 ; J. Söderberg (2013), « Imprimentes 3 D, Dernière solution magique, Illusoire émancipation par la technologie ». *Le Monde diplomatique*, p. 3 ; N. Aurray et S. Ouardi (2014), « Numérique et émancipation. De la politique du code au renouvellement des élites », *Mouvements*, Éditions La Découverte, 3, n° 79, p. 13-27.
 31. V. Scardigli (1983), *La consommation, culture du quotidien*, Paris : Presses universitaires de France.
 32. F. De Singly (2005), *L'individualisme est un humanisme*, Éditions de l'Aube, La Tour d'Aigues ; O. Bobineau (2011), « La troisième modernité, ou "l'individualisme confinitaire" », *Sociologie, Théories et recherches*, <<https://journals.openedition.org/sociologies/3536>>.
 33. P. Rasse (2013), « Diversité et dynamique du progrès ». *Revue française des sciences de l'information et de la communication*, 2, <<https://journals.openedition.org/rfsic/280>>.
 34. P. Peretti-Watel (2010), *La société du risque*, Paris : La Découverte.

individuel, y compris les conséquences de la dégradation de notre environnement, qui prévaut dans nos sociétés.

En l'état des transformations sociales, on pourrait même conclure que, dans certains domaines (la famille, la santé, l'alimentation), le mariage entre les nouvelles technologies et l'individualisme n'a pas généré les catastrophes que certains oracles avaient prévues³⁵, mais aussi, pourquoi ne pas l'avouer, les résultats que les promoteurs d'un progrès rapide avaient fait miroiter (par exemple, dans le domaine de la thérapie génique).

En fait, ce qui nous fait peur dans la bataille annoncée entre humanisme et « post-humanisme », ce n'est pas tant les « manipulations » biologiques et génétiques sur l'homme que la manière dont elles peuvent affecter notre perception de la personne humaine et de l'exercice de sa liberté individuelle et qui pourrait se résumer par la devise « *Omnes pro uno, Omnis homo sibi* » (Tous pour un et chacun pour soi)³⁶.

Il en va de même de la notion d'intelligence artificielle. Nous savions déjà avec Freud et l'essor des sciences cognitives³⁷ la part que l'inconscient, les émotions et les désirs jouent dans la constitution de « la nature humaine », et les neurosciences nous confirment que l'homme n'est pas qu'un cerveau « rationnel »³⁸ dans les processus de prise de décision. Cela nous oblige à nous interroger sur la centralité de l'intelligence dans la définition de la personne humaine par rapport à d'autres types de mécanismes de conduite. Remettre en question cette vision des choses pourrait bien faire passer au second plan

-
35. J.-B. Fressoz (2014), « Eugène Huzar et l'invention du catastrophisme technologique », *Romantisme, revue du XIX^e siècle*, 4, n° 150, p. 150; J.-P. Dupuy (2004), *Pour un catastrophisme éclairé*, coll. « Points », Paris: Seuil.
36. Il s'agit bien sûr d'un détournement à des fins ironiques de la devise traditionnelle de la Suisse, « *Unus pro omnibus, omnes pro uno* », dont Alexandre Dumas a fait celle des Trois mousquetaires (« Un pour tous, tous pour un »).
37. O. Houdé et L. Degos (2010), « La psychologie cognitive et le rapport science/politique pour mieux comprendre l'homme », Colloque *Regard sur l'homme contemporain*, Institut de France, 7 juin 2010, <<http://www.canalacademie.com/ida5771-Regard-sur-l-Homme-contemporain-a-travers-la-science-la-morale-et-la-politique.html>>; F. Clément et L. Kaufmann (2011), *La sociologie cognitive*, Paris, Maison des Sciences de l'Homme, coll. « Cogniprisme ».
38. A. Damasio (1995), *L'erreur de Descartes : la raison des émotions*, Paris: Odile Jacob; A. Damasio (2003), *Spinoza avait raison : joie et tristesse, le cerveau des émotions*, Paris: Odile Jacob; A. Damasio (2010), *L'autre moi-même, les nouvelles cartes du cerveau, de la conscience et des émotions*, Paris: Odile Jacob; A. Noë (2009), *Out of our heads, why we are not our brain, and other lessons from the biology of consciousness*, Hill and Wang; D. Van Hoorebeke (2008), « L'émotion et la prise de décision », *Revue française de gestion*, Ed. Lavoisier, 2, n° 282, p. 200; F. Wolff (2012), « L'homme n'est pas un animal comme les autres ». *Études*, 11, tome 417, p. 144.

notre peur que robots et cyborgs, qui ne font que mettre en œuvre l'intelligence artificielle dont nous les avons dotés, ne nous surpassent.

Cela impliquerait-il aussi de renoncer à « l'humanisme historique³⁹ », qui constitue la morale politique dont se réclament les sociétés occidentales ? À l'inverse, perçoit-on les prémices d'un retour de l'humanisme⁴⁰ face à la tentation du « post-humanisme⁴¹ » ?

B) « Il faut qu'une porte soit ouverte ou fermée⁴² »

La relation entre l'homme et la technoscience ressemble à un badinage amoureux fait de multiples faux départs avant que la sincérité ne triomphe.

En l'espèce, la sincérité, c'est reconnaître que nous évoluons, nous nous transformons, avec la technologie en même temps que nos connaissances nous permettent de développer plus encore celle-ci.

Mais cette relation entre l'homme, la science et la technologie ne peut plus avoir une finalité purement égoïste, anthropocentrique, et oublier le monde dans lequel nous vivons et auquel nous sommes liés⁴³.

C'est pourquoi « la notion de personne en anthropologie définit donc le savoir d'une société par rapport à l'être humain en tant que situé à une place sociale qui est authentifiée, corrélative de droits et d'obligations et localement conçue comme "naturelle". Une telle conception n'est pas séparable des autres ordres de représentation – cosmogoniques, symboliques, biologiques – propres à toute société⁴⁴ ».

39. B. Quillet (2002), *La tradition humaniste, VIII^e siècle av. J.-C. – XX^e siècle ap. J.-C.*, Paris : Fayard ; A. Bidar (2014), *Histoire de l'humanisme en Occident*, Paris : Armand Colin.

40. J.-M. Besnier (2002), « Un humanisme paradoxal », *La Télémaque*, Presses universitaires de Caen, 1, n° 21, p. 80 ; P. Rabhi (2008), *Manifeste pour la Terre et l'Humanisme*, Arles : Actes Sud.

41. E. Kleinpeter (dir.) (2013), *L'humain augmenté, les essentiels d'Hermès*, Paris : CNRS Éditions.

42. A. De Musset (1845), « Il faut qu'une porte soit ouverte ou fermée ». *La Revue des deux mondes*.

43. L. Ferry (1992), *Le nouvel ordre écologique. L'arbre, l'animal et l'homme*, Paris : Grasset ; C. Larrère et R. Larrère (1997), *Du bon usage de la nature. Pour une philosophie de l'environnement*. Paris : Aubier ; R. Argullos (2004), « Vers un humanisme polycentrique ». *Diogène*, 2, n° 206, p. 176 ; pour le concept d'« anthropocentrisme méthodologique » : G. Hottos (2004), *Philosophie des sciences, philosophie des techniques*. Collège de France, Paris : Odile Jacob, p. 199 ; M.-E. Blackburn (2013), *Le développement durable : une solution éthique pour contrer la détérioration de l'environnement*. Mémoire, Université du Québec à Montréal, <<https://archipel.uqam.ca/5880/1/M13224.pdf>>.

44. Définition de la personne (extrait), *Encyclopedia Universalis*.

Cette définition doit être rapprochée de la « vision du monde de Claude Lévi-Strauss⁴⁵ », pour qui l'être humain est partie intégrante de la nature dont il ne saurait être séparé artificiellement. Il en tire la conséquence que l'esprit et la pensée sont partie intégrante de la nature⁴⁶ : « l'esprit ne peut comprendre le monde que parce qu'il est un produit et une partie de ce monde⁴⁷ », ce qui fait dire aux analystes de sa pensée que « la dépendance des êtres humains par rapport à la nature ne se situe pas seulement sur le plan de la biologie ou de la physiologie ; elle est aussi d'ordre intellectuel⁴⁸ ». Or, depuis Descartes, l'homme est devenu « le maître et possesseur de la nature », ce que la philosophie politique occidentale a traduit par le triomphe du sujet et l'exaltation de « l'humanisme ».

Faut-il alors rompre avec la philosophie humaniste qui imprègne notre philosophie morale et politique ? Mais pour prendre quel chemin ? Celui d'une « gouvernance » de nos sociétés sous l'impulsion d'un sensible réduit à l'émotion, aux passions et combiné avec la quête d'un illusoire repli sur des « identités immuables », ou celle d'un « post-humanisme » faisant de la science et de la technologie le moteur essentiel des transformations humaines au profit d'une minorité ?

Ne faut-il pas plutôt redéfinir le concept d'humanisme ? Appeler à « un nouvel humanisme », à un « humanisme élargi⁴⁹ », c'est alors, comme en bioéthique⁵⁰, prendre acte de la nécessité du « conflit permanent de deux tendances contradictoires, l'une vers l'homogénéité, l'autre vers la production incessante de nouvelles distinctions⁵¹ » et en favoriser la continuité.

CONCLUSION :

L'HOMME MÉDIATEUR D'UN MONDE EN MOUVEMENT

Si l'« Homo sapiens » est depuis 30 000 ans la seule espèce connue sur terre, l'évolution qui a conduit à lui n'est pas pour autant linéaire, et cette

45. E. Terray (2010), « La vision du monde de Claude Lévi-Strauss. L'Homme ». *Revue française d'anthropologie*, n° 193, p. 23-44, <<https://journals.openedition.org/lhomme/24346>>.

46. *Ibid.*, para. 3.

47. C. Lévi-Strauss (1983), *Le regard éloigné*, Plon, p. 163.

48. E. Terray (2010), « La vision du monde de Claude Lévi-Strauss. L'Homme », *op. cit.*, para. 19.

49. Le nouvel humanisme impliquerait que « les droits de l'homme » soient désormais fondés non plus sur sa qualité d'être moral, mais sur sa qualité d'être vivant, en sorte que ses droits s'arrêteraient « au moment précis où leur existence met en cause celle d'une autre espèce » : E. Terray (2010), « La vision du monde de Claude Lévi-Strauss. L'Homme », *op. cit.*, p. 23-44 ; C. Lévi-Strauss (1983), *Le regard éloigné*, *op. cit.*, p. 374-376.

50. C. Byk (2005), « Bioéthique, universalisme et mondialisation : la dynamique des contradictions ». *La Semaine Juridique*, Éd. Générale, n° 7.

51. C. Lévi-Strauss (1978), *Myth and Meaning*, Londres : Routledge, p. 10.

évolution n'est pas terminée à ce jour⁵². Elle dépend, en outre, d'une variété de facteurs : biologiques, génétiques et environnementaux, mais aussi sociaux et culturels. C'est de cette interaction que naît l'homme transformé. Or l'action de ces facteurs obéit à des conditions de temps, d'espace, d'énergie qui sont différentes et dont nous ne connaissons que faiblement la manière dont elles s'interpénètrent. À cet égard, les capacités que la technoscience nous donne ne nous font pas seulement sujets de ces transformations, elles nous en instituent les acteurs et, au regard de notre dimension sociale et culturelle, elles justifient que nous exercions pleinement des responsabilités, elles aussi transformées. C'est à l'humanité, comprise désormais comme englobant l'ensemble du monde vivant, l'environnement et la biosphère, les générations présentes comme les générations futures, que nous devons penser pour mener de façon équilibrée et raisonnée nos actions.

Et c'est dans ce cheminement incertain que nous devons essayer de construire le futur sans avoir la prétention de croire que nous y serons encore présents, mais aussi sans porter en nous un nihilisme destructeur de l'avenir des autres.

Pour suivre la suggestion de Pierre Le Coz à élargir la palette de nos affects⁵³, je vous livre ces mots du poète Jacques Brel, appel où se mêlent l'ironie et le désespoir sur la manière dont nous organisons nos vies : « Vivre, c'est très mauvais pour la santé. Il n'y a rien qui use plus un homme que vivre⁵⁴. »

52. J.-J. Hublin (2016-2017), *Homo sapiens : l'espèce orpheline*. Cours au Collège de France, <<https://www.college-de-france.fr/site/jean-jacques-hublin/course-2016-2017.htm>>.

53. P. Le Coz (2014), *Le gouvernement des émotions*, Paris : Albin Michel.

54. J. Brel (1971), « Interview », Knokke, RTBF, <<https://vimeo.com/34195512>>.

CHAPITRE 4

Le transsexualisme : transformation fondamentale de la personne

Pascale Boucaud

L'idée d'utiliser la technique pour réparer l'homme est apparue très tôt dans l'histoire de l'humanité, lorsque les êtres humains ont refusé de conserver dans leur chair les conséquences des accidents de la vie ou des malheurs de la naissance en remplaçant un membre déficient ou absent par un appareillage. L'inventif chirurgien Ambroise Paré conçut des prothèses articulées de bras, de main ou de jambe¹. Mais il fallut attendre les années 2000 pour voir apparaître les premières prothèses robotisées copiant une fonction, comme la marche et la préhension. Les plus répandues sont les prothèses myoélectriques. Des capteurs sont implantés dans le corps du patient afin de détecter les signaux électriques des muscles qui sont alors envoyés aux moteurs de la prothèse. Cette prothèse robotisée n'est pas pour autant bionique. En cours de recherche, la prothèse bionique devrait mimer le fonctionnement physiologique du membre perdu. Elle permet de résoudre le problème des douleurs fantômes et de recouvrer les sensations du toucher ou de la marche. Si ces prothèses externes sont les plus développées, des recherches prometteuses sont en cours pour les prothèses robotisées internes, destinées à remplacer un organe défaillant ou absent. Les technosciences laissent penser que cela pourrait aller plus loin qu'une simple amélioration de l'homme. Les implants

1. H. Dorvil (2001), « Le handicap. Origines et actualité d'un concept », *In Problèmes sociaux. Tome I. Théories et Méthodologies*, Québec, Presses de l'Université du Québec, coll. « Problèmes sociaux et Interventions sociales », p. 213.

bioniques surpassent les prothèses en mimant les fonctions organiques et en permettant des performances supérieures. Que ce soit à l'extérieur, à la surface de la peau ou à l'intérieur de l'organisme, tous les moyens sont envisagés et utilisés pour augmenter les capacités du corps et du psychisme. Greffes, prothèses, appareillages et implants constituent l'ensemble de ces techniques d'*enhancement*, qui préparent déjà la venue du posthumain.

Se pose bien évidemment la question des effets juridiques de la transformation du corps par la robotisation. La personne deviendrait-elle l'accessoire d'un corps idéalisé et magnifié par la technique? Quel statut juridique pour l'androïde et le robot? La difficulté de l'adaptation du droit à la situation de la personne transformée et les débats doctrinaux qui en résultent trouvent un précédent avec l'exemple du transsexualisme, transformation fondamentale de la personne, rendue également possible grâce aux avancées scientifiques, dans le but d'une réparation ou d'une amélioration psychique vitale.

Les développements qui suivent montrent ainsi le questionnement juridique ouvert par cette transformation ayant un impact réel sur l'identité de la personne, à travers le dialogue des jurisprudences en France et en Europe.

1. UNE TRANSFORMATION IMPRÉVUE PAR LA LOI

Le phénomène du transsexualisme n'apparaît dans la recherche médicale qu'après la Seconde Guerre mondiale. Le docteur Cauldwell, en 1949, intitule un article «*Psychopathia transsexualis*». En 1953, Harry Benjamin donne au mot la forme «*transsexualism*»; il affirmera que la psychothérapie est impuissante devant le transsexualisme, le seul traitement possible étant le traitement hormonal et la chirurgie. Cette même année Christian Hamburger publia le récit de traitement de Georges Jorgenson, qui devint Christine Jorgenson. Ce n'était pas la première tentative de transformation hormono-chirurgicale, mais c'était la plus réussie; d'abord parce que le patient avait survécu, et ensuite parce qu'il avait pu passer pour une femme le reste de sa vie.

Jusque-là, le fait pour un homme de se conduire en femme, ou plus précisément de prétendre être une femme, était considéré par la psychiatrie comme une forme d'aliénation mentale.

Au sein d'une équipe de l'hôpital John Hopkins de Baltimore, J. Money franchira un pas de plus, en définissant l'identité sexuée comme le produit d'un développement biologique et psychosocial où peut se produire une erreur dans la différenciation sexuelle au cours des phases prénatales ou postnatales. L'identité est le produit d'une longue séquence de différenciation biopsychologique qui naît des chromosomes sexuels, se manifeste par des organes génitaux externes et se poursuit par les conduites culturelles liées au genre.

Ainsi identifié par Benjamin et Money comme une entité clinique autonome et moralement justifiable, le syndrome transsexuel ne se caractérise par aucune anomalie sexuelle physiologique. Allant plus loin, le psychiatre américain Richard Stoller va formuler l'hypothèse du transsexualisme « vrai », permettant d'imposer définitivement le phénomène. Il propose le concept d'identité de genre, désignant le fait psychologique par lequel un sujet se sent homme ou femme, « la quantité de féminité ou de masculinité que l'on trouve chez une personne » ou encore « la connaissance ou la perception consciente ou inconsciente que l'on appartient à un sexe et non à l'autre »².

Avec le concept d'identité de genre, nous sommes devant le syndrome transsexuel, marque du « vrai » transsexualisme : ni anomalie de la sexualité, ni perversion ou psychose, mais une conviction inébranlable, irrésistible et surtout précoce d'appartenir au sexe opposé, malgré la reconnaissance par le sujet de sa conformité sexuelle d'origine.

En France, la première équipe hospitalière fut constituée à l'hôpital Saint Louis à Paris en 1979, et la première opération de réassignation sexuelle fut réalisée cette même année par cette équipe composée du psychiatre Jacques Breton, du chirurgien Pierre Banzet et de l'endocrinologue Jean-Pierre Luton.

Une fois la transformation physique réalisée, la personne demande la rectification de son acte d'état civil, la transformation juridique de son identité.

Le droit est interrogé. Ce cas n'est pas prévu expressément par la loi.

Les juges sont saisis. Le transsexuel entre sur la scène judiciaire en sachant qu'elle l'autorise à faire entendre son interrogation et à transformer l'image invisible qu'il a de lui-même en une réalité socialement et juridiquement acceptée. Au bout de la quête où le conduit le trouble de son identité sexuelle, c'est au droit qu'il demande la révélation de son identité. Les demandes de rectification d'actes d'état civil adressées au juge témoignent d'un besoin irrésistible d'être reconnu dans le sexe opposé à celui que l'ordre biologique leur assigne et là où l'ordre juridique ne leur en reconnaît aucun.

À partir d'une méthode fondée essentiellement sur une étude de cas, nous suivrons le grand débat jurisprudentiel qu'a suscité la demande des candidats au changement de sexe tant en France qu'en Europe, jusqu'au triomphe récent du droit au respect de la vie privée dans les jurisprudences de la Cour européenne des droits de l'homme et de la Cour de cassation.

Sans nier l'intérêt du traitement des transsexuels comme moindre mal thérapeutique, la tâche du juge est de rechercher la juste identité entre une

2. R. Stoller (1968), *Recherche sur l'identité sexuelle*, Paris, Gallimard, p. 28-29.

aspiration individuelle et un monde de significations partagées où cette aspiration doit s'inscrire.

Dans cette perspective, l'hypothèse est que les transsexuels ne peuvent être sujets de droit par la simple transformation de leur corps protégée par le respect dû à la vie privée. Ce n'est qu'au terme d'une recherche d'équilibre entre les droits nés de leur nouveau statut et les devoirs que leur imposent les liens de filiation et d'alliance qu'ils pourront trouver une place dans l'espace public.

2. L'ASSOUPPLISSEMENT DU GRAND PRINCIPE D'INDISPONIBILITÉ DE L'ÉTAT DES PERSONNES

La 1^{re} Chambre civile de la Cour de cassation, dans deux arrêts de 1975, fait d'emblée référence à une vérité chromosomique : la demande de changement de sexe doit être acceptée quand elle résulte d'une modification subie (cas d'un déporté qui aurait subi des interventions chirurgicales en captivité), mais elle doit être rejetée quand elle résulte d'un choix délibéré, au nom de l'ordre public et du principe d'indisponibilité de l'état des personnes³. Le transsexualisme était alors trop peu identifié comme entité clinique pour asseoir sa jurisprudence.

Pourtant, le Tribunal de Toulouse, peu après, en donnant satisfaction à un candidat au changement de sexe, inaugurerait une jurisprudence radicalement nouvelle. Les motifs sont consacrés essentiellement au commentaire des expertises diligentées. Questionné sur le point de savoir « si le changement sollicité est susceptible d'apporter au sujet une amélioration de son état de santé », l'expert diagnostique que le « changement de sexe le libérerait d'une angoisse permanente dont on ne peut mesurer les aspects évolutifs fâcheux », et le juge décide que « X ne peut sans risque grave pour lui, supporter le statut social correspondant au sexe déclaré à la naissance »⁴. Le droit paraît remplir ici une fonction explicitement thérapeutique : tout individu qui ne peut supporter son identité sexuée, pour peu qu'il arrive à convaincre un expert, serait en droit d'obtenir une rectification de son état civil.

Au début des années 1980, une nouvelle vague de décisions apparaît, alors que le transsexualisme est mieux identifié et fait l'objet des premiers protocoles. Les jugements se détachent d'une fonction exclusivement curative pour travailler la question dans le champ des relations sociales, que le droit a

3. Cass. 1^{re} civ. 16 décembre 1975, Dalloz, 1976, p. 397.

4. TGI. 24 mai 1978 et Cour d'appel de Toulouse, 11 octobre 1978; JCP, 1981, II, 19528, note Penneau.

pour mission d'organiser. La tendance dominante tient compte prioritairement du sexe psychologique pour assouplir le principe de l'indisponibilité de l'état des personnes⁵.

L'idée qui va dominer est qu'il faut choisir entre les différentes composantes génétique, hormonale, morphologique et psychologique du sexe, celle qui peut réconcilier l'individu avec son corps. Ainsi serait-il ouvert un nouveau droit de la personnalité permettant l'intégration sociale du sujet, dès lors qu'il peut retrouver une image corporelle correspondant à son sexe psychologique. L'indisponibilité de l'état serait dépassée par le recours à une notion psychosociale du sexe où le juge « trouve dans le silence de la loi des solutions humaines qui peuvent être créatrices de droit »⁶. Fait capital, beaucoup de ces décisions n'accordent pas d'effet rétroactif au changement accordé et introduisent une réserve sur les droits des tiers. Elles entendent la voix des tiers, dont elles refusent de léser les intérêts une fois le changement réalisé. On prend donc en compte la responsabilité induite par le changement d'état. On découvre le rapport à l'autre qui institue le sujet dans une dimension familiale ou généalogique. On n'est plus en tout cas dans le soin apporté à un malade, mais plutôt dans l'ébauche du statut d'un sujet non réductible à son individualité.

Face à cette fronde des juges du fond, comment va se positionner la Cour de cassation ?

3. UN TROISIÈME SEXE ?

Dans l'arrêt B⁷, qui sera ensuite soumis à la Cour européenne des droits de l'homme, elle considère que la demande de changement de sexe n'est pas recevable dès lors qu'elle résulte d'une volonté délibérée du sujet.

La première chambre civile de la Cour de cassation maintient sa position par trois arrêts rendus en termes identiques, le 21 mai 1990⁸. La solution était brutalement claire :

Le transsexualisme, même lorsqu'il est médicalement reconnu, ne peut s'analyser en un véritable changement de sexe, le transsexuel, bien qu'ayant perdu certains caractères de son sexe d'origine, n'a pas pour autant acquis ceux du sexe opposé. Le sexe est un élément objectivement déterminé et intangible dont le meilleur critère est celui de la formule chromosomique.

5. TGI Paris, Gaz. Pal., 1983, p. 603, note G. Sutton.

6. Concl. Guéman, TGI Nanterre, 16 octobre 1980, Actes, n° 35, 1982, p. 41.

7. Cass. 31 mars 1987, Bulletin des arrêts de la Cour de cassation, chambres civiles, I, 1987, n° 116, p. 87.

8. Cass. Civ., 21 mai 1990, J.C.P., II, 21588, Rapport M. Massip et Conclusions de l'avocat général FLIPO.

Suzanne, Joséphine Z..., née le 6 mai 1943, a été déclarée sur les registres de l'état civil comme étant de sexe féminin ; après s'être soumise à des traitements médicaux et avoir subi plusieurs opérations chirurgicales, elle a saisi le tribunal de grande instance en lui demandant de dire qu'elle était de sexe masculin, d'ordonner la mention de ce sexe en marge de son acte de naissance et de lui attribuer le prénom de « Jean-Marc » par substitution à ses prénoms féminins ; l'arrêt attaqué (Lyon, 19 novembre 1987) l'a déboutée de son action ; Suzanne J. Z... fait grief à la cour d'appel, d'une part, d'avoir, en refusant de reconnaître son identité sexuelle masculine telle qu'elle résulte de sa morphologie modifiée, de son psychisme et de son rôle social, violé l'article 8, alinéa 1^{er}, de la Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales ; d'autre part, de n'avoir pas donné de base légale à sa décision en ne précisant pas en quoi la discordance entre le sexe génétique et le sexe psychologique, situation qui s'impose à l'intéressé et dont la cour d'appel énonce elle-même qu'elle ne procède pas d'une claire conscience, peut néanmoins être considérée comme volontaire.

Mais la cour d'appel a estimé, au vu du rapport des experts et par une appréciation qui est souveraine, que les circonstances de l'espèce ne permettaient pas de parler, à propos de Suzanne J. Z..., d'état transsexuel et que le changement physiologique intervenu était le résultat d'opérations chirurgicales qu'elle a voulues.

La formule « même lorsqu'il est médicalement reconnu » semble clore une période d'incertitude par un arrêt de fermeture, comme le disent à l'époque d'éminents commentateurs. C'était sans compter sur l'intervention de la Cour européenne des droits de l'homme.

4. REVIREMENT DE JURISPRUDENCE: LE DROIT AU RESPECT DE LA VIE PRIVÉE L'EMPORTE SUR LE PRINCIPE DE L'INDISPONIBILITÉ DE L'ÉTAT DES PERSONNES

La Cour européenne des droits de l'homme est saisie dans l'affaire B contre France le 25 mars 1992⁹. B est un ressortissant français à qui la Cour de cassation a eu tort de refuser la censure d'un arrêt de Bordeaux ne lui accordant ni la modification de l'acte de naissance constatant le sexe masculin, ni le changement de ses prénoms de Norbert en Lyse Antoinette. L'opération chirurgicale, réalisée à l'étranger, n'en a pas moins entraîné l'abandon irréversible des marques extérieures du sexe d'origine de M^{lle} B. La Cour estime que

9. CEDH, 25 mars 1992, n° 13343/87, *France*, AJDA 1992. 416, chron. J.-F. Flauss ; D. 1993. 101, note J.-P. Marguénaud ; RTD civ. 1992. 540, obs. J. Hauser ; JCP 1992. II. 21955, note T. Garé.

la détermination dont a témoigné l'intéressée constitue, dans les circonstances de la cause, un élément assez important pour entrer en ligne de compte, avec d'autres, sur le terrain de l'article 8. La Cour en arrive donc à conclure que la requérante se trouve quotidiennement placée dans une situation globale incompatible avec le respect dû à sa vie privée.

À la suite de cet arrêt de la Cour européenne des droits de l'homme, l'Assemblée plénière de la Cour de cassation opère un revirement total de jurisprudence le 11 décembre 1992¹⁰.

Prenant acte de la condamnation de la France et suivant les conclusions de l'avocat général Jéol, qui lui demande « d'arracher la minorité souffrante des transsexuels au véritable ghetto social dans lequel elle se trouve », l'Assemblée plénière de la Cour de cassation prend clairement position dans un sens radicalement opposé à la première chambre civile :

Lorsque, à la suite d'un traitement médico-chirurgical, subi dans un but thérapeutique, une personne présentant le syndrome du transsexualisme ne possède plus tous les caractères de son sexe d'origine et a pris une apparence physique la rapprochant de l'autre sexe, auquel correspond son comportement social, le principe du respect dû à la vie privée, posé par l'article 8 de la Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales et par l'article 9 du Code civil, justifie que son État civil indique désormais le sexe dont elle a l'apparence, le principe de l'indisponibilité de l'état des personnes ne faisant pas obstacle à une telle modification.

Le revirement est total puisque désormais, les prémisses du raisonnement sont à rechercher dans le droit au respect de la vie privée et non plus dans le principe de l'indisponibilité de l'état des personnes. La référence n'est plus une conception substantialiste où l'état des personnes est indisponible mais, au nom du droit à la vie privée, l'adoption d'un double critère où prévalent l'aspect extérieur (apparence physique et comportement social) et le but thérapeutique (syndrome et traitement).

La place accordée dans le choix de l'argumentation à la théorie de l'apparence permet de reconnaître et de situer la transformation corporelle à sa juste place. Le juge ne dit pas la vérité de la conviction, mais affirme sa vraisemblance, ce qui a le mérite de relativiser le changement de sexe :

L'idée que l'on puisse changer un homme en femme est une idée folle, rappelle C. Chiland ; on ne change pas le sexe, on ne change que l'apparence ; mais c'est cette apparence qui permet au sujet d'occuper la place qu'il souhaite dans le réseau des échanges symboliques : le langage et les rôles sociaux¹¹.

10. Cass. Plénière, 11 décembre 1992, Bull. inf., Cour de cassation, n° 360, 1^{er} février 1993, Concl. Jéol ; note G. Mémeteau, RTD civ. 1993. 97.

11. C. Chiland (1989), « Homosexualité et transsexualisme ». In *Adolescence*, 7, p. 140.

Les juges reformulent ce que dit cette clinicienne du transsexualisme, en donnant une valeur normative à une croyance partagée. Ce positionnement est tout à fait conforme à la fonction du juge qui cherche le juste partage entre l'ordre symbolique du droit et la conviction du sujet partagée par l'observation médicale. On n'insistera jamais assez sur l'importance du diagnostic médical de transsexualisme, préalable à toute candidature au changement d'état civil. Lui seul positionne le raisonnement du juge sur le cas à partir de la conviction argumentée de l'équipe médicale.

5. EXIGENCE PRIMORDIALE DE L'IRRÉVERSIBILITÉ DE LA TRANSFORMATION PHYSIQUE

L'Assemblée plénière de la Cour de cassation posait ainsi quatre conditions à la modification du sexe sur l'acte de naissance : la réalité du syndrome transsexuel (sexe psychologique), le caractère irréversible de la transformation physique, ce qui supposait une intervention chirurgicale, l'apparence physique proche de l'autre sexe et l'adoption du comportement social correspondant (sexe social).

L'exigence de l'irréversibilité était primordiale. Par la suite, une jurisprudence plus libérale émanant des juges du fond s'est développée, notamment en raison de l'évolution des traitements proposés aux transsexuels¹². En effet, les opérations chirurgicales de réassignation sexuelle comme la phalloplastie ou la vaginoplastie avaient montré leurs limites et l'hormonothérapie offrait une véritable alternative aux transsexuels en vue de modifier leur apparence physique (pilosité, voix, etc.). Une circulaire de 2010 est même venue conforter cette jurisprudence en enjoignant aux procureurs de ne plus exiger une ablation et une reconstruction des organes génitaux, mais de se satisfaire d'une seule hormonothérapie¹³.

L'abandon de l'exigence de la chirurgie des organes génitaux posait cependant une difficulté majeure, la réversibilité du processus de changement de sexe. C'est ainsi que la Cour de cassation, dans deux arrêts de 2012¹⁴, n'avait pas véritablement approuvé cette évolution libérale, semblant être encore sensible à l'exigence de stérilité du transsexuel à travers la notion d'irréversibilité¹⁵.

12. V., sur ce point, Ph. Reigné, La reconnaissance de l'identité de genre divise la jurisprudence, JCP 2011. 480 ; F. Vialla, Transidentité, en attendant la Cour de cassation, JCP 2012. 697.

13. Circ. 14 mai 2010.

14. Civ. 1^{re}, 7 juin 2012, 2 arrêts, nos 11-22.490 et 10-26.947, D. 2012. 1648, note F. Vialla ; JCP 2012. 753, note Ph. Reigné. RDSS 2012. 880, obs. S. Paricard.

15. Par exemple. TGI Montpellier, 24 mars 2016, n° 15/03425, LPA 2016, n° 151 p. 8, chron. Vialla, qui retient que le « traitement hormonal peut entraîner le caractère irréversible de la

6. ASSOUPLISSEMENT DE CETTE EXIGENCE AU NOM DU RESPECT DE L'INTÉGRITÉ PHYSIQUE DES PERSONNES

La Cour européenne des droits de l'homme est saisie sur ce point, dans l'affaire Y.Y. contre la Turquie¹⁶.

En droit turc, l'incapacité de procréer est une exigence qui doit être satisfaite avant le processus de changement de sexe, de sorte qu'elle conditionne l'accès du requérant à la chirurgie de conversion. Au vu des éléments du dossier, et notamment des témoignages des proches du requérant devant les instances nationales, la Cour observe que celui-ci mène depuis de nombreuses années sa vie sociale en tant qu'homme. Il apparaît également qu'il a fait l'objet d'un suivi psychologique dès l'adolescence, qu'il a été diagnostiqué comme transsexuel par un comité d'experts en psychologie et que ceux-ci ont par ailleurs conclu qu'il était nécessaire pour lui de poursuivre sa vie sous une identité masculine. Au moment où il sollicite pour la première fois l'autorisation judiciaire de recourir à une opération de changement de sexe, le requérant s'inscrivait donc déjà, depuis des années, dans un parcours de conversion sexuelle. En dépit de ces faits, les juridictions internes lui ont d'abord refusé l'autorisation requise pour le changement physique auquel il aspirait.

La Cour constate que les juridictions internes ont justifié ce refus initial par la seule circonstance qu'il n'était pas incapable de procréer. Elle ne voit pas comment, sauf en se soumettant à une opération de stérilisation, le requérant aurait pu satisfaire à l'exigence d'infertilité définitive dès lors que, biologiquement, il était apte à procréer. Elle considère que le respect dû à l'intégrité physique de l'intéressé s'opposerait à ce qu'il doive se soumettre à ce type de traitements.

Elle estime que l'exigence de l'incapacité définitive de procréer, préalable à une autorisation de changement de sexe, n'apparaît aucunement nécessaire et qu'en déniaut au requérant, pendant de nombreuses années, la possibilité d'accéder à la chirurgie de changement de sexe, l'État a méconnu le droit de l'intéressé au respect de sa vie privée.

La Cour européenne des droits de l'homme est à nouveau saisie contre la France le 6 avril 2017, dans l'Arrêt A.P., Garçon et Nicot. Je m'en tiendrai ici à la présentation du parcours du premier requérant. Inscrit à la naissance sur les registres d'état civil comme étant de sexe masculin, il s'est cependant toujours comporté comme une fille, et son apparence physique a toujours été très féminine. Durant son adolescence, il a été victime de graves troubles de

transformation exigé par la jurisprudence, lorsqu'il entraîne une impossibilité définitive de procréer dans son sexe d'origine».

16. CEDH, 10 mars 2015, Y.Y. c. Turquie.

l'identité. En 2006, alors qu'il est âgé de 23 ans, il amorce une phase de transition en se présentant socialement comme une femme et il entreprend un traitement hormonal suivi par un endocrinologue et un neuropsychiatre. Il produit plusieurs certificats médicaux reconnaissant son syndrome. Le requérant souligne qu'il n'entendait pas initialement recourir à une opération mutilante de réassignation sexuelle, mais qu'il s'y est résigné parce que la jurisprudence française en avait fait une condition de changement d'état civil.

Il décide de se faire opérer en Thaïlande par un médecin qualifié de spécialiste mondialement reconnu. L'opération réalisée a combiné une orchietomie, une vaginoplastie, une clitoroplastie et une labiaplastie effectuées en une seule et même opération. À la fin de cette opération, les organes sexuels mâles ont été remplacés par des organes d'apparence et de fonctionnement féminins, à l'exception des organes reproducteurs.

Dans le respect de toutes les définitions médicales et légales connues, l'opération est irréversible et change de manière permanente l'identité sexuelle masculine pour une identité sexuelle féminine. Le médecin, dans un certificat, souligne que la demande de changement d'état civil est impérative et recevable et fait partie intégrante de sa prise en charge.

Une seconde chirurgienne indique que le requérant a subi une laryngoplastie cosmétique et une orthophoniste signale avoir fait pendant deux ans un travail de féminisation de la voix ; elle précise aussi qu'actuellement, sa voix et son apparence sont parfaitement féminines et concordantes. Deux autres certificats émanant d'un psychiatre et d'un psychothérapeute indiquent qu'il est légitime que M^{lle} A.P. puisse obtenir, pour son insertion sociale et professionnelle, la mise en concordance de son état civil avec son apparence et son souhait.

Le requérant assigne le procureur de la République devant le TGI de Paris pour pouvoir dire qu'il est désormais de sexe féminin et qu'il se prénomme A. Il produit l'ensemble des certificats médicaux. Le procureur demande qu'une expertise pluridisciplinaire soit réalisée, au motif que le requérant a été opéré à l'étranger.

Le TGI jugea que la demande d'expertise était totalement justifiée et nomma trois experts, un psychiatre, un endocrinologue et un gynécologue. Le requérant refusa de se soumettre à cette expertise, considérant qu'elle était irrespectueuse de son intégrité physique et morale. Devant ce refus, le TGI le débouta de sa demande de rectification de la mention du sexe figurant sur son acte de naissance.

La Cour d'appel de Paris confirma ce jugement, soulignant que « l'atteinte à sa vie privée qui pourrait lui être portée du fait de l'expertise demandée, était

proportionnée à l'exigence du constat de l'identité sexuelle, laquelle est une composante de l'état de la personne soumise au principe d'ordre public d'indisponibilité». Elle jugea toutefois que le fait que le requérant soit connu sous un prénom féminin, la réalité de sa vie sociale, justifiaient son intérêt légitime à changer de prénom. Elle ordonna donc la rectification de ses prénoms sur son acte de naissance.

Le requérant saisit la Cour de cassation, qui rejeta son pourvoi.

Invoquant l'article 8 (droit au respect de la vie privée), A. P., E. Garçon et S. Nicot se plaignaient devant la Cour européenne des droits de l'homme de ce que la rectification de la mention de leur sexe sur leur acte de naissance était conditionnée au caractère irréversible de la transformation de leur apparence. E. Garçon se plaignait en outre de ce que la condition de preuve du syndrome transsexuel portait atteinte à la dignité des personnes concernées. A.P. se plaignait enfin de ce que les expertises médicales imposées par les juridictions internes constituaient, au moins potentiellement, des traitements dégradants.

S'agissant de la contestation de la condition d'irréversibilité de la transformation de l'apparence pour faire droit aux demandes de changement de la mention du sexe à l'état civil :

La Cour constate que pour obtenir la reconnaissance de leur identité, les personnes se trouvant dans la situation des requérants n'avaient d'autre choix que de subir préalablement un lourd traitement médical ou une opération chirurgicale qui devait avoir pour conséquence, selon le droit positif français à l'époque des faits, une transformation irréversible de leur apparence. Cela conduisait, selon un très fort taux de probabilité, à exiger leur stérilité. Elle relève à cet égard que dans son avis du 27 juin 2013, la Commission nationale consultative des droits de l'homme souligne que des personnes qui ne souhaitent pas avoir recours à ces traitements et à ces opérations en acceptaient néanmoins la contrainte dans l'espoir de voir aboutir la procédure judiciaire relative à la modification de leur état civil.

Or de tels traitements et opérations touchent à l'intégrité physique de la personne, laquelle est protégée par l'article 3 de la Convention européenne des droits de l'homme.

La Cour observe que le fait de conditionner la reconnaissance de l'identité sexuelle des personnes transgenres à la réalisation d'une opération ou d'un traitement stérilisant qu'elles ne souhaitent pas subir revenait à conditionner le plein exercice du droit au respect de la vie privée à la renonciation au plein exercice du droit au respect de leur intégrité physique.

Elle constate qu'au nom de l'intérêt général, le droit positif français, tel qu'établi à l'époque des faits, mettait les personnes transgenres ne souhaitant pas suivre un traitement de réassignation sexuelle intégral devant un dilemme insoluble : soit subir malgré elles une opération ou un traitement stérilisant et renoncer ainsi au plein exercice de leur droit au respect de leur intégrité physique ; soit renoncer à la reconnaissance de leur identité sexuelle et donc au plein exercice de leur droit au respect de leur vie privée. Elle juge donc qu'il y a là rupture du juste équilibre que les États parties sont tenus de maintenir entre l'intérêt général et les intérêts des personnes concernées.

Elle observe par ailleurs que le 12 octobre 2016, le législateur français a expressément exclu la stérilisation des conditions exigées des personnes transgenres pour l'obtention de la reconnaissance de leur identité¹⁷.

Le nouvel article 61-6 du Code civil précise en effet que « le fait de ne pas avoir subi des traitements médicaux, une opération chirurgicale ou une stérilisation ne peut motiver le refus de faire droit à la demande de modification de la mention relative à son sexe dans les actes de l'état civil ».

La Cour estime donc dès lors que la condition d'irréversibilité de la transformation de l'apparence s'analyse en l'espèce, en un manquement par l'État défendeur à son obligation positive de garantir le droit au respect de la vie privée. Elle conclut à cet égard à la violation de l'article 8.

S'agissant de la condition de réalité du syndrome transsexuel posée par le droit français avant 2016, pour faire droit aux demandes de changement de sexe, la Cour observe qu'il existe un large consensus sur ce point entre les États membres du Conseil de l'Europe : sur les quarante États qui reconnaissent juridiquement l'identité de genre des personnes transgenres, seuls le Danemark, l'Islande, Malte et la Norvège ont adopté une législation mettant en place une procédure qui exclut un tel diagnostic. Dans tous les autres États, un diagnostic préalable est exigé. La Cour relève d'ailleurs que contrairement à la condition de stérilité, l'obligation de ce diagnostic ne met pas en cause l'intégrité physique des individus.

17. La stérilisation est également exclue en Autriche, au Portugal, en Allemagne, en Suède, aux Pays-Bas, au Danemark, à Malte, en Irlande, en Italie, en Norvège, en Estonie, en Espagne, en Hongrie, en Islande, en Moldavie, en Pologne et au Royaume-Uni.

7. ÉVOLUTION ULTIME DU DROIT FRANÇAIS : ABANDON DE L'EXIGENCE DE TRAITEMENTS INDUISANT DES MODIFICATIONS DIFFICILEMENT RÉVERSIBLES

La nouvelle loi française précitée supprime également cette condition de diagnostic préalable. Elle est donc plus libérale encore que la position de la Cour européenne des droits de l'homme. Elle vient expressément détacher la modification de la mention du sexe de tout aspect médical, pour le fonder sur l'apparence sociale choisie par l'intéressé, ce qui suscite des interrogations essentielles non résolues par le texte.

L'absence de limites relatives au nombre de changements de sexe et l'abandon de l'exigence de traitements induisant des modifications difficilement réversibles autorisent dès lors ces personnes à faire des allers-retours d'un sexe à l'autre, en fonction de leur ressenti ponctuel. Le sexe n'aurait donc plus vocation à s'inscrire dans la durée, mais à évoluer au gré des flottements de leur sentiment d'identité.

Les magistrats vont-ils suivre cette interprétation et ainsi autoriser plusieurs changements de sexe successifs, donc plusieurs changements d'identité, ou bien exiger, en se fondant sur le principe d'indisponibilité de l'état civil, des qualités propres à la possession d'état, notamment la continuité des faits établis, permettant ainsi d'ancrer dans une certaine mesure la mention du sexe dans le temps ?

Telle est l'une des questions majeures posées par cette réforme du droit français.

8. LES QUESTIONS EN SUSPENS : INCIDENCES DU CHANGEMENT DE SEXE

Comment ne pas penser aux conséquences de la décision sur les nouveaux liens d'alliance ou de parenté dès lors que l'identité sexuée est inversée ? Qu'en est-il des effets du changement de sexe de l'un des parents sur les enfants déjà nés ou à naître ?

Dans la nouvelle loi du 12 octobre 2016, le législateur français a incontestablement voulu s'affranchir de toutes les conditions posées par la jurisprudence dans un texte très libéral, au risque cependant de faire de la mention du sexe un élément flottant de l'état civil et de permettre une situation pourtant redoutée, que neutralisait la Cour de cassation grâce à l'exigence de stérilité, la procréation dans son sexe d'origine, de la personne ayant changé de sexe.

Le lien de filiation par le sang repose en effet sur le principe fondamental que l'enfant naît biologiquement d'un homme et d'une femme dont il est déduit que l'enfant peut établir un lien de filiation paternel à l'égard d'un homme seulement et un lien de filiation maternel à l'égard d'une femme seulement. Chacun de ces liens est d'ailleurs établi en réalité selon des modalités spécifiques : l'accouchement établit généralement le lien maternel, tandis que la présomption de paternité et la reconnaissance permettent d'établir le lien paternel.

Puisque la stérilité n'est plus une condition du changement de sexe, un homme ayant conservé son utérus pourra accoucher et demander à établir un lien de filiation en tant que « mère », tandis qu'une femme ayant pareillement conservé sa capacité de reproduction dans son sexe d'origine pourra demander à reconnaître un enfant accouché d'une autre femme¹⁸. L'enfant pourrait donc être inscrit à l'état civil comme né de deux parents de même sexe.

Dans un arrêt du 14 novembre 2018, la Cour d'appel de Montpellier a officialisé la spécificité de la filiation biologique du parent transsexuel – homme marié devenu femme – ayant conservé et utilisé ses capacités reproductives masculines pour concevoir un enfant avec son épouse. Saisie de la question de la transcription et donc de la qualification – masculine ou féminine – d'une telle filiation, la cour crée une « troisième voie » en inventant, aux côtés des filiations biologiques maternelle/paternelle, une filiation neutre via la mention « parent biologique ».

Les juges du fond étaient en l'espèce saisis d'une affaire concernant un couple marié qui avait donné naissance à deux enfants nés en 2000 et 2004. En 2011, le mari, qui avait fait état d'un syndrome transsexuel, avait bénéficié d'un jugement portant changement de la mention de son sexe à l'état civil, tout en conservant la fonctionnalité de ses organes sexuels masculins. Le couple, toujours marié et continuant à entretenir des relations intimes, donnait naissance en 2014 à un troisième enfant.

Monsieur B, devenu M^{me} B, avait, quelques jours avant la naissance de son troisième enfant, pris soin d'établir devant notaire une « reconnaissance prénatale de nature maternelle, non gestatrice ». À sa naissance, l'enfant avait été déclaré à l'état civil comme né de M^{me} A et portait son nom de famille. Mais M^{me} B avait également demandé ultérieurement la transcription sur l'acte de naissance de l'enfant de la reconnaissance de maternité souscrite devant

18. Plusieurs grossesses de transsexuels FtH ont ainsi été recensées dans le monde, et en France, l'Académie de médecine précise qu'une personne MtF aurait même procréé naturellement : Académie de médecine, Autoconservation des gamètes des personnes transsexuelles et projet parental éventuel, séance du 25 mars 2014.

notaire, ce qui devait lui être refusé par l'officier d'état civil agissant sur instruction du procureur de la République. Il lui était opposé que la transcription demandée doterait l'enfant d'une double filiation maternelle interdite par la loi, ce que M^{me} B devait contester judiciairement.

En 2016, le législateur s'attache, notamment à travers l'article 68-1 du Code civil, à préserver l'intérêt supérieur et la vie privée des enfants déjà nés avant le changement de sexe de l'un de leurs parents. Il prévoit en effet que « le changement de sexe d'un parent est sans incidence sur la filiation des enfants nés de la personne avant ce changement ». Par dérogation à l'article 61-4 du Code civil, les modifications de prénoms consécutives à ce changement ne sont portées en marge des actes de l'état civil des enfants qu'avec leur consentement ou ceux de leurs représentants légaux.

Cependant, le législateur n'a pas réglé la situation des enfants nés après le changement de sexe d'une personne présentant une « assignation sexuelle alternative » (l'une publique résultant d'un changement de sexe accordé par le juge ; l'autre intime/biologique liée aux capacités procréatives du sexe d'origine conservé).

Par rapport à cette situation, les juges d'appel font œuvre prétorienne. En première instance, le Tribunal considère que M^{me} B a fait le choix de revenir de façon unilatérale sur le fait qu'elle est désormais reconnue comme une personne de sexe féminin, et elle doit en assumer les conséquences, à savoir soit procéder à une reconnaissance de paternité sur l'enfant et revenir dans son sexe masculin d'origine, soit engager une procédure d'adoption plénière de l'enfant de sa conjointe et rester dans son sexe féminin.

En appel, la Cour, s'appuyant sur les articles 3-1 et 7¹⁹ de la Convention internationale relative aux droits de l'enfant, relève que :

- la filiation ne peut être établie par la voie de l'adoption en raison de l'intérêt supérieur de l'enfant de voir reconnaître la réalité de sa filiation biologique qui ne correspond ni à une adoption ni à une filiation maternelle « non gestatrice », laquelle nierait à l'enfant toute filiation paternelle et brouillerait la réalité de sa filiation maternelle ;
- la possibilité juridique pour M^{me} B, tout en conservant son identité de femme, de se reconnaître comme le père biologique de l'enfant, n'est pas envisageable : imposer à M^{me} B un retour à son ancien sexe, exiger de sa part un renoncement partiel à sa nouvelle identité

19. Article 3-1 : « l'intérêt supérieur de l'enfant doit être une considération primordiale dans toutes les décisions qui concernent les enfants » ;

Article 7 : « l'enfant a le droit, dans toute la mesure du possible, de connaître ses parents et d'être éduqué par eux ».

sexuelle, l'obliger à révéler la complexité de la situation lors de chaque production du livret de famille seraient contraires au principe du respect de sa vie privée.

Elle estime donc qu'il y a lieu de faire droit à la demande subsidiaire de M^{me} A, en établissant judiciairement la filiation de l'enfant à l'égard de ses deux parents biologiques. Pour ce faire, la cour choisit un véhicule juridique prétoire : la mention, sur l'acte de naissance de l'enfant, de M^{me} B comme « parent biologique ».

Cet arrêt soulève de nombreuses interrogations :

Si une femme devenue juridiquement un homme peut accoucher d'un enfant et si un homme devenu juridiquement une femme peut être biologiquement le père d'un enfant né d'une autre femme, comment l'enfant peut-il établir sa double filiation maternelle et paternelle ? Quelles sont les incidences de l'émergence d'une filiation « neutre » sur les modes de procréation assistée (GPA, PMA) ? Comment, dans toutes ces situations, concilier le droit positif avec les principes posés par la Convention internationale relative aux droits de l'enfant et la Convention européenne des droits de l'homme : vie privée, droit de connaître ses origines biologiques, etc. ?

CONCLUSION

Le rôle du juge consiste à rechercher la juste identité entre aspiration individuelle ou technologique et monde de significations partagées où cette aspiration doit s'inscrire.

Il y a trente ans, la question posée au juge était celle-ci : le transsexuel peut-il changer d'identité juridique par la simple transformation de son corps protégée par le respect dû à sa vie privée ? La réponse des juges était la suivante : le transsexuel, bien qu'ayant perdu certains caractères de son sexe d'origine, n'a pas pour autant acquis ceux du sexe opposé. La transformation de l'identité juridique était donc refusée.

Trente ans plus tard, le grand principe d'indisponibilité de l'état des personnes est battu en brèche par la référence à la protection de la vie privée des personnes, de leur autonomie. L'irréversibilité du changement d'identité n'est même plus exigée par les textes ni par la jurisprudence la plus récente. Homme ou femme ? Sera-t-il possible de faire des allers-retours au gré des flottements du sentiment d'identité ?

Cet exemple inaugure l'exigence d'une réflexion quant à l'incidence de la robotisation de l'humain sur l'identité de la personne.

Avec la technologie du contrôle du robot par la pensée²⁰, les contours de la personne sont encore davantage floutés. Les multiples hybridations robotisées du corps au moyen d'électrodes, d'implants, de puces, etc., ou à tout le moins d'un casque d'électroencéphalographie, ne transformeraient toutefois pas le cyborg en chose, puisqu'il conserverait sa nature première de personne en vertu de l'adage « l'accessoire suit le principal ». Mais cet adage pourrait-il encore servir à modifier la nature juridique des technologies robotiques en personne ? Que ce soit avec un casque d'électroencéphalographie ou avec des électrodes implantées dans le cerveau, toutes les approches réparatrices de la personne par le système du contrôle par la pensée impliquent un transfert de l'intentionnalité de la personne vers des artefacts robotisés situés à distance. Il devient alors difficile de distinguer le sujet agissant de l'objet manipulé, puisque le sujet apparaît passif, tandis que l'objet devient actif et semble doté d'intentions. Pourrait-on alors considérer que le robot contrôlé par la pensée serait une personne par destination, car affecté à son service ?

Quant au robot humanisé, quelle pourrait être son identité juridique ? Wakamaru de Mitsubishi est capable de reconnaître deux « maîtres » et huit autres personnes ; il comprend plus de 10 000 mots et peut prendre l'initiative de parler, en utilisant les informations qu'il a obtenues par observation ou via l'agenda que son propriétaire aura préalablement rempli. Dans son livre *Love and sex with Robots*, David Levy suggère que nous tomberons amoureux de robots affectueux et câlins et que nous aurons même des relations sexuelles avec eux.

S'en tiendra-t-on, comme pour le transsexuel, à la théorie de l'apparence ? Certains auteurs optent déjà pour la reconnaissance de la personnalité juridique aux robots²¹, tandis que d'autres considèrent que l'existence de cette personnalité suppose une double condition : la reconnaissance comme humain et le critère de la conscience morale reconnue à l'humanité²². Selon cet auteur, un androïde qui « pense » selon des programmes, même imitant parfaitement les raisonnements humains, serait dépourvu de la conscience de soi qui fait l'humanité.

Mais qu'en sera-t-il de ce questionnement dans dix ans seulement ?

-
20. A. Phillip, « A paralysed woman flew an F-35 fighter jet in a simulator – using only her mind », *The Washington Post*, 3 mars 2015.
 21. Rapport « Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics », du consortium Robolaw du 22 septembre 2014, <www.robolaw.eu> ; A. Bensoussan (2015), *Droit des robots : science-fiction ou anticipation ?* D.1640 ; « Plaidoyer pour un droit des robots : de la personne morale à la personne robot », *La Lettre des juristes d'affaires*, 23 octobre 2013, n° 1134.
 22. X. Bioy (2013), « Chapitre 7. Vers un statut juridique des androïdes ? », *Journal international de bioéthique*, vol. 24, n° 4, p. 85-98.

Le rôle du juge, gardien de la règle commune, est particulièrement ardu face à tous ces défis. Il s'agit moins pour lui de choisir une vérité dans un code normatif extérieur que de trouver une fiction acceptable – avec la création de nouveaux concepts juridiques – qui réponde à la demande en neutralisant le risque de désarticulation de la société entière au-delà du droit lui-même. Au fil de ce débat jamais clos est réaffirmé un principe d'humanité où s'enracine la représentation de notre identité.

CHAPITRE 5

Personne transformée par l'intelligence artificielle et droit de la responsabilité

Marjolaine Monot-Fouletier

Tout doit paraître naturel, jusqu'au dialogue avec une machine.

Mais la simplicité voire la simplification apparente n'annule pas la complexité invisible des technologies qui s'offrent à nous¹.

Qui agit vraiment en moi quand j'agis? Voici une question philosophique souvent, psychologique parfois, qui va devenir toujours plus juridique à mesure que l'intelligence artificielle, incarnée dans le corps même de l'humain ou sous la forme d'un robot quelconque, assistera toujours plus, voire remplacera, le geste comme la décision et le jugement, humains. Notre corps, « point de rencontre avec la réalité (...) [peut devenir] une simple interface, désirée ou non, avec la seconde réalité, une seconde peau du monde, numérique et vertigineuse »². Notre corps pourrait devenir ainsi le support d'une technologie dont nous ne maîtrisons avec certitude ni le fonctionnement, ni les intentions potentielles, ses concepteurs

-
1. B. Razon (2018), *Philosophie magazine*, n° 119.
 2. *Ibid.*

revendiquant son autonomie sous la forme d'une capacité d'apprentissage, puis de décision indépendante.

La personne sera nécessairement transformée physiquement par l'introduction d'une technologie en son corps, mais elle le sera aussi socialement par le fait qu'elle n'est plus celle qui décide de façon indépendante de l'ensemble de ses comportements. Et ceci vaut également, indirectement, pour les personnes morales qui seront assistées dans leurs démarches par des intelligences artificielles. On pense ici notamment aux *smart contracts*³, ou encore aux perspectives ouvertes par la justice prédictive⁴.

Au-delà de la question de l'identité, de la définition de la personne juridique, se pose ainsi la question de la responsabilité de la personne transformée en son corps ou assistée par une intelligence artificielle déterminante, qui pourra agir sur elle ou à travers elle sans la médiation d'une décision humaine. Dans ce cas, la personne engage-t-elle encore – et sous quelle forme – sa responsabilité, cette responsabilité qui est au cœur des relations humaines, au cœur du contrat social ?

Le droit positif de la responsabilité se compose de divers régimes qui, dans leur très large majorité, reposent sur la détermination d'une imputabilité et d'un lien de causalité entre l'acte ou l'inaction et le dommage : celui qui décide d'agir – ou de ne pas agir – ou celui qui dirige et contrôle ce qui agit est responsable. Au fil du temps, l'aménagement de ces régimes de responsabilité, en particulier l'émergence de la notion de responsabilité du fait des choses, s'est opéré pour permettre l'indemnisation des dommages liés au développement de nouvelles technologies (la machine à vapeur en 1896⁵, le véhicule terrestre à moteur en 1930⁶) : c'est la prise en compte de la part active de la technologie comme intermédiaire complexe entre la décision de l'homme et l'action préjudicielle qui a mené à ce que le droit de la responsabilité évolue, pour pallier les difficultés nouvelles de détermination de la source du dommage, de l'imputabilité⁷, du lien de causalité faute/dommage et donc de l'indemnisation des victimes.

3. Voir Théocharidi (2018), « La conclusion des smart contracts : révolution ou simple adaptation », *Rev. Lamy droit des affaires*, n° 138, p. 28.

4. Voir J.-B. Declercq (2018), « Les algorithmes en procès », *RFDA*, p. 131 ; F. Rouvière (2018), « Le raisonnement par algorithme : le fantasme du juge-robot », *RTD civ.*, p. 530.

5. Cass. civ., 16 juin 1896, *Oriolle, Guisnez et Cousin c/ Tefaine*.

6. Cass. Ch. réun., 13 févr. 1930, *Jand'heur*.

7. Cass. Civ. 2^e, 21 juill. 1982, *Desmares*, arrêt ayant eu une influence majeure sur l'adoption postérieure de la loi n° 85-677 du 5 juillet 1985 relative à l'indemnisation des victimes d'accidents de la circulation dite *Loi Badinter*, dont l'objectif premier tient dans la volonté de faciliter l'indemnisation des victimes.

Or l'émergence des intelligences artificielles contemporaines, autonomes et apprenantes, nous dit-on, fait franchir un nouveau palier de complexité technologique qui renouvelle potentiellement la question de la responsabilité, sachant qu'il n'existe pas une forme d'intelligence artificielle, que par conséquent il serait utopique et peu pertinent de poser un régime juridique de responsabilité unique pour des réalités plurielles.

Si l'intelligence artificielle ne fait que simuler l'intelligence humaine, reproduire un schéma de conduite prédéterminé, sa responsabilité ne saurait être détachée de celle de la ou des personnes, morales et/ou physiques, qui ont permis son émergence et garantissent son fonctionnement, de sorte que le droit commun de la responsabilité s'applique sans difficulté. En revanche, si l'intelligence artificielle prend la forme d'une intelligence individualisée, elle renvoie à une capacité propre de décision (ce qui reste à prouver), à une forme d'émancipation qui met en doute l'applicabilité des régimes actuels de responsabilité. Cette intelligence artificielle constitue en réalité un système faisant intervenir un nombre important d'éléments technologiques eux-mêmes intelligents (des algorithmes) d'origines diverses qui interagissent, sans que l'on puisse déterminer lequel d'entre eux incarne le lieu d'émergence du dommage.

Sécrtant ainsi une responsabilité diffuse impliquant potentiellement les concepteurs des diverses composantes de l'intelligence artificielle, l'autorité l'ayant habilitée, le fabricant du « produit fini » et l'utilisateur, le recours à l'intelligence artificielle défie les cadres de responsabilité existants en ce qu'elle bute sur l'exigence d'établissement d'une faute, d'une imputabilité, d'un lien de causalité.

Il s'agira alors ici de proposer des pistes de réflexion qui puissent entrer dans le cadre des enjeux politiques et sociaux actuels en la matière, à savoir garantir l'indemnisation des victimes, d'une part, et ne pas mettre de frein à l'innovation technologique de l'autre.

Au regard de ces deux enjeux, le droit commun de la responsabilité est-il adaptable au contexte d'action des nouvelles technologies (I) ? Faut-il envisager l'application d'un régime spécial de responsabilité (II) ?

1. DE L'ADAPTABILITÉ DU DROIT COMMUN DE LA RESPONSABILITÉ

C'est la proposition qui peut sembler la plus raisonnable : plutôt que de vouloir à tout prix créer un nouveau régime de responsabilité, convoquons les régimes classiques de la responsabilité pour faute personnelle (art. L. 1240 du code civil) (A) ou de la responsabilité du fait de la garde (art. L. 1242 du code civil) (B).

A) L'hypothétique recours à la responsabilité pour faute personnelle

En sus de la mise en cause des personnes, physiques et morales, directement impliquées dans le processus de survenance du dommage (en particulier le concepteur du système, le fabricant de l'interface technologique et l'utilisateur lui-même, autrement dit la personne transformée) (1), il faudra s'interroger sur la pertinence de vouloir engager la responsabilité de l'intelligence artificielle elle-même (2).

1) La faute humaine

a) Du sachant

Le système d'intelligence artificielle avancée se présente comme une garantie d'amélioration de la personne ; c'est bien en ce sens que l'on parle de personne augmentée. En ce sens, il prétend à une haute fiabilité, voire à une infaillibilité. Par conséquent, ses concepteurs et fabricant devraient se situer au premier rang de ceux susceptibles d'en assumer les failles dommageables, en postulant qu'elles sont liées à une erreur de conception, de programmation.

Toutefois, tant l'identification que l'imputabilité de la faute seront particulièrement délicates à mener dès lors que le système d'intelligence artificielle repose sur une pluralité de composants, donc de concepteurs, et que la traçabilité du processus menant au dommage sera nécessairement limitée. Même dans l'hypothèse où des « boîtes noires » seraient censées assurer la traçabilité du fonctionnement de l'intelligence artificielle, l'utilisation des données recueillies générera sans nul doute des batailles d'expertises complexes⁸ peu adaptées aux nécessités d'une indemnisation rapide des victimes.

Par ailleurs, dès lors que le ou les concepteurs de l'intelligence artificielle l'ont élevée à l'autonomie la plus complète grâce à l'apprentissage profond, comment pourra-t-on prétendre les tenir pour personnellement responsables d'un dommage lié à une décision « inventée » par l'intelligence artificielle selon un processus que, par définition, ils n'auront pas programmé ?

b) De la personne transformée

Cette personne bénéficie d'une technologie modifiant ses capacités et son comportement, mais elle n'en a pas la pleine maîtrise, précisément parce que cette technologie peut agir sans qu'aucune intervention de sa part ne soit requise. En fonction du niveau d'autonomie du système d'intelligence artificielle

8. Voir par exemple Gregor Gössler e.a. « Apport des méthodes formelles pour l'exploitation de logs informatiques dans un contexte contractuel », *Technique et science informatiques* – n° 1/2012, 1-22.

utilisé, la maîtrise de celui-ci devient illusoire et la personne utilisatrice s'efface pour devenir un simple support (!) de l'action de l'intelligence artificielle. Il devient dès lors de moins en moins évident de la tenir pour responsable du comportement dommageable du système qu'elle abrite, mais qui agit sans elle, sauf à considérer que la personne conserve toujours a minima la capacité et, par suite, l'obligation d'arrêter son fonctionnement lorsqu'une telle décision s'impose. Mais dans quelle mesure saura-t-elle dégager à temps la pertinence d'une telle décision ? Dans quelle mesure son temps de réaction, plus ou moins rapide, pourrait-il être pris en compte dans la reconnaissance de son éventuelle responsabilité ?

Ces problématiques liées aux conditions d'imputabilité de la responsabilité à l'utilisateur d'une technologie ne sont pas nouvelles ; on les retrouve d'ores et déjà dans l'hypothèse de la mise en cause de conducteur de véhicule terrestre à moteur lors d'un accident causé par celui-ci. À cet égard, on peut noter que selon la Cour de cassation, le conducteur, à qui reviendra la charge d'indemniser la victime de l'accident dans le cadre de la loi Badinter⁹, ne peut être que celui qui dispose d'une certaine maîtrise du véhicule sans que cette maîtrise soit absolue. Qu'en sera-t-il du degré de maîtrise nécessaire sur l'intelligence artificielle pour continuer à pouvoir sérieusement envisager de retenir la responsabilité pour faute de son utilisateur ?

c) La tentation de la responsabilité pour faute présumée

En toute hypothèse, que l'on souhaite retenir la responsabilité pour faute de l'utilisateur de l'intelligence artificielle ou de son concepteur/fabricant, les difficultés auxquelles devront se confronter les victimes en matière de preuve de la faute commise nous amènent à envisager la nécessité d'instaurer un régime de responsabilité pour faute présumée.

Celui-ci devrait peser prioritairement, nous semble-t-il, soit sur le concepteur, soit sur le fabricant, qui sont les plus aptes à maîtriser l'intelligence artificielle. L'intérêt d'une telle approche, outre bien entendu de faciliter l'indemnisation de la victime, tient dans le fait d'inciter ces acteurs essentiels à proposer des systèmes particulièrement fiables et dont la traçabilité du fonctionnement devra être assurée le plus finement possible.

Toutefois, dans l'hypothèse de technologies évoluant rapidement (mises à jour de logiciels, inventions opérées en autonomie par l'intelligence artificielle elle-même), la preuve contraire d'une faute de conception sera infiniment complexe à apporter. Ainsi, faire le choix d'une faute présumée du concepteur/fabricant revient en pratique à lui faire assumer l'ensemble des risques. Ce

9. *Op. cit.*, note 7.

choix est envisageable, mais risque de constituer un frein majeur à l'innovation, sauf d'une part à établir une répartition de la responsabilité entre les différents acteurs de la conception/fabrication du système intelligent selon leur contribution individuelle effective à celle-ci et, d'autre part, à instaurer une clause d'exonération de responsabilité fondée sur un aléa technologique imprévisible au moment de la conception de l'intelligence artificielle, à l'image de ce que prévoit déjà l'article L. 1245 du Code civil en matière de responsabilité du fait des produits défectueux.

On le voit, le renversement de la charge de la preuve ainsi envisagé ne fait pas pour autant disparaître la difficulté de prouver le lien de causalité entre la faute et le dommage. Pratiquement donc, le mécanisme de la responsabilité pour faute, y compris présumée, ne sera pertinent que pour les dommages liés à des systèmes d'intelligence artificielle simples, d'exécution.

2) *La faute du système d'intelligence artificielle « individualisé » : responsabiliser le robot*

Le système de responsabilité pour faute serait-il plus pertinent si l'on impute celle-ci à l'intelligence artificielle elle-même ?

On peut tout d'abord s'étonner du paradoxe qui consiste à chercher la responsabilité d'une entité dont on postule par ailleurs la perfection, l'infailibilité, la supériorité par rapport à l'homme et ses faiblesses, sources de dommages et justifiant sa responsabilité.

Un courant doctrinal assez soutenu porte pourtant cette position. Une résolution du Parlement européen du 16 février 2017¹⁰ en particulier propose de donner une personnalité électronique aux robots autonomes.

Or l'existence d'une telle personnalité juridique suppose la reconnaissance d'une intention propre justifiant l'engagement de sa responsabilité, alors même que nulle unanimité ne se cristallise aujourd'hui chez les scientifiques¹¹ pour affirmer avec certitude que l'intelligence artificielle décide seule, invente sa décision indépendamment d'un esprit humain, d'une intention humaine. Pour nous, le robot n'a pas de libre arbitre ou de volonté propre, il ne règle pas son comportement selon ses propres règles et valeurs (il n'en a pas), il n'est au mieux

10. Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL)), consultable sur : <<http://www.europarl.europa.eu>>.

11. Voir J.-G. Ganascia (2017), *Le Mythe de la Singularité – Faut-il craindre l'intelligence artificielle ?*, Paris : Seuil ; R. Chatila (2016), « Intelligence artificielle et robotique : un état des lieux en perspective avec le droit », *Daloz IP/IT*, n° 6, p. 284.

qu'une interface dépendant d'algorithmes qui déclenchent son action selon une procédure et dans un but qu'un ou des humains lui ont assigné.

Par ailleurs, nous ne croyons pas que donner un tel statut – et par suite une responsabilité juridique – au robot soit d'un quelconque intérêt pour les victimes, d'une part parce que donner un statut à l'intelligence artificielle ne garantit aucunement l'indemnisation de la victime potentielle d'un dommage causé par celle-ci dans la mesure où l'intelligence artificielle n'est pas solvable (elle n'a pas de patrimoine, ou si elle en a un, celui-ci relève en réalité de personnes, physiques ou plus vraisemblablement morales); d'autre part, parce que poser la responsabilité du robot en tant que personne induit la déresponsabilisation du concepteur et du fabricant, ce qui n'est guère incitatif en matière d'exigences de sécurité des systèmes mis sur le marché.

Surtout, l'humain et l'intelligence artificielle ne sont pas interchangeables; l'intelligence artificielle n'intervient que dans un rapport de complémentarité dans lequel il faudra prendre garde à s'assurer de la dissociabilité humain/robot, de la réversibilité du processus d'association. Si l'intelligence artificielle peut remplacer l'humain dans certaines hypothèses pratiques, elle ne saurait le remplacer en droit, et plus particulièrement en droit de la responsabilité. Ses actions ne sont pas détachables des choix de celui qui l'a conçu et/ou de ceux qu'il sert. À cet égard, le droit commun de la responsabilité permet déjà de chercher la responsabilité du gardien de la chose ayant provoqué le dommage.

B) Les incertitudes liées au recours à la responsabilité du fait de la garde

L'applicabilité de la responsabilité du fait de la garde aux intelligences artificielles nous semble fort incertaine. Elle dépend en premier lieu de la qualification de celle-ci, quelle que soit sa forme (corporelle ou incorporelle), comme chose. En l'occurrence, le droit positif évoque bien les choses inanimées, les choses pourvues d'un dynamisme propre et dangereux... Mais ici la difficulté repose sur le fait que la responsabilité est liée non pas à l'interface de l'intelligence artificielle (au produit fini donc), mais plutôt aux technologies qu'elle contient, aux algorithmes: sont-ils des choses (un logiciel virus a été qualifié de bien¹² en jurisprudence)? Ne sont-ils pas plutôt des services?

Il faut ensuite résoudre la question de savoir qui est le gardien de quoi dans le système complexe, sachant que la garde, juridiquement, n'est pas cumulative. On pourrait chercher à déterminer le titulaire de la garde, en établissant une distinction entre la garde de la structure (relevant du concepteur et/ou du

12. Voir notamment Cass. Com., 8 février 1994: Bull. civ. IV n° 56.

fabricant) et la garde du comportement (relevant de l'utilisateur) pour répartir équitablement la charge de la responsabilité liée à un dommage causé par l'intelligence artificielle. Mais saura-t-on attribuer avec suffisamment de certitude la cause du dommage à la structure ou au comportement de celle-ci ? Vraisemblablement pas, eu égard à l'opacité et à la complexité de son fonctionnement.

Par ailleurs, le gardien, par principe, est celui qui dispose du pouvoir d'usage, de contrôle et de direction de la chose¹³. Or s'agissant du fonctionnement d'un système précisément conçu et développé pour son autonomie, ses facultés d'apprentissage et de décision, sa capacité à se passer d'un gardien en somme, on ne peut pas véritablement invoquer un pouvoir de direction et de contrôle de l'utilisateur de l'intelligence artificielle qui, le plus souvent, ne maîtrisera qu'une infime partie de son fonctionnement ; il en va de même et pour les mêmes raisons s'agissant du fabricant.

Quant au concepteur, en première intention on pourrait être tenté de le désigner comme le gardien du système qu'il a conçu (« éduqué » en quelque sorte)... Mais là encore, comment le considérer comme disposant d'un pouvoir de direction alors même qu'il a programmé le système précisément pour qu'il se dirige seul ? Tout au plus pourrait-on considérer qu'il dispose, comme l'utilisateur d'ailleurs, d'un pouvoir de contrôle, de direction et d'usage s'incarnant dans sa possibilité, qui devra toujours être garantie, d'arrêter, de débrancher le système d'intelligence artificielle. Mais même cette orientation nous semble devoir être exclue dans l'hypothèse d'un système développé sous la forme d'une *open source*, qui dilue plus encore la responsabilité.

On le voit, que ce soit sur le fondement de la faute personnelle ou sur celui de la garde, les régimes de droit commun de la responsabilité ne parviennent pas à saisir toute la complexité de l'action impliquant une personne soutenue, complétée par une intelligence artificielle censée être autonome. Ces régimes s'ouvrent sur des impasses, tant en termes d'imputabilité de la faute qu'en termes d'établissement du lien de causalité entre l'action mixte homme/machine et le dommage en résultant. En découle un risque majeur sur le plan de la garantie d'indemnisation des victimes.

Cette situation n'est pas nouvelle pour le droit : la jurisprudence puis la loi ont su concevoir des régimes juridiques spéciaux pour prendre en compte de telles difficultés. Il convient à présent d'analyser leur déclinaison possible à l'hypothèse des dommages causés par des intelligences artificielles autonomes supplantant l'action humaine.

13. Cass. Ch. réunies, *Franck*, 2 déc. 1941, *Bull.* ch. mixte, n° 292.

2. LES PERSPECTIVES OUVERTES PAR DES RÉGIMES SPÉCIAUX DE RESPONSABILITÉ

Deux types de régimes sont invocables ici : ceux qui restent essentiellement fondés sur une logique de responsabilité, d'une part (A) ; ceux qui changent de paradigme en se fondant sur un principe de solidarité, d'autre part (B).

A) La déclinaison des régimes spécifiques de responsabilité existants

Deux régimes nous paraissent particulièrement intéressants : celui de la responsabilité liée à un produit défectueux, et celui de la responsabilité sans faute pour risque.

1) *Le cas particulier du « produit défectueux »*

Si l'on peut établir que le dommage consécutif à une interaction entre l'homme et l'intelligence artificielle est lié à un défaut de sécurité de celle-ci antérieur à sa mise sur le marché, on pourra chercher à engager la responsabilité de son producteur sur le fondement de l'article 1245 du Code civil visant les dommages liés à des produits défectueux au sens de la directive européenne du 25 juillet 1985¹⁴.

La pertinence d'un tel régime, pour les victimes, tient dans le fait que la responsabilité du producteur est alors engagée de plein droit. Toutefois, plusieurs difficultés d'application peuvent être relevées.

Tout d'abord, qui est, dans notre hypothèse, le « producteur » visé par l'article 1245 ? Le(s) concepteur(s) des intelligences artificielles en cause ? Le fabricant du produit fini ? Sans doute faudra-t-il en toute hypothèse retenir une définition élargie du « producteur » afin que la mobilisation d'un tel régime juridique soit efficiente.

En second lieu, comme on a pu le soulever précédemment, est-on véritablement en présence d'un produit ? Si l'on considère l'interface matérielle du système, sans doute, mais qu'en est-il de la technologie embarquée, du simple logiciel ? Il faudra éclaircir ce point et caractériser précisément ce qui sépare le produit du service, dans la mesure où l'on ne saurait établir un régime de responsabilité différent selon que l'intelligence artificielle en cause est incarnée dans un bien matériel ou n'existe que sous la forme d'une chose incorporelle.

Surtout, le régime de responsabilité du fait des produits défectueux n'est pas applicable aux dommages survenus après la mise sur le marché, or l'une des spécificités liées à l'usage des intelligences artificielles tient justement dans

14. Directive n° 85/374/CEE.

le caractère évolutif de leur comportement au fur et à mesure de l'usage, dès lors qu'elles apprennent et s'adaptent à leur environnement, en autonomie. Il faudrait alors élargir l'applicabilité de l'article 1245 au défaut survenant dans le processus d'apprentissage postérieurement à la mise sur le marché, ce qui ne manquerait pas de donner naissance à un contentieux fort complexe relatif à la distinction entre le dommage lié à un tel défaut et celui lié à une mauvaise utilisation.

Par ailleurs, l'article 1245-10 du Code civil exonère de sa responsabilité le producteur qui ne pouvait pas anticiper ou déceler l'existence du défaut dommageable eu égard à « *l'état des connaissances scientifiques et techniques au moment où il a mis le produit en circulation* ». Pour des systèmes d'intelligence artificielle par définition évolutifs, l'exonération pourrait être systématique, sauf à opposer au producteur une disposition prévoyant qu'en matière d'intelligence artificielle avancée, le « défaut » permettant l'application de l'article 1245 tient justement dans le fait d'une conception ou d'une programmation laissant la possibilité au robot d'adopter une conduite dommageable. Cela reviendrait à créer une obligation complémentaire de surveillance relative au risque de développement de l'intelligence artificielle commercialisée. En amont, on pourrait également interpréter largement l'article 1245-3 du Code civil, qui pose qu'un « *produit est défectueux lorsqu'il n'offre pas la sécurité à laquelle on peut légitimement s'attendre* » : s'agissant des intelligences artificielles, ne pourrait-on considérer que ce niveau de sécurité légitimement attendu est un fonctionnement infaillible, qualité qui semble bien fonder le recours à ces systèmes.

Enfin, resurgit dans le cadre de la responsabilité du fait des produits défectueux la difficulté récurrente liée à l'établissement du lien de causalité entre le défaut de l'intelligence artificielle et le dommage causé : la complexité du système impliqué permet difficilement l'établissement de ce lien. Une adaptation de l'article 1245 du Code civil est toutefois envisageable, qui prévoirait une présomption de défectuosité s'agissant de certaines intelligences artificielles particulièrement complexes. Un tel mécanisme de présomption de causalité existe déjà, s'agissant par exemple de la mise sur le marché de médicaments, dès lors que les présomptions sont graves, précises et concordantes¹⁵.

Nonobstant ces aménagements possibles du régime de responsabilité du fait des produits défectueux, il ressort une nouvelle fois de l'analyse que l'autonomie consubstantielle aux intelligences artificielles susceptibles de transformer la personne, qui n'est pas une autonomie « morale » mais une autonomie

15. Voir l'arrêt rendu par la 2^e ch. de la Cour de justice de l'Union européenne le 21 juin 2017, N° C-621/15 (Renvoi préjudiciel – Directive 85/274/CEE).

technique, rend en pratique illusoire la détermination d'une responsabilité individualisée. En effet, elle entraîne inmanquablement une imprévisibilité, non au sens d'un comportement hiératique lié à un dysfonctionnement, mais une imprévisibilité liée d'abord à la complexité invoquée, renforcée par « l'apprentissage » du système qui pourra évoluer, sans pour autant qu'un comportement puisse être qualifié de défectueux ; or pour cette imprévisibilité non défectueuse, ni les règles générales de la responsabilité, ni celles relatives aux produits défectueux¹⁶ ne pourront s'appliquer de façon pertinente pour les raisons sus-évoquées (imputabilité complexe, voire impossible, lien de causalité tout aussi complexe à établir).

On pourra alors se tourner vers les mécanismes historiquement dégagés par la jurisprudence, puis repris dans la loi, de responsabilité sans faute pour risque.

2) *La prise en compte du risque*

Lorsque la complexité de la preuve de la faute expose la victime à une absence d'indemnisation, que la recherche même d'une faute est impropre à rendre compte du fait générateur du dommage, reste la possibilité de s'inspirer du régime de la responsabilité sans faute fondée sur le risque. L'indemnisation repose alors sur une socialisation du risque par la mobilisation des deniers publics, une collectivisation indirecte impliquant l'ensemble du corps social, dans une logique juridique qui reste encore celle de la responsabilité : mettre en présence un dommage, un lien de causalité, une imputabilité, quand bien même elle ne se caractériserait pas par une faute.

Une telle approche de l'indemnisation des victimes à raison de la gravité du dommage subi et surtout du contexte de sa survenance s'est développée en particulier dans le droit de la responsabilité hospitalière à partir de l'arrêt du Conseil d'État *Bianchi* du 9 avril 1993¹⁷, relatif aux dommages graves causés par des actes médicaux à risques. Le juge administratif y pose que « *lorsqu'un acte médical nécessaire au diagnostic ou au traitement du patient présente un risque dont l'existence est connue mais dont la réalisation est exceptionnelle et dont aucune*

16. À cet égard, on lit dans la Résolution du Parlement européen du 16 février 2017 relative aux règles du droit civil sur la robotique, point AI, que « nonobstant le champ d'application de la directive 85/374/CEE [Produits défectueux], le cadre juridique actuellement en vigueur ne suffirait pas à couvrir les dommages causés par la nouvelle génération de robots, puisque celle-ci peut être équipée de capacités d'adaptation et d'apprentissage qui entraînent une certaine part d'imprévisibilité dans leur comportement, étant donné que ces robots tireraient, de manière autonome, des enseignements de leurs expériences, variables de l'un à l'autre, et interagiraient avec leur environnement de manière unique et imprévisible ».

17. CE, Ass., 9 avril 1993, *Bianchi*, n° 69336 : Rec. p. 127.

raison ne permet de penser que le patient y soit particulièrement exposé, la responsabilité du service public hospitalier est engagée si l'exécution de cet acte est la cause directe de dommages sans rapport avec l'état initial du patient comme avec l'évolution prévisible de cet état, et présentant un caractère d'extrême gravité». De la sorte, on institue une responsabilité objective détachée de l'existence d'une faute, fondée sur le risque thérapeutique qui a été pris à l'égard de la victime. Cette jurisprudence a été développée par la loi Kouchner du 4 mars 2002¹⁸, qui requalifie les dommages causés lors de ces actes comme étant des « aléas thérapeutiques » et prévoit leur indemnisation par un fonds d'indemnisation spécifique, à condition qu'ils soient directement imputables à des actes de prévention, de diagnostic ou de soin, qu'ils aient eu pour le patient des conséquences anormales au regard de son état de santé comme de l'évolution prévisible de celui-ci, qu'ils dépassent un certain seuil de gravité. La même approche d'adaptation du droit de la responsabilité à raison de la complexité de survenance du dommage considéré se retrouve à propos des accidents causés par les accidents de la circulation¹⁹. Dans le cadre de ces régimes de responsabilité objective, dès lors que la situation visée par la loi survient (accident impliquant un véhicule/dommages consécutifs à un acte médical nécessaire à risque), la victime est en droit d'obtenir une indemnisation de la part d'un fonds d'indemnisation dédié alimenté par une communauté prédéfinie. Ainsi l'aléa, le risque thérapeutique ou le risque technologique automobile font-ils l'objet d'un régime d'indemnisation spécifique, qui nous semble pouvoir être décliné s'agissant du risque technologique ou algorithmique lié aux intelligences artificielles avancées. En effet, ces régimes innovent en se détachant de la logique de responsabilité, concrètement impraticable, pour faire prévaloir une logique de solidarité permettant à la fois une prise en charge équitable des dommages et une responsabilisation effective des différents acteurs du système complexe.

B) Un nécessaire changement de paradigme : de la responsabilité à la solidarité

Le changement de paradigme, de la responsabilité vers la solidarité, nous paraît devoir être inspirant pour la prise en charge de l'indemnisation des victimes d'intelligences artificielles. En effet, c'est en réalité l'anormalité pour la victime de supporter seule la charge potentielle des progrès technologiques de la société, qui peuvent finalement être définis comme des risques socialement

18. Loi n° 2002-303 du 4 mars 2002 *relative aux droits des malades et à la qualité du système de santé*. V. également l'article L.1142-1 § II du code de la santé publique.

19. Loi n° 85-677 du 5 juillet 1985 *tendant à l'amélioration de la situation des victimes d'accidents de la circulation et à l'accélération des procédures d'indemnisation*.

conditionnés²⁰, qui fonde la recherche d'alternatives aux régimes de responsabilité existants.

Les systèmes d'intelligence artificielle avancés se développent et se développeront parce qu'ils promettent une qualité de vie supérieure, des capacités supérieures, une sécurité supérieure dans les relations sociales.

Or du fait qu'ils impliquent non seulement une chaîne complexe d'intervenants, mais également une imprévisibilité du comportement de la chose ayant participé au dommage, ils engendrent une responsabilité insaisissable, qui ne s'adapte pas aux impératifs classiques de la responsabilité civile ou administrative, exigeant si ce n'est la preuve d'une faute, du moins celle d'un lien de causalité entre le dommage et la situation en cause.

Dès lors, si l'on veut satisfaire le besoin de sécurité ou d'amélioration des conditions de vie que promet ce système, il convient de garantir l'indemnisation des victimes d'accidents (par définition rares) qui pourront encore survenir sans que l'on puisse précisément identifier un responsable, une imputabilité du dommage, du fait qu'ils ne sont pas liés à une action isolée, mais à une pluralité d'actes infimes intimement corrélés dont aucun, seul, ne permettra d'imputer le dommage à son auteur.

Pour ce faire, il faut s'engager dans une démarche pragmatique de prise en charge sociale des risques auxquels chacun est soumis dans l'intérêt de tous. L'indemnisation des victimes sera fondée non pas sur la faute ou l'imputabilité, mais sur l'existence même du dommage, non pas à raison de sa nature, mais à raison de l'intérêt général du moyen de sa survenance, du caractère équitable de l'indemnisation individuelle de la charge potentielle du progrès collectif.

Il s'agit de retenir un principe d'indemnisation plutôt qu'une responsabilité de principe; de ne pas s'arc-bouter sur la recherche d'un responsable, mais de reconnaître l'intérêt social de certaines transformations de la personne en faisant en sorte que l'indemnisation intervienne rapidement et certainement, sans peser sur un seul des acteurs de la chaîne (en particulier la victime) grâce à une prise en charge collective du risque algorithmique/technologique, sans déresponsabiliser pour autant les différents acteurs du système: au lieu d'imposer la recherche d'un lien de causalité, on postulera l'absence de non-causalité entre le dommage subi et l'implication d'une intelligence artificielle complexe.

Dans ce contexte, la prise en charge collective du préjudice se fonde sur une logique de compensation solidaire du dommage que fait courir à chacun la marche du progrès social. Faute d'imputabilité certaine, il faut garantir aux

20. J. Knetsch, *Le droit de la responsabilité et les fonds d'indemnisation*, Thèse, Paris, 2011, p. 280 et suiv.

individus une indemnisation rapide et égalitaire du préjudice supporté dans l'intérêt de tous.

Telle est la fonction du fonds de garantie que l'on propose de mettre en place à l'image de ceux qui existent déjà s'agissant du risque thérapeutique, des accidents impliquant des véhicules terrestres à moteur, mais aussi de certains risques technologiques²¹ : il s'agit, en déclinant la même logique assurantielle, de répartir collectivement la charge du gain social représenté par certaines transformations de la personne permises par les intelligences artificielles, transformation du corps, ou transformation des modes de décision, des agissements. Un tel fonds devra nous semble-t-il revêtir une dimension prospective²², c'est-à-dire anticiper la nécessité d'une indemnisation rapide par la constitution d'un capital établi en amont grâce à la mise en place d'un système d'assurance obligatoire permettant de responsabiliser les différents acteurs impliqués (concepteurs, fabricants, utilisateurs), à l'instar du mécanisme mis en place pour l'Office national d'indemnisation des accidents médicaux²³ (ONIAM).

Ainsi, la seule circonstance de l'implication du système d'intelligence artificielle avancée suffirait par principe à justifier l'intervention du fonds, une intervention à titre exclusif ou à titre subsidiaire, en fonction du degré d'autonomie de l'intelligence artificielle impliquée, et donc de la complexité à établir une responsabilité selon les régimes juridiques du droit commun. C'est ce que suggère déjà en substance la Résolution du Parlement européen²⁴ relative aux règles civiles de la robotique, en posant que la responsabilité « *devrait être proportionnelle au niveau réel d'instructions données au robot et à l'autonomie de celui-ci, de sorte que, plus un robot est autonome, plus sa capacité d'apprentissage est grande, plus sa période de formation a été longue, plus grande devrait être la responsabilité de la personne qui l'a formé* » ; la résolution précise encore que lorsqu'il s'agira de déterminer « *qui est la personne réellement responsable du*

21. Ainsi le Fonds de garantie des assurances obligatoires (FGAO) indemnise en particulier les victimes d'accidents de la route quand le responsable de l'accident est inconnu, insolvable ou n'est pas assuré, sur le fondement d'un devoir de solidarité nationale fondé sur le préambule de la Constitution de 1946 (Voir Cass. 2^e civ., 21 oct. 1987, Bull. civ. II, n° 204). Son champ d'intervention n'a cessé de s'accroître, incluant notamment l'indemnisation de risques technologiques non pris en charge par l'assurance. Il pourrait désormais accueillir, sur le même fondement de solidarité et d'équité, la compensation de l'aléa algorithmique.

22. Jonas Knetsch, *op. cit.*, p. 125 et suiv.

23. Voir la loi n° 2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de soins.

24. Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL)), en particulier § 56.

comportement dommageable du robot, les compétences acquises par un robot au cours de sa formation ne devraient pas être confondues avec les compétences strictement dépendantes de sa capacité à apprendre de manière autonome».

Retenir un tel système de fonds de garantie solidaire pour les dommages les plus graves impliquant l'usage des intelligences artificielles les plus avancées permettrait d'assurer une indemnisation à la fois équitable et raisonnable de la victime : équitable dans la mesure où le dommage résultant d'un système complexe ne permettant pas d'identifier précisément le responsable pourra être indemnisé sur le fondement de la solidarité, mais sans déresponsabiliser totalement les responsables potentiels ; raisonnable car tout préjudice n'est pas automatiquement indemnisé, de sorte qu'on n'instaure pas un frein majeur au développement technologique.

En outre, la création d'un fonds de garantie va dans le sens de l'histoire du droit de la responsabilité : après avoir délaissé largement les régimes de responsabilité pour faute au profit de régimes de responsabilité objective caractérisés par l'absence de faute, la multiplication des fonds d'indemnisation dans des champs de plus en plus variés²⁵ apparaît comme la solution d'une société complexe qui n'admet plus l'absence d'indemnisation des victimes, qui cherche moins un coupable qu'un débiteur solvable.

Cette adaptation du droit de la responsabilité aux technologies les plus avancées nous paraît absolument fondamentale et indispensable, car on ne saurait se contenter de mettre entre les seules mains des concepteurs de systèmes d'intelligence artificielle la responsabilité de décider de ce qui doit se faire ou pas, des garde-fous à installer.

À cet égard, les chartes éthiques qui se multiplient ces dernières années ne suffisent pas, surtout lorsqu'elles sont élaborées par des acteurs privés qui sont à la fois juges et parties. L'enjeu est tel pour les personnes, pour l'Homme, qu'une régulation publique à cet égard est urgente : les intelligences artificielles les plus avancées, celles qui agiront à la place de la personne et selon leur propre « intention », ne sont en pratique pas encore présentes dans la vie sociale, de sorte que l'on peut et doit urgemment engager, en amont des aménagements des régimes de responsabilité, une réflexion sur l'instauration de principes directeurs préventifs permettant de garantir la moindre faillibilité des

25. Voir par exemple le fonds de garantie des victimes du terrorisme, issu de l'article 9 de la loi n° 86-1020 du 9 sept. 1986 relative à la lutte contre le terrorisme ; le fonds d'indemnisation des transfusés et hémophiles, mis en place par l'article 47 de la loi n° 1991-1406 du 31 janvier 1991 portant sur diverses mesures d'ordre social ; le fonds d'indemnisation des victimes de l'amiante, créé par l'article 53 de la loi n° 2000-1257 du 23 décembre 2000 de financement de la sécurité sociale.

systèmes²⁶, et donc de limiter la survenance de dommages technologiques. L'encadrement juridique commence là. On pense tout particulièrement aux principes d'expérimentation préalable obligatoire, d'évaluation et d'audit tout au long de la « vie » de l'intelligence artificielle. On pense également à un principe de transparence, non pas qu'il s'agisse de dévoiler des secrets industriels, mais d'être capable – si toutefois cela s'avère possible²⁷ – d'expliquer le fonctionnement de l'intelligence artificielle, le cheminement des corrélations (plus que des causalités) qui la font agir ou ne pas agir, les principes gouvernant ses choix. En toute hypothèse, le droit doit s'ancrer dans la perspective que face aux intelligences artificielles transformant son corps ou son environnement, l'humain reste aux commandes.

26. Voir notamment : Avis n° 2017/C 288/01 du Comité économique et social européen, « L'intelligence artificielle : les retombées de l'intelligence artificielle pour le marché unique (numérique), la production, la consommation, l'emploi et la société » ; V. également le document de l'IEEE « Ethically Aligned Design : A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems » et en particulier les principes généraux mentionnés dans le document <http://standards.ieee.org/develop/indconn/ec/ead_general_principles.pdf>.

27. Voir G. Berry, « L'Hyperpuissance de l'informatique : Algorithmes, données, machines, réseaux » de Gérard Berry pour les nombreuses mentions relatives aux inquiétudes générées par l'absence de rigueur dans les développements de l'algorithmique des véhicules.

CHAPITRE 6

De la signification éthique des limites

Mathieu Guillermin et Mathieu Lefort

INTRODUCTION

Au cours de la dernière décennie, le domaine de l'intelligence artificielle (IA) a connu un essor impressionnant. De nombreuses percées ont été réalisées, notamment pour le traitement et l'analyse de grands ensembles de données (*data mining* et *big data*), pour l'analyse d'image, la reconnaissance vocale ou encore pour le traitement du langage naturel. Ces progrès sont susceptibles d'induire de profondes transformations de nos sociétés, des personnes et de leur rapport à autrui et au monde. Les nouvelles technologies d'IA influencent notre manière de vivre. Elles modifient par exemple en profondeur notre rapport à l'information à travers les processus de recommandation et de filtrage déployés par les moteurs de recherche ou les réseaux sociaux. Elles interrogent aussi le sens et les contours de notre vie privée. Les algorithmes d'IA deviennent capables, éventuellement par le truchement de robots, d'assurer de plus en plus de tâches complexes. Les assistants vocaux fleurissent sur nos ordinateurs et téléphones intelligents. Les véhicules et les armes autonomes sont en plein développement. Des avancées sont aussi en cours dans le domaine médical avec le développement d'algorithmes performants capables de suggérer des diagnostics (par exemple en dermatologie, cancérologie ou ophtalmologie) dont la fiabilité atteint le niveau des experts¹.

1. A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau et S. Thrun (2017), « Dermatologist-level classification of skin cancer with deep neural networks », *Nature*, 542,

Ces avancées technologiques ne vont pas sans soulever de nombreuses interrogations éthiques. Comment vivra-t-on demain ? Quelles sont les voies de développement les plus intéressantes, et celles à éviter ? Jusqu'où pousser les progrès techniques ? Doit-on seulement prolonger l'humain ou bien tenter de le remplacer par les technologies IA ? Quelle sera la place respective des humains et des IA ou des robots dans l'activité économique de demain ? Un travail est un élément important pour se construire en tant que personne. Comment partagerons-nous les richesses produites dans un tel contexte transformé par l'IA ? Quelle sera la place respective de l'humain et de la machine demain, dans les processus de décision impliquant directement des vies humaines ? Les dilemmes que l'on doit affronter lorsqu'on tente de définir les comportements des véhicules autonomes en cas d'accidents sont redoutables². L'élaboration et la mise en service d'armes autonomes capables de prendre « seules » des décisions stratégiques impliquant des conséquences de vie ou de mort de personnes ou de population génèrent des problèmes éthiques considérables³ (Bonnemains *et al.*, 2018). Quel rôle demain pour les professionnels de la médecine ? Aura-t-on encore besoin de former des médecins ? Le médecin du futur ne sera-t-il plus qu'un « courtier en information » ? L'enjeu éthique de la place de l'humain face aux IA semble encore plus mis en tension lorsque l'on se projette dans le futur. Pour certains commentateurs, le temps est proche où les IA vont devenir autonomes et vont dépasser le niveau de l'intelligence humaine⁴. L'humain serait alors rapidement menacé d'obsolescence et n'aurait d'autre échappatoire que de se transformer technologiquement pour s'émanciper de ses propres limitations.

Ces quelques exemples illustrent bien l'impérieuse nécessité de réaliser une réflexion éthique approfondie remettant en question le développement des nouveaux outils d'IA et leur impact sur la personne humaine. Comment les outils technologiques nous transforment-ils ? Comment modifient-ils ce que signifie être une personne humaine ? L'IA

va-t-elle tuer le travail ? Quelles

1. P. F. F. F., N. Coudray, S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Feny, A.L. Moreira, N. Razavian et A. Tsigirgos (2018), « Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning », *Nature Medicine*, 24, p. 1559-1567 ; J. De Fauw *et al.* (2018), « Clinically applicable deep learning for diagnosis and referral in retinal disease », *Nature Medicine*, 24, p. 1342-1350.
2. Le lecteur pourra consulter à ce sujet le site du MIT (<moralmachine.mit.edu>), qui propose de se confronter à de nombreux dilemmes de ce type (variations autour du fameux dilemme du tramway).
3. V. Bonnemains, C. Saurel et C. Tessier (2018), « Embedded ethics: Some technical and ethical challenges », *Ethics and Information Technology*, 20, p. 41-58.
4. R. Kurzweil (2005), *The singularity is near: When humans transcend biology*, New York : Penguin Books ; L. Alexandre (2017), *La guerre des intelligences : Comment l'intelligence artificielle va révolutionner l'éducation*, Paris : JC Lattès.

transformations sont acceptables ? Doit-on augmenter l'humain ? Faut-il interdire le développement de certains systèmes ou de certains types de technologies ? Se confronter à ces interrogations est un travail titanesque. Il demande de rassembler de nombreux contributeurs (scientifiques, philosophes, mais aussi utilisateurs ou acteurs économiques et politiques, etc.). De plus, il nécessite une contextualisation afin de respecter les particularités des questions éthiques telles qu'elles se déploient *in situ*.

L'ambition de ce texte sera donc modeste. Il s'agira d'interroger la manière dont on peut mobiliser la notion de limite dans ce type d'investigation éthique. Nous tenterons tout particulièrement de réfléchir à ce qui fait qu'une limite devient problématique ou éthiquement significative. Une première réponse peut être qu'une limite devient éthiquement pertinente quand elle tient une place prépondérante pour l'humain, et que les pouvoirs technologiques permettent d'envisager sa transgression. Les biotechnologies fournissent de nombreux exemples de techniques qui mettent en tension ce type de limites (transgénèse, biologie de synthèse, procréation médicalement assistée, etc.). On attend alors dans cette perspective de la réflexion éthique qu'elle se positionne sur l'acceptabilité ou non de la transgression considérée (section 1). Nous verrons dans la section suivante (section 2) que les développements technologiques dans le domaine de l'IA génèrent promesses et anticipations qui font un puissant écho à cette approche de la question de la limite (autonomie des machines, singularité technologique et transhumanisme). Nous tenterons ensuite d'introduire progressivement l'idée que la prise en compte de la limite en éthique sous l'angle de la transgression n'épuise pas la pertinence de cette question de la limite. Notre conviction est qu'une limite peut devenir extrêmement problématique en éthique, non seulement lorsqu'elle est susceptible d'être transgressée, mais aussi (et peut-être de manière plus commune) quand elle risque d'être ignorée. Avec la section 3, nous montrerons tout d'abord l'intérêt de clarifier les limites des pouvoirs technologiques afin de caractériser correctement les possibilités réelles de transgression ouverte par les technologies d'IA. Une telle mobilisation de la question des limites est indispensable pour maintenir la démarche éthique en prise avec la réalité technologique. Elle nous invite en conclusion à prendre nos distances avec la thématique de la transgression pour étudier des limites, dont le dépassement n'est pas en question, mais qui n'en apparaissent pas moins extrêmement pertinentes du point de vue éthique.

1. VERS UNE TRANSGRESSION RADICALE DES LIMITES ?

Certains discours prospectifs sur l'IA prédisent un développement rapide et quasi illimité de cette dernière⁵. Bien que ces approches ne nous semblent pas pertinentes à court ou moyen terme (qui peut prédire le futur des possibles technologiques à très long terme?)⁶, nous tenons quand même à les aborder, car leur pouvoir de fascination est loin d'être négligeable et, par de nombreux aspects, elles entrent en résonance avec, ou adoptent le schéma de la transgression des limites. Au-delà de l'argument classique basé sur l'accroissement exponentiel de la puissance de calcul des microprocesseurs prédit par la Loi de Moore, la renaissance de l'enthousiasme pour l'IA répond aux succès de l'apprentissage automatique dont l'idée remonte au moins à Turing (Turing, 1950). Nous parvenons enfin à mettre en œuvre cet apprentissage machine, en particulier dans sa version profonde (*deep learning*, à base de réseaux de neurones artificiels avec un nombre élevé de couches intermédiaires). Pris dans l'élan généré par ce renouveau, on peut rapidement avoir l'impression d'avoir instauré un développement de l'IA voué à franchir certaines limites significatives.

En premier lieu, l'apprentissage machine permettrait de franchir le cap de l'autonomie dans la machine. En effet, on peut comprendre l'apprentissage automatique comme signifiant qu'il n'est plus besoin de programmer intégralement la machine. On peut seulement donner un problème dont la machine va chercher seule la solution en se paramétrant elle-même. C'est ce type de technique qui a révolutionné l'analyse d'image automatique en 2012⁷, ainsi que le monde du go en 2016⁸. Ces tâches – reconnaître des objets dans des images ou jouer à un niveau expert au jeu de go – étaient reconnues comme particulièrement résistantes aux techniques classiques d'intelligence artificielle, basées sur une programmation explicite, logico-symbolique. Mais par apprentissage automatique, la machine a trouvé elle-même des solutions à ces problèmes en exploitant des apports humains sous forme d'exemples de réalisation des tâches concernées (ensembles de photos pré-étiquetées selon leur contenu, ensembles de parties de go) et/ou sous forme de systèmes de récom-

5. *Ibid.*

6. Comme le mettra en lumière notre discussion de la section suivante, certains éléments de ces discours spéculatifs nous paraissent, si ce n'est invalides, à tout le moins beaucoup trop flous pour permettre une saine réflexion éthique. Les fantasmes que ces ambiguïtés attisent seraient même plutôt de nature à égarer et à entraver la démarche éthique.
7. A. Krizhevsky, I. Sutskever et G.E. Hinton (2012), « ImageNet Classification with Deep Convolutional Neural Networks », In F. Pereira, C.J.C. Burges, L. Bottou et K.Q. Weinberger (dir.) *Advances in Neural Information Processing Systems 25*, Curran Associates Inc., p. 1097-1105.
8. D. Silver *et al.* (2018), « A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play », *Science*, 362, p. 1140-1144.

pense (apprentissage *par renforcement*)⁹. Dans les deux cas, l'humain joue le rôle du professeur qui «éduque» le programme apprenant¹⁰. Mais c'est la machine qui se paramètre seule afin de parvenir à reproduire les exemples fournis et/ou en se laissant guider par un système de récompenses... Une certaine dose d'autonomie donc? En tout cas, une technique qui a permis à Alpha go de battre les meilleurs joueurs au monde de go.

Mais l'approche de l'apprentissage automatique peut être poussée encore plus loin en restreignant l'apport des humains. En particulier, il est parfois possible de réaliser un apprentissage *non supervisé* au cours duquel la machine se paramètre sans l'aide d'exemples de tâches fournis par des humains (ou d'objectifs définis au moyen de systèmes de récompenses). Dans ce type d'apprentissage, l'humain ne définit plus un but précis au moyen d'exemples de tâches à reproduire ou en récompensant certains schémas d'action. Il fournit seulement des objectifs très généraux (formulés mathématiquement au moyen d'une fonction de coût qui ne mobilise que les données d'entrée, sans s'appuyer sur une caractérisation «externe» introduite par l'humain des données de sortie à atteindre). La machine produit alors seule des résultats qui s'avèrent particulièrement féconds. C'est ainsi qu'en 2012-2013, un algorithme capable de détecter les faces de chat et les corps humains a été élaboré à partir d'un entraînement basé non sur des exemples de détection par des humains, mais sur la consigne de compresser les images puis de les reconstruire avec une fidélité maximale (principe de l'autoencodeur¹¹). Toujours dans cette optique de restriction de l'apport des humains, une version améliorée d'Alpha Go a été développée en 2017 en réduisant l'apport de connaissance experte. L'apprentissage a cette fois été réalisé en partant seulement des règles du jeu de go, sans l'apport de parties-exemples jouées par des humains. Alpha Go Zéro, c'est son nom, a largement battu son grand frère Alpha Go¹². La même technique a été employée pour élaborer Alpha Zéro, une version capable de jouer aussi aux échecs, qui a battu Stockfish 8, le programme le plus avancé à ce jour¹³. Limiter l'apport de connaissance experte permet aux algorithmes de maximiser

-
9. Le programme qui a révolutionné le monde du go, Alpha Go, a été élaboré au moyen d'une combinaison de ces techniques d'apprentissage supervisé et par renforcement.
 10. S. Shalev-Shwartz et S. Ben-David (2014), *Understanding Machine Learning: From theory to Algorithms*, Cambridge University Press.
 11. Q. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean et A. Ng (2012), *Building high-level features using large scale unsupervised learning*, Proceedings of the 29th International Conference on Machine Learning.
 12. D. Silver *et al.* (2017), «Mastering the game of Go without human knowledge», *Nature*, 550, p. 354.
 13. Ce type de technique est actuellement appliqué avec des résultats prometteurs à des jeux vidéo de stratégie en temps réel comme *Starcraft II* (voir: <<https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii>>). D. Silver *et al.* (2018), *loc. cit.*

l'exploration de coups originaux et surprenants pour leurs adversaires humains, de stratégies viables et performantes que les humains n'avaient pas encore identifiées comme telles.

Avec la mise en œuvre de ces techniques d'apprentissage machine, les algorithmes d'IA connaissent un nouvel essor impressionnant, voire vertigineux... Les machines atteignent des niveaux de performances humaines dans de nombreuses tâches. Elles dépassent même ce seuil symbolique dans certains cas tels que ceux que nous venons de survoler. Une autre limite apparaît donc menacée (en sus de celle de l'attribution d'une certaine forme d'autonomie à la machine) : l'IA semble en passe de surpasser l'intelligence humaine. On pourrait toutefois objecter que le fait n'est pas nouveau. Cela n'est pas en soi une révolution... On se souvient du «traumatisme» causé par la victoire de Deep(er) Blue contre Kasparov en 1997. Traumatisme dont on s'est relevé en notant, à juste titre, que cette victoire ne signalait pas une défaite de l'intelligence humaine, mais manifestait plus sobrement le fait que l'ordinateur peut calculer plus rapidement que les humains. Il peut donc résoudre plus efficacement des problèmes difficiles, mais formulables mathématiquement (comme les conséquences d'un coup particulier aux échecs)... L'humain restait néanmoins supérieur, et de loin, pour toutes les tâches plus complexes, demandant de l'intuition ou de la sensibilité, les tâches ne pouvant faire l'objet d'une mise en équation et d'un traitement analytique. Le jeu de go était alors souvent invoqué comme illustration de la supériorité de l'intelligence humaine (ce jeu était et demeure beaucoup trop complexe en matière de combinatoire pour qu'une approche analytique puisse fonctionner). L'apprentissage machine a donc permis de contourner cette limitation. L'IA est aussi devenue performante dans des tâches complexes comme la reconnaissance de formes dans des images ou des flux vidéo, la reconnaissance vocale, etc. Autant de tâches que l'on ne sait pourtant pas mettre en équation. Les machines deviennent peu à peu convaincantes dans le traitement du langage naturel. Elles commencent même, nous le mentionnions en introduction, à envahir le domaine du diagnostic médical.

L'apprentissage automatique permettrait donc de dépasser les limites de l'intelligence humaine lorsque les humains ne savent pas traiter analytiquement un problème en le mettant en équation dont on peut calculer la solution dans un temps raisonnable. De plus, certains commentateurs affirment que c'est aussi dans l'autonomie – que confère l'apprentissage automatique aux machines – que réside la supériorité naissante et future de l'IA sur l'intelligence humaine. Un des tenants les plus provocateurs de cette thèse affirme que «le déluge de

données rend la méthode scientifique obsolète¹⁴». L'idée est la suivante : l'humain est obligé de faire des hypothèses pour résoudre des problèmes (notamment pour élaborer des théories scientifiques et les valider). Il injecte aussi inévitablement des biais subjectifs ou socio-culturels¹⁵. Selon les plus enthousiastes comme Chris Anderson, la machine serait en mesure de s'émanciper de ces limitations de l'intelligence humaine grâce à l'autonomie de l'apprentissage automatique. Elle pourrait analyser les données de manière neutre et parfaitement rationnelle, sans s'appuyer sur des choix posés par les humains. L'apprentissage non supervisé serait ainsi à la pointe de l'autonomie et donc de l'émancipation de l'IA vis-à-vis des limites de l'intelligence humaine. En franchissant la limite de l'attribution d'autonomie à la machine, on se rapprocherait du dépassement par l'IA de la limite de l'intelligence humaine.

À partir de ce type de discours sur la puissance de l'IA et de l'apprentissage automatique, la réflexion éthique est canalisée. La question éthique posée dans ce cadre par l'IA est : doit-on transgresser ces limites de l'autonomie et de l'intelligence humaine ? Doit-on créer des IA douées d'autonomie et susceptibles de nous dépasser sur le plan de l'intelligence ? Ainsi orientée, la réflexion éthique se retrouve face à l'horizon de l'obsolescence de l'humain. Les machines, devant autonomes, passent du statut de prolongement technologique de l'humain (pour le meilleur ou pour le pire) au rang d'entités en compétition avec l'humain, et en passe de remporter la compétition sur le plan de l'intelligence.

Cette thématique du dépassement et de l'obsolescence de l'humain est poussée à son paroxysme avec la notion de « singularité technologique », formulée initialement par des auteurs comme Irvin John Good ou Vernor Vinge¹⁶. L'idée de singularité technologique part de l'hypothèse que l'humain parviendrait à créer une IA générale (capable d'assurer toutes les tâches cognitives réalisées par l'homme) qui dépasserait (même insensiblement) l'intelligence humaine. Nous nous retrouverions alors confrontés à une explosion d'intelligence artificielle,

-
14. C. Anderson (2008), « The end of theory: The data deluge makes the scientific method obsolete », *Wired*, 23 juin 2008 (<<https://www.wired.com/2008/06/pb-theory>>). Pour une exposition plus modérée bien qu'enthousiaste de l'apport du *big data* à l'investigation scientifique, voir : T. Hey, S. Tansley et K. Tolle (dir.) (2009), *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Corporation.
 15. T.S. Kuhn (1996), *The Structure of Scientific Revolutions*, 3^e éd., The University of Chicago Press ; A. Bird (2014), « The historical turn in the philosophy of science », In M. Curd et S. Psillos (dir.) *The Routledge Companion to Philosophy of Science* (Second edition), Routledge, p. 79-89 ; P.A. Roth (2014), « The epistemology of science after Quine », In M. Curd et S. Psillos (dir.), *The Routledge Companion to Philosophy of Science* (2^e édition), Routledge, p. 15-26.
 16. A.H. Eden, E. Steinhart, D. Pearce et J.H. Moor (2012), « Singularity Hypotheses : An Overview », In A.H. Eden, J.H. Moor, J.H. Søraker et E. Steinhart (dir.), *Singularity Hypotheses: A Scientific and Philosophical Assessment*, Springer Berlin Heidelberg, p. 1-12.

qui nous échapperait et nous deviendrait, très soudainement, opaque et incompréhensible. En effet, l'IA ainsi créée serait elle-même capable de faire tout ce que les humains peuvent faire, mais de manière plus performante. En particulier, elle serait donc capable de créer une autre IA plus performante qu'elle-même (ou de se mettre à jour pour atteindre ce nouveau palier d'intelligence). Étant donné que cette IA est (par hypothèse) plus intelligente que l'humain, le gain à ce stade serait plus important que le premier écart entre l'humain et cette première IA. Le processus pourrait se reproduire très rapidement (à la vitesse de réécriture du code dans les mémoires informatiques). Même avec un dépassement minime de l'intelligence humaine par la première IA, il se produirait alors un accroissement exponentiel de l'intelligence des machines s'améliorant successivement de génération en génération. Dans un tel scénario, l'humain, limité par sa constitution biologique (durée de vie, reproduction biologique, temps de gestation et d'éducation des enfants, etc.), se verrait rapidement dépassé jusqu'à devenir obsolète, incapable de comprendre sa création. L'ensemble de ce scénario vertigineux repose sur l'hypothèse de la création d'une super-IA (une IA dépassant, même modérément, le stade de l'intelligence humaine). Certains auteurs prédisent l'occurrence d'un tel événement dans un futur relativement proche, notamment en s'appuyant sur la loi de Moore, qui laisse espérer la persistance du taux de croissance exponentiel de la puissance de calcul des microprocesseurs¹⁷. L'horizon de l'obsolescence humaine face à l'essor et à l'émancipation de ses créations technologiques serait donc proche.

Cette perspective de l'obsolescence humaine, en particulier dans sa version extrême associée à la singularité technologique, conduit à différentes interrogations plutôt radicales quant à la nécessité d'imposer, de respecter ou de transgresser certaines limites. Une première réaction relève des courants trans-humanistes¹⁸. Afin de ne pas devenir obsolète, il est nécessaire que l'humain se libère de ses limites biologiques, qu'ils les dépasse ou les transgresse¹⁹. En particulier, il pourrait s'augmenter artificiellement en utilisant des implants neuronaux, dépassant ainsi lui-même l'ancienne limite de l'intelligence humaine. Il pourrait aussi tenter de transférer sa conscience dans un support informatique, se libérant peut-être même ainsi de la limite de la mort.

17. R. Kurzweil (2005), *The singularity is near: When humans transcend biology*, New York: Penguin Books; A.H. Eden, E. Steinhart, D. Pearce et J.H. Moor (2012), « Singularity Hypotheses: An Overview », In A.H. Eden, J.H. Moor, J.H. Søraker et E. Steinhart (dir.), *Singularity Hypotheses: A Scientific and Philosophical Assessment*, Springer Berlin Heidelberg, p. 1-12; L. Alexandre (2017), *La guerre des intelligences: Comment l'intelligence artificielle va révolutionner l'éducation*, Paris: JC Lattès.
18. F. Damour, S. Deprez et D. Doat (dir.) (2018), *Généalogies et nature du transhumanisme: État actuel du débat*, Montréal: Liber.
19. R. Kurzweil (2005), *op. cit.*; L. Alexandre (2017), *op. cit.*

La réponse transhumaniste à la perspective de l'obsolescence serait donc de flouter (voire d'abolir) les frontières entre l'humain et la machine, d'hybrider intelligence humaine et IA, pour profiter, par cette transgression, de l'explosion d'intelligence artificielle de la singularité technologique.

Une réaction alternative possible est de rejeter les possibilités transhumanistes. Deux raisons au moins peuvent motiver ce choix. Premièrement, on peut considérer que la perspective d'hybridation relève du pur fantasme et que l'humain est nécessairement condamné à l'obsolescence dans l'éventualité de l'occurrence de la singularité technologique. Deuxièmement, on peut juger les projets transhumanistes moralement ou éthiquement inacceptables. Pour qui rejetterait ainsi les possibilités transhumanistes tout en admettant l'occurrence dans un futur proche de la singularité technologique, une manière de se prémunir de l'obsolescence de l'humain serait d'imposer des limites au développement de l'IA : bannir tel ou tel type de technologie, comme l'apprentissage machine. Une autre posture possible serait de juger que l'obsolescence de l'humain n'est pas nécessairement un problème, mais que nous devons intégrer des limites éthiques dans les IA afin que leur autonomie ne nous mette pas en danger²⁰.

Toutes ces interrogations autour de l'acceptabilité éthique et la nécessité de préserver et respecter, d'imposer ou de transgresser certaines limites (limites de l'autonomie des machines, de l'intelligence humaine, des capacités cognitives humaines, du corps humain, limites au développement technologique ou au comportement et à l'autonomie des IA, etc.) génèrent de passionnants et indispensables débats, plus ou moins prospectifs et spéculatifs. Elles relèvent d'une dimension centrale de la réflexion en éthique des technologies, une dimension structurée autour de la question de la transgression de la limite. Les développements technologiques ouvrent des possibles susceptibles de conduire à la transgression de telle ou telle limite importante, et il s'agit d'évaluer ces possibilités. Cette approche de la notion de limite en éthique est indispensable. Notre propos ne vise en aucun cas à mettre ce point en doute. Notre but est plutôt de pointer que cette optique de la transgression n'épuise pas la pertinence et la signification de la notion de limite en éthique. Nous souhaitons montrer que les limites n'ont pas nécessairement besoin d'être (effectivement) dépassables pour mériter leur place au sein de la démarche éthique.

20. Dans cette perspective, les trois lois de la robotique d'Asimov sont souvent mises en avant. Il est intéressant de noter qu'une partie du travail d'Asimov vise (à notre sens) à illustrer les limites de ces règles de comportement qui semblent, à première vue, infaillibles.

2. RECONNAÎTRE LES LIMITES DES POUVOIRS TECHNOLOGIQUES

Parmi les limites qui semblent pertinentes pour la réflexion éthique sans pour autant faire l'objet d'une possible transgression, on peut mentionner les limites des pouvoirs technologiques. Nous allons voir que l'identification et la prise en compte de ces dernières ont une influence forte sur le contenu de la réflexion en éthique des technologies. Cette influence est particulièrement importante pour les questions liées à l'IA, domaine technologique qui, en plus d'être largement opaque pour les non-spécialistes, constitue un terreau fertile pour toutes sortes de fantasmes (aussi bien enthousiasmants qu'inquiétants) susceptibles d'entraver la démarche éthique.

Il convient donc de s'interroger : quelles limites sont effectivement mises en question avec les technologies IA (aujourd'hui et dans un futur proche) ? Quelle est cette rupture qu'apporte l'apprentissage automatique ? Quelle est l'autonomie que cette technique confère aux machines ? Dans quel sens risquons-nous de nous retrouver confrontés à un dépassement de l'intelligence humaine par l'IA ? Ces questionnements sont salutaires, car ils nous rappellent que nous ne pouvons nous focaliser exclusivement sur les succès récents de l'IA (succès réels et impressionnants), mais que nous devons aussi développer une compréhension minimale du comment, de ce qu'on peut ou ne peut pas faire... Dans cette perspective, il convient déjà de rappeler que l'emploi commun que nous faisons du terme IA est un abus de langage : *stricto sensu*, l'IA n'est pas un programme, mais une discipline scientifique et un domaine technologique qui visent à étudier la cognition et à écrire des programmes intéressants²¹. Mais, s'il est identifié comme tel, l'abus de langage n'est pas nécessairement problématique. Ce qu'il convient par contre d'affirmer haut et fort, c'est que, jusqu'à preuve convaincante du contraire, un programme ou un algorithme d'IA (une IA pour alléger le discours) est un ensemble d'opérations algorithmiques ou logico-mathématiques qui transforme des données (des données d'entrée en données de sortie) d'une manière intéressante (notamment pour automatiser des processus qui autrement mobiliseraient de l'intelligence humaine).

Cette clarification a déjà le mérite de bien montrer que les performances de l'IA ne dépendent pas uniquement de la puissance de calcul. On peut utiliser le plus puissant des ordinateurs pour exécuter en boucle des instructions triviales (par exemple, fixer en boucle une variable *i* à une valeur constante). Une avalanche de puissance de calcul peut très facilement s'avérer parfaitement inutile.

21. Par « intéressants », on entend notamment des programmes permettant d'automatiser des tâches qui sont reconnues comme demandant de l'intelligence quand elles sont réalisées par des humains (conformément à la définition pionnière de Marvin Minsky).

La puissance de calcul est un élément indispensable, mais encore faut-il savoir en faire quelque chose... Encore faut-il être capable d'écrire un programme intéressant. Pour ce faire, une première approche est la programmation « explicite », l'approche dite logico-symbolique. C'est l'approche intuitive de la programmation... L'humain écrit l'ensemble du programme de transformation des données *à la main*. Cette approche est tout à fait efficace lorsque l'on est capable d'intégralement décrire la tâche à réaliser de manière analytique, lorsque l'on sait la mettre en équation. La programmation logico-symbolique permet l'élaboration de nombreux programmes intéressants. On peut notamment mentionner les programmes *experts*, ainsi nommés car ils modélisent la connaissance d'un expert (historiquement, *Mycin* en médecine et *Dendral* en chimie sont des exemples célèbres). La programmation logico-symbolique permet aussi d'encoder des règles de raisonnement éthique²². Néanmoins, un recours exclusif à la programmation logico-symbolique laisse démuni lorsqu'on tente d'élaborer des programmes pour des tâches ou des problèmes que l'on ne sait pas traiter analytiquement. Il est, par exemple, particulièrement difficile de décrire analytiquement comment il faut transformer des données d'entrées sous forme d'image ou de flux vidéo en données de sortie décrivant les objets présents dans les images. De même, la transformation de données à réaliser pour déterminer les intérêts d'une personne (afin de lui recommander des vidéos, du contenu sur son réseau social préféré, ou encore de l'espace publicitaire ciblé) à partir de son activité sur Internet échappe à toute formalisation purement analytique de son fonctionnement psychologique.

Ce sont ces limites dans nos capacités de formulation mathématique que l'apprentissage automatique peut parfois nous permettre de contourner. L'idée générale est la suivante : à partir d'une définition de l'objectif à atteindre, on peut tenter de paramétrer de manière automatique une architecture algorithmique au sein de laquelle on laisse des paramètres libres²³. Ainsi compris,

22. V. Bonnemains, C. Saurel et C. Tessier (2018), *loc. cit.*

23. Les fameux réseaux de neurones artificiels sont de telles architectures algorithmiques. Ils se composent d'unités de calcul qui effectuent des opérations mathématiques (comme des sommes pondérées et des opérations non linéaires) sur les valeurs qu'on leur donne en entrée. Très flexibles et performants dans de nombreux cas (en particulier lorsqu'on superpose de nombreuses couches de neurones artificiels, dans le cadre de l'apprentissage profond), ils n'épuisent pas, tant s'en faut, les architectures possibles. On peut par exemple mentionner aussi les arbres de décision ou les algorithmes de *clustering* (comme l'algorithme *k-means*). Il convient de noter que le domaine de l'apprentissage automatique est vaste et mobilise de nombreuses techniques différentes. La notion de paramétrage automatique d'architecture algorithmique est adéquate dans de nombreux cas. Néanmoins, certaines formes d'apprentissage machine se conçoivent difficilement sous cet angle. C'est le cas par exemple de l'algorithme *knn* (dit « des *k* plus proches voisins »), qui tire des prédictions à propos d'une nouvelle instance

l'apprentissage automatique relève d'une classe particulière de problèmes mathématiques, les problèmes dit « d'optimisation ». Dans cette démarche, le programmeur va tout de même devoir choisir et écrire *à la main* une architecture d'algorithme avec des paramètres libres (un programme à trou, en quelque sorte). Il écrira ensuite, toujours *à la main*, un autre programme – parfois appelé programme heuristique – qui sera chargé de trouver les meilleurs jeux de paramètres possibles²⁴. Ceci implique que le programmeur devra spécifier ce que veut dire « bon jeu de paramètres ». Avec l'apprentissage supervisé, cette définition est fournie par un ensemble d'exemples d'une tâche déjà réalisée par des humains, par exemple un étiquetage de photos en fonction de leur contenu (on donne alors un ensemble de couples données d'entrée-données de sortie que l'on souhaite être en mesure de reproduire en optimisant l'architecture – ici les photos constituent les données d'entrée et les étiquettes tiennent le rôle des données de sortie). Avec l'apprentissage par renforcement, on vise à résoudre les problèmes non pas en reproduisant les tâches dans leur globalité, mais en trouvant des séquences d'actions qui, mises bout à bout, atteignent le but recherché. On guide donc l'optimisation « pas à pas » en établissant une fonction de récompense. Contrairement à ce que l'on pourrait penser intuitivement, l'apprentissage non supervisé ne devrait pas être conçu comme une sorte de « super » version de l'apprentissage supervisé, avec une machine qui atteint les mêmes objectifs mais avec une plus grande autonomie. En réalité, l'apprentissage non supervisé ne répond pas aux mêmes objectifs que son alter ego supervisé. Il ne vise pas à construire des algorithmes afin de réaliser des tâches précises (spécifiées par des exemples ou des récompenses), afin d'obtenir des données de sorties prédéfinies (diagnostics médicaux, coups à jouer au go ou aux échecs, identification de formes, etc.) en transformant adéquatement des données d'entrée. L'apprentissage non supervisé vise plutôt l'élaboration de représentations compressées des données d'entrée qui préservent l'information

en exploitant directement sa ressemblance avec les exemples pré-donnés. On peut aussi mentionner l'apprentissage développemental qui ne vise pas à résoudre des problèmes prédéfinis, mais plutôt à réaliser des agents apprenants des comportements intéressants. Voir J. Xue, O.L. Georgeon et M. Guillermin (2019), « Causality Reconstruction by an Autonomous Agent », In A.V. Samsonovich (dir.), *Biologically Inspired Cognitive Architectures 2018*, Springer International Publishing, p. 347-354.

24. De manière générale, ce type de choix concernant les architectures ou les modèles (réseau de neurones ou arbre de décision par exemple), leurs spécificités (nombres de paramètres libres, structures, nombres de couches de neurones), ou encore les détails des programmes heuristiques (algorithmes de régression, algorithmes génétiques, optimisation par rétropropagation du gradient...) revient à fixer l'ensemble des *hyperparamètres* (par opposition aux paramètres qui feront l'objet de l'optimisation proprement dite). Certains de ces hyperparamètres peuvent parfois eux-mêmes faire l'objet d'une optimisation par un algorithme spécifiquement élaboré par le programmeur pour cette tâche.

contenue dans ces dernières. Ces représentations peuvent ensuite être employées pour améliorer la réalisation algorithmique de tâches bien définies, soit comme *inputs* pour des algorithmes établis de manière logico-symbolique, soit comme *pre-processing* nourrissant un processus d'apprentissage supervisé (éventuellement avec renforcement).

Ces distinctions entre les différentes formes d'apprentissage machine sont délicates et largement débattues. Néanmoins, l'exposition ci-dessus suffit, nous l'espérons, à rendre évident le fait que l'apprentissage automatique est bien loin de l'idée d'une machine devenant autonome dans un sens radical, ou d'un bébé machine qu'on devrait éduquer. À aucun moment nous ne sortons du domaine de l'élaboration et de la mise en œuvre d'algorithmes qui transforment des données (on passe souvent sous silence cette question de la mise en œuvre qui cache pourtant de considérables défis scientifiques et techniques). La notion de non-supervision ne signifie en aucun cas que la machine fait ce qu'elle veut, qu'elle se débrouille toute seule. Cette notion indique surtout l'absence d'étiquetage des données par l'humain, l'absence de bases d'exemples des tâches à accomplir réalisées par des humains. Ceci est bien naturel, car aucune tâche précise n'est visée. Mais l'humain doit tout de même injecter une intention dans le processus non supervisé (à travers la fonction de coût et aussi, indirectement, à travers le choix des données d'apprentissage et des différents hyperparamètres). L'apprentissage automatique ne revient donc pas à créer une machine douée d'intentionnalité et d'autonomie au sens fort et humain du terme. La machine ne veut pas, n'a pas d'intention²⁵. Elle exécute des instructions algorithmiques, ni plus ni moins²⁶. L'apprentissage automatique ne signifie donc pas non plus que la machine se trouve dotée d'une capacité d'adaptation à de nouveaux problèmes. Qui dit nouveau problème, dit nouvelle fonction d'évaluation à définir ou nouveaux exemples à communiquer. Dans certains cas, une fonction d'évaluation définie pour un problème peut fonctionner pour traiter un autre problème (la logique de l'autoencodage permet de traiter des problèmes de compression de l'information dans différents contextes, et est aussi utile pour la reconnaissance de formes et la classification automatique).

25. J.R. Searle (1980), «Minds, brains, and programs», *Behavioral and Brain Sciences*, 3, p. 417.

26. Notons que ce constat ne signifie pas nécessairement que le programme de l'intelligence artificielle forte est voué à l'échec. D'une part, il n'interdit pas, en lui-même, l'élaboration d'algorithmes capables de jouer le *jeu de l'imitation* à la perfection. D'autre part, il ne met pas non plus automatiquement la spécificité de l'humain à l'abri. Il reste possible de défendre que la cognition humaine elle-même n'est qu'exécution d'algorithmes. Néanmoins, ce constat est salubre dans le sens où il permet de comprendre que l'apport de l'apprentissage automatique dans le débat sur la possibilité de l'IA forte n'est peut-être pas aussi radical qu'il n'y paraît en première lecture. Il ne change en particulier rien à la nécessité de considérer, par exemple, la difficile question des liens entre calcul ou computation et causalité.

Mais au stade où nous en sommes, nous ne savons pas écrire une fonction d'évaluation unique qui résumerait tous les objectifs que nous souhaitons atteindre (une sorte de méta-formulation de tous les problèmes ou d'une grande variété d'entre eux). Cet obstacle est très épais et revient à un problème classique bien connu, le problème du cadre²⁷. Enfin, et peut-être plus prosaïquement, l'apprentissage automatique ne permet en aucun cas au processus d'élaboration de programme de s'émanciper des choix humains. Que l'on soit en apprentissage supervisé ou non, les choix humains restent indispensables : il faut indiquer un but, il faut injecter une intention, il faut choisir une architecture algorithmique et définir les différentes logiques et autres hyperparamètres guidant le processus d'optimisation.

Cet aperçu des réalités technologiques associées à la notion d'IA, et surtout à celle d'apprentissage machine (la technique la plus médiatisée à l'heure actuelle dans le domaine de l'IA), est extrêmement éclairant par rapport à notre réflexion sur le rôle des limites en éthique. Elle nous permet déjà de mieux caractériser la technologie en question et les limites qu'elle est susceptible de nous mettre en capacité de transgresser. Rien n'indique par exemple qu'une transgression soit possible en matière d'attribution d'autonomie au sens fort aux machines. L'apprentissage automatique est d'ailleurs à concevoir comme un complément de la programmation logico-symbolique pure, et non comme une alternative radicale reposant sur l'éducation d'entités devenues autonomes. De même, il y a peu de sens à parler d'une transgression par l'IA de la limite de l'intelligence humaine. Certes, l'humain parvient à écrire des algorithmes susceptibles de traitements de données qui ne pourraient être réalisés *de tête*. Mais cela est déjà vrai de la calculatrice. Étant donné que l'humain, le programmeur, est partout dans son algorithme, même en apprentissage automatique, il n'y a pas réellement compétition entre l'IA et l'intelligence humaine. Les algorithmes d'IA apparaissent bien plutôt comme des outils, des prolongements qui nous permettent de réaliser plus efficacement certaines tâches ou qui nous en rendent de nouvelles accessibles²⁸.

Un premier enseignement éthique est donc que nous ne sommes pas nécessairement, avec l'IA, à l'aube d'une transgression technologique (en matière d'autonomie radicale et de dépassement de, ou de compétition avec, l'intelligence humaine) imposant de remettre en cause des limites ou des

27. M. Shanahan (2016), « The Frame Problem », In E.N. Zalta (dir.), *The Stanford Encyclopedia of Philosophy*, Stanford University.

28. Dans cet esprit, il convient de concevoir la problématique de l'automatisation du travail au moyen d'algorithmes d'IA, non sous l'angle de la compétition entre l'humain et la machine, mais bien plutôt sous l'angle des rapports d'humains à humains (éventuellement comme une compétition dans laquelle certains humains sont dotés d'outils d'IA puissants).

vulnérabilités profondes essentielles à l'humain (notamment des limites biologiques). Une vision clarifiée des limites des pouvoirs technologiques effectivement disponibles nous permet donc d'orienter adéquatement la réflexion éthique. En prenant en considération ces limites, nous nous rendons bien compte que nous ne pouvons focaliser la réflexion éthique (que) sur la thématique (transhumaniste) de la compétition avec la machine, sur celle de l'obsolescence humaine, et donc sur la nécessité de l'augmentation humaine par des moyens technologiques pour dépasser nos limites biologiques ou cognitives. D'autres limites doivent être considérées, et peut-être pas seulement sous l'angle de la transgression. Nous allons y revenir dès la section suivante. Mais avant, il convient d'être parfaitement clair. La thématique de la transgression des limites *est* éthiquement pertinente, et même cruciale. La prise en compte des limites technologiques nous permet néanmoins de mieux cerner les limites que nous sommes susceptibles de transgresser ou non. Une limite franchie avec l'apprentissage automatique est, par définition, celle du paramétrage automatique de nos algorithmes. Ce franchissement est hautement significatif et éthiquement problématique, car il implique que nous ne pouvons valider le paramétrage automatique point par point, paramètre par paramètre (l'optimisation automatique n'aurait alors que peu d'intérêt). Franchir cette limite soulève des problèmes très ardues : comment s'assurer que le code ne va pas générer d'effets inattendus et indésirables ? En cas de défaillance de l'algorithme et du processus qu'il anime, avec des conséquences négatives significatives, qui est responsable²⁹ ? Il s'agit d'un exemple parmi tant d'autres des difficultés soulevées par le dépassement de certaines limites avec les technologies d'intelligence artificielle³⁰.

29. Il est tentant en ces matières d'opposer la clarté de la programmation logico-symbolique à l'opacité des codes générés par apprentissage automatique. Bien que la compréhension et l'évaluation de ces derniers soient délicates, il ne faut pas les considérer comme des « boîtes noires » dont on ne pourrait rien tirer pour alimenter une réflexion sur la question de la responsabilité. En outre, il faut aussi se garder d'idéaliser la programmation logico-symbolique. La plupart des codes sont si complexes (des millions de lignes de code écrites par de nombreux programmeurs différents) qu'il est souvent très difficile de s'y retrouver. Voir J.A. Kroll (2018), « The fallacy of inscrutability », *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376 p.

30. On aurait par exemple aussi pu mentionner le cas des armes létales autonomes et la limite, extrême mais techniquement franchissable, serait la mise en œuvre d'apprentissage automatique pour l'identification et la sélection des cibles.

3. SIGNIFICATION ÉTHIQUE DES LIMITES

Insistons donc encore une fois : la thématique de la transgression des limites est cruciale du point de vue éthique. La thèse développée ici est simplement que cette thématique n'épuise pas la manière dont la notion de limite doit être mobilisée : certaines limites, dont la transgression n'est pas en question, doivent pourtant être reconnues et prises en compte dans la réflexion éthique. Nous venons de voir un exemple à l'appui de cette thèse : la reconnaissance de certaines limites techniques oriente la réflexion éthique en permettant d'identifier plus clairement le pouvoir transformatif de ces techniques et les ruptures qu'elles rendent effectivement possibles. Mais la reconnaissance et la prise en compte des limites (toujours en dehors de la question de leur transgression) pourraient aussi avoir un impact plus direct sur le contenu de la réflexion éthique et sur les prescriptions qui en émergent. Tentons de fournir quelques exemples.

De manière générale, un des objectifs de la réflexion éthique est de caractériser les transformations que peuvent engendrer les avancées technologiques sur les personnes et leur environnement : seront-elles acceptables ? Apporteront-elles une amélioration ? Bien mesurer les forces et les limites des technologies concernées semble de nouveau crucial. Ce constat paraît particulièrement fécond pour analyser les transformations induites par l'IA. Nous l'avons vu, l'idée que l'IA atteint une forme supérieure de rationalité – en s'émancipant des biais et des limitations humaines grâce à l'apprentissage automatique nourri par le *big data* – a acquis ces dernières décennies une impulsion non négligeable. Pris dans ce courant, on pourrait croire en une transformation positive de nos processus de décision par une délégation aux systèmes d'IA, capables de nous surpasser en neutralité et en objectivité. Via l'IA, les décisions scientifiques pourraient s'améliorer. Ce mouvement pourrait aussi concerner les choix financiers, les choix stratégiques militaires, ou encore les diagnostics médicaux et les décisions thérapeutiques... On pourrait aussi croire en une transformation positive sur le plan éthique : un algorithme n'ayant pas d'état d'âme, ne ressentant pas de fatigue ou de baisse de concentration pourrait s'avérer plus neutre et impartial, plus juste, qu'un humain.

La reconnaissance des limites techniques incite pourtant à une évaluation plus modérée des transformations induites. Nous l'avons abordé lors de notre discussion sur la question de l'autonomie des machines et du dépassement de l'intelligence humaine, les progrès apportés au domaine de l'IA par l'apprentissage automatique ne correspondent pas à une purge des choix et de la subjectivité des humains. Les technologies d'IA ne constituent pas une super rationalité indépendante et autonome. Au mieux, elles peuvent efficacement prolonger l'humain lorsqu'elles sont bien construites... Et c'est déjà beaucoup.

Mais rien ne laisse penser qu'elles soient plus ou moins infaillibles que les humains qui les élaborent. La reconnaissance de ce type de limite est cruciale du point de vue éthique : elle alerte contre toute confiance aveugle envers les algorithmes. Elle nous conduit à réfléchir à nos choix techniques, à les superviser et à les évaluer, pour s'assurer qu'ils conduisent à une fiabilité suffisante. À l'opposé de la logique de « boîte noire », la reconnaissance de la réalité et des limites des techniques d'apprentissage automatique nous enjoint à mettre l'accent sur la thématique de l'évaluation de la qualité des algorithmes. Accorder sa confiance à un algorithme, mettons un algorithme de diagnostic médical, ne doit pas se faire de manière aveugle. Le simple fait que ce soit un algorithme issu d'un apprentissage machine n'est pas suffisant (ni même, en soi, un gage de qualité). La qualité des données ayant servi à l'apprentissage, ainsi que la qualité du processus d'apprentissage reposent sur la qualité de choix humains. Et pourtant, cela fonctionne dans une certaine mesure. Comme nous l'avons vu en introduction, des algorithmes parviennent à reproduire les jugements experts, notamment en dermatologie, oncologie ou ophtalmologie³¹. C'est un très bon signe de leur fiabilité. Il est même pensable qu'un algorithme, produisant un jugement divergent par rapport à un expert, conduise ce dernier à réévaluer son jugement. Mais là réside précisément un point crucial. C'est l'expert qui révise ou non son jugement. L'algorithme, lui, se limite à s'exécuter et à transformer les données. C'est l'humain qui peut prendre de la distance par rapport aux calculs effectués pour s'interroger sur leur validité³².

Les mêmes enseignements, et la logique de responsabilisation qui en découle, s'appliquent aussi à la question de l'impartialité morale des algorithmes. Les algorithmes ne sont pas exempts de choix humains, et ces choix peuvent avoir des conséquences éthiques et morales. S'il est vrai qu'une programmation logico-symbolique peut, en vertu de son aspect explicite et transparent (pour qui a accès au code), garantir une limitation des discriminations induites par le code, cet avantage ne s'étend pas nécessairement aux algorithmes générés par apprentissage automatique. Il est par exemple maintenant bien connu que les biais présents dans les bases d'apprentissage (des

31. A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau et S. Thrun (2017), « Dermatologist-level classification of skin cancer with deep neural networks », *op. cit.* ; N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Feny, A.L. Moreira, N. Razavian et A. Tsirigos (2018), « Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning », *op. cit.* ; J. De Fauw *et al.* (2018), « Clinically applicable deep learning for diagnosis and referral in retinal disease », *op. cit.*

32. Porter une attention sérieuse à ces limites techniques permet d'envisager l'encodage de fonctions de sécurité dans les systèmes obtenus par apprentissage automatique, fonctions qui pourraient par exemple conduire le programme à signaler à l'utilisateur qu'un apprentissage complémentaire pourrait être souhaitable.

choix humains implicites et socialement sédimentés) vont être propagés au comportement du code³³. L'apprentissage machine n'apporte aucune garantie de neutralité ou d'impartialité morale. Ignorer cette limite serait particulièrement dangereux en légitimant une déresponsabilisation éthique sous couvert de neutralité algorithmique supposée, alors même que c'est tout le contraire qui serait souhaitable. Corriger les biais dans les processus d'apprentissage est un réel défi, réclamant de solides processus inter et transdisciplinaires.

Étant donné les limites techniques des processus d'apprentissage automatique rappelées ci-dessus, il semble donc crucial que l'humain soit réaffirmé et renforcé dans son rôle d'évaluateur, dans sa capacité à accepter ou à rejeter en conscience les résultats des procédés algorithmiques³⁴. Des transformations positives de nos processus cognitifs (notamment de prise de décision) sont tout à fait envisageables dans la mesure où l'on préserve cette place pour l'humain. L'IA peut conduire à un prolongement de l'intelligence humaine augmentée, mais non à son remplacement. Autant d'enseignements centraux pour la réflexion éthique qui découlent d'une prise en compte des limites techniques. Le même type de raisonnement peut être appliqué à d'autres limites.

On peut par exemple s'interroger sur la valeur épistémologique de la neutralité et de l'émancipation par rapport à tout biais subjectif, culturel ou conventionnel apporté par l'humain (Guillermin, 2019)³⁵. Cette valeur résonne avec, et rend attractives, les promesses (certes illusoire) de super rationalité de l'IA et de l'apprentissage automatique. Elle a notamment été mise en avant par Descartes, puis par Kant, avec l'idéal de l'enquête pure (une procédure d'investigation parfaitement neutre, purgée de tout choix ou de toute contingence, et produisant donc des résultats absolument certains ; voir Williams, 2008). La méthode scientifique est souvent comprise comme une réalisation de cet idéal, en vertu d'un enracinement dans le donné empirique brut et dans la pureté mathématico-logique. L'image et l'attraction sont peut-être puissantes (expliquant sûrement une partie de la fascination pour les promesses de super rationalité algorithmique), elles n'en sont pas moins inadéquates. La philosophie des sciences du XX^e siècle a clairement mis en évidence l'irréductibilité de la contingence et du choix humain au sein des procédures d'investigation scientifique³⁶. Bien plus que de conduire à un rejet de la science comme

33. D. Boyd et K. Crawford (2012), « Critical Questions for Big Data », *Information, Communication & Society*, 15, p. 662-679.

34. J.-M. Besnier (2017), « Comment dire "non" quand les machines triomphent ? », *Esprit*, p. 76.

35. M. Guillermin (2019), « Non-neutralité sans relativisme ? », In L. Brière, M. Lieutenant-Gosselin et F. Piron (dir.), *Et si la recherche scientifique ne pouvait pas être neutre ?*, Édition Science et Bien Commun, Chapitre 15.

36. A. Bird (2014), « The historical turn in the philosophy of science », In M. Curd et S. Psillos (dir.), *The Routledge Companion to Philosophy of Science* (2^e édition), Routledge, p. 79-89;

irrationnelle, la reconnaissance de ces limites épistémologiques encourage plutôt à rejeter l'idéal de rationalité comme neutralité. Une procédure d'investigation rationnelle ne peut exister sans choix humains. Être rationnel ne veut pas dire s'interdire de choisir. L'enjeu est bien plutôt de reconnaître cette indispensabilité du choix, d'admettre le caractère faillible de nos enquêtes et d'endosser la responsabilité qui vient avec : la nécessité de l'évaluation de nos choix. Être rationnel, c'est (notamment) avoir conscience des choix posés, c'est prendre du recul, c'est viser les choix les plus adéquats possible.

On retrouve bien le schéma qui conduit de la reconnaissance et de la prise en compte d'une limite (dont la transgression n'est pas nécessairement en question) à des enjeux et des prescriptions éthiques (c'est-à-dire des enjeux et des prescriptions concernant ce qui doit être, ce que nous devons faire ou non). Pour revenir à la thématique des transformations induites par l'IA, cette conclusion d'épistémologie des sciences peut nous aider à nous préserver des fantasmes de super rationalité algorithmique et autonome. Choix humains et rationalité n'ont pas nécessairement à être opposés. Le développement d'algorithmes ne peut se faire sans choix humains, mais cela n'équivaut pas automatiquement à une perte de rationalité, qu'il faudrait tenter de circonvenir en purgeant les choix en question. Ce constat amène bien plutôt à reconnaître une limite et à en tirer les enseignements éthiques. Le caractère faillible des choix des humains est propagé à leurs productions algorithmiques. Nous pouvons développer de merveilleuses technologies apportant de très positives transformations si nous gardons cette possibilité de l'erreur dans un coin de notre réflexion, comme une injonction à constamment interroger nos artefacts numériques, à les améliorer pour en tirer le meilleur.

Pour terminer cette section, nous pouvons, à titre exploratoire, considérer une autre limite dont la prise en compte pourrait aussi s'avérer pertinente pour la réflexion éthique sur les transformations induites par les technologies d'IA. L'être humain est limité, dans le sens où il présente une certaine fragilité ou vulnérabilité. Il peut être blessé et souffrir (aussi bien physiquement que moralement). Cette vulnérabilité est une limite dans ce que l'humain peut subir et endurer. Parce qu'il peut être affecté (« affectible » en anglais), l'humain peut rompre. Mais cette vulnérabilité est aussi dans une certaine mesure constitutive. Elle est ouverture au monde et condition de possibilité de la vie et de la personne³⁷. La vulnérabilité humaine constitue une limite dont la signification est complexe. Bien comprendre et reconnaître cette limite pourrait éclairer et

P.A. Roth (2014), « The epistemology of science after Quine », In M. Curd et S. Psillos (dir.), *The Routledge Companion to Philosophy of Science* (2^e édition), Routledge, p. 15-26.

37. T. Magnin (2015), « Vulnerability at the Heart of the Ethical Implications of New Biotechnologies », *Human and Social Studies*, 4, p. 13-25.

informer l'évaluation éthique de certaines transformations rendues possibles par les technologies IA. On peut par exemple penser aux armes autonomes, en particulier aux systèmes d'armes létales autonomes (ou SALA). Ces systèmes pourraient permettre de minimiser les traumatismes subis par les soldats ou de prévenir les accidents ou les problèmes attribuables aux émotions négatives ressenties par les acteurs sur le terrain. Néanmoins, la prise en compte complète de cette limite qu'est la vulnérabilité humaine nous permet de comprendre que recourir à des armes autonomes pourrait aussi signifier le retrait, du terrain opérationnel, de la capacité de l'humain à entrer dans un authentique contact avec autrui, à éprouver un sentiment d'appartenance à la communauté des vivants malgré la situation de combat. En tentant de préserver l'humain des expériences traumatisantes, on risque aussi de s'interdire de bénéficier de sa capacité à être en prise avec les conséquences de ses actes et de ses choix, de son aptitude à se sentir responsable. Dans un registre peut-être moins dramatique, on peut aussi penser aux transformations induites par le déploiement des algorithmes de recommandation ou de filtrage qui conditionnent radicalement notre expérience de navigation sur Internet et sur les réseaux sociaux. La tentation est grande en ce domaine de ne fournir à l'utilisateur (ou, pour l'utilisateur, de ne souhaiter consulter) que des contenus qui correspondent à ses préférences et à ses intérêts. Parce qu'il peut être affecté et vulnérable, l'humain peut être énervé ou perturbé par certaines idées. L'expérience peut être difficile et même, conséquence dramatique, le conduire à s'éloigner pour un temps de son réseau social ou de son moteur de recherche préféré. Les technologies d'IA permettent de modéliser un utilisateur et ses préférences afin d'adapter le contenu qui lui est proposé. L'utilisateur peut ainsi être préservé des images ou des idées qui pourraient lui déplaire, le déstabiliser, voire le choquer. On peut néanmoins se demander si cette logique de protection est souhaitable. Une prise en compte adéquate de la limite de la vulnérabilité peut nous pousser à une certaine retenue en la matière. Bien que potentiellement désagréable ou traumatisante, l'expérience de l'altérité (par rapport à ses préférences, et peut-être encore plus par rapport à une modélisation automatique des préférences) pourrait bien se révéler indispensable pour se constituer pleinement en tant qu'être humain et en tant que citoyen.

On retrouve peut-être dans cette discussion exploratoire le schéma dont nous avons tenté de mettre en évidence la pertinence pour la réflexion éthique sur les transformations induites par l'IA. Même s'il n'est pas question, avec les technologies algorithmiques, de rendre l'humain invulnérable ou impossible à affecter, une compréhension approfondie de la limite de la vulnérabilité nous permet de comprendre certains enjeux tels que la préservation de la possibilité pour une personne de faire des expériences décisives pour apprivoiser sa propre vulnérabilité, cette vulnérabilité qui fait d'elle un être humain authentique.

Une attention accrue à cette limite de la vulnérabilité nous pousse aussi à réfléchir à l'importance de maintenir une présence authentiquement humaine dans certaines situations (nous avons parlé de situations impliquant le recours à la force létale, mais nous aurions aussi pu évoquer le rapport aux malades ou aux aînés).

CONCLUSION: UNE ÉTHIQUE DE LA LIMITE EN QUÊTE DE PERFORMANCE GLOBALE

En guise de conclusion, tentons une synthèse en deux temps. Nous nous sommes interrogés sur ce qui rend une limite pertinente pour l'éthique des nouvelles technologies. S'il est clair qu'une limite peut devenir problématique en éthique lorsque certaines transformations générées par les évolutions techniques en induisent la possible transgression, nous nous sommes demandés si cette optique de la transgression épuisait ce qu'il pouvait y avoir à dire à propos des limites en éthique des technologies. Nous avons défendu une réponse négative en tentant de mettre en évidence le rôle de l'approfondissement de notre compréhension de limites dont la transgression n'est pas en question. Nous avons vu l'importance de la prise en compte de la réalité effective des technologies d'IA, avec leurs forces mais aussi avec leurs limites. Nous avons aussi insisté sur la pertinence de la notion de vulnérabilité comme limite intrinsèque et constitutive de l'humain. Reconnaître et explorer ces limites, non comme des frontières dont le franchissement est problématique, mais comme des éléments en eux-mêmes significatifs pour les problématiques considérées, permet d'enrichir la réflexion sur nos choix techniques et leurs conséquences éthiques et sociétales. Ce type d'approche nous a permis de considérer les technologies d'IA non comme donnant naissance à des entités autonomes entrant en compétition avec l'humain, mais bien plutôt comme des prolongements de ce dernier, comme des outils porteurs d'un pouvoir de transformation que nous avons la responsabilité de piloter en conscience afin d'avancer vers plus d'éthique et d'humain.

Il est possible de reformuler le même message en s'interrogeant sur ce que signifie conduire une démarche éthique. Le but de la démarche éthique est parfois de caractériser les possibles transgressions d'une limite donnée, afin de déterminer si cette transgression est acceptable ou non. Ce rôle est crucial. Il est parfois indispensable de poser des limites à ne pas dépasser. Notre propos dans cette contribution a été de suggérer que le sens et la motivation de la démarche éthique ne se posent pas exclusivement en ces termes, à travers le filtre de la transgression. La démarche éthique ne peut se limiter à un rôle de censeur. Nous nous devons, en plus, de prendre à bras le corps nos responsabilités dans la réflexion évaluative sur nos choix technologiques. Dans cette

optique, certaines limites peuvent être convoquées, non seulement pour restreindre et freiner, mais aussi (ou peut-être plus généralement) pour aider à identifier et à sélectionner les chemins de développement technologique les plus souhaitables. En ce sens, le but de la démarche éthique peut se concevoir comme intégrée à la quête de performance par l'innovation technologique, avec la charge d'élargir cette notion de performance (notamment à partir de l'identification des limites pertinentes), de contribuer à forger une notion de performance globale, traversée et infusée par l'exigence de l'épanouissement humain.

CHAPITRE 7

Quelle(s) forme(s) de gouvernance juridique pour l'encadrement de l'IA et de la robotique interactive ?

Charles-Étienne Daniel¹

INTRODUCTION

Depuis les récentes avancées qu'ont connu les techniques d'apprentissage machine, les domaines de l'intelligence artificielle (IA) et de la robotique se sont vus propulsés sous les feux de la rampe médiatique². Les récents progrès réalisés en matière d'apprentissage profond et de réseaux neuronaux artificiels participent à l'essor du développement de ces technologies, lesquelles proposent désormais d'amorcer une véritable transformation en profondeur de l'humain et du milieu social qui l'entoure. Un bref constat permet d'apercevoir l'ampleur des bouleversements tant réels que potentiels qui affecteront de manière positive ou négative la personne humaine. D'une part, l'intégration de dispositifs connectés tels que des puces, des prothèses et des implants de toutes sortes lui permettent

-
1. L'auteur tient à remercier chaleureusement Georges A. Legault pour ses précieux conseils et commentaires.
 2. G. Lewis-Kraus (14 décembre 2016), «The Great A.I. Awakening», *The New York Times Magazine*, <<https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html>>.

déjà de modifier certains sens comme l'ouïe et la vision, tout en promettant d'améliorer sensiblement ses capacités cognitives. D'autre part, à plus large échelle cette fois, l'intégration de différents systèmes d'intelligence artificielle à titre d'aide décisionnelle dans une foule d'applications modifiera de plus en plus nos relations interpersonnelles basées jusqu'à présent sur la liberté décisionnelle. Pour l'heure, il est difficile de préciser l'étendue exacte des conséquences que tous les dispositifs issus de l'Internet des objets, de la robotique interactive et des systèmes d'IA entraîneront, tant sur le plan individuel que collectif.

Néanmoins, la perspective d'une transformation tous azimuts de la personne humaine et de la société soulève une certaine inquiétude par rapport aux conséquences anticipées des technologies intégrant sous une forme ou une autre la robotique et l'intelligence artificielle. De quelle manière nos rapports humains seront-ils affectés par ces technologies émergentes? Quelle considération accorderons-nous aux humains transformés par l'incorporation de dispositifs technologiques? Sont-ils ou seront-ils avantagés ou désavantagés par rapport aux humains « biologiquement conventionnels »? Quels emplois sont ou seront menacés par l'intégration croissante de ces technologies? Quelle place occuperont les machines dans notre société et quelles en seront les conséquences?

À titre de technologies de rupture, les développements de la robotique et de l'IA promettent ainsi plusieurs bénéfices, tout en étant susceptibles d'entraîner différents risques. La grande question qui se pose à l'aube de la diffusion croissante de ces nouvelles technologies dans la société est donc la suivante: comment bénéficier des avantages qu'elles nous offrent, sans subir pour autant les conséquences adverses qu'elles pourraient entraîner? Cette question s'est également retrouvée dans le contexte des développements de la génétique, des biotechnologies et des nanotechnologies. Pour chacune de ces technologies, la prise en compte des bénéfices et des risques soulevés par leur développement interpelle nécessairement les modalités de leur encadrement normatif: que voulons-nous encadrer et comment pouvons-nous le faire?

Le débat national public français sur les nanotechnologies de 2009-2010 représente à cet égard une illustration du défi lié à la mise en place d'un encadrement dédié au développement des technologies émergentes. Par cet exercice de démocratie participative, le gouvernement français désirait alors informer et consulter la population au sujet des différents enjeux soulevés par le développement des nanotechnologies³. Les autorités publiques souhaitaient

3. Commission particulière du débat public Nanotechnologies, « Débat public sur les options générales en matière de développement et de régulation des nanotechnologies », <http://cpdp.debatpublic.fr/cdpd-nano/debat/debat_public.html>.

également déterminer grâce au débat public les orientations de l'action étatique en matière d'élaboration réglementaire. Toutefois, plusieurs séances publiques ont été fortement perturbées et certaines annulées par l'intervention du groupe d'opposition aux nanotechnologies Pièces et Main d'Œuvre. En refusant le cadre de discussion proposé pour le débat public, ce groupe a surtout fondé ses reproches sur le fait que les décisions étaient prises d'avance, qu'il s'agissait d'un exercice de promotion publique, et que ce « faux débat » n'était pas légitime⁴. Les discussions tenues lors des séances publiques ont également fait l'objet d'aspects techniques complexes et vastes, ce qui a entraîné de sérieux défis dans la compréhension publique des enjeux scientifiques et la délimitation des questions débattues par le public consulté. De l'avis même du président de la commission particulière chargée d'organiser le débat public sur les nanotechnologies, l'exercice n'a pas donné les effets escomptés et la participation publique a été somme toute limitée⁵.

Le débat public français sur les nanotechnologies a surtout permis de mettre en lumière deux visions de l'encadrement normatif qu'ont souhaité les différents acteurs ayant contribué aux échanges⁶. D'une part, plusieurs acteurs issus du gouvernement et du secteur privé ont prôné un développement technologique permettant de favoriser l'innovation et les retombées économiques. Dans cette optique, la réglementation est souvent considérée comme un obstacle qu'il convient d'adapter au contexte du progrès technologique, en limitant les interventions contraignantes et excessives de l'État. D'autre part, des acteurs issus des domaines social et environnemental ont exprimé leurs inquiétudes liées à une perte de contrôle des risques potentiels soulevés par le développement des nanotechnologies. C'est pour cette raison que plusieurs d'entre eux ont invoqué le principe de précaution pour justifier la prise de mesures destinées à limiter ces risques et l'imposition d'un moratoire pour prévenir toute conséquence catastrophique.

Nous retrouvons dans la teneur des réflexions et des débats sur l'essor de la robotique et de l'IA les mêmes interrogations qui ont accompagné les développements des autres technologies émergentes quant à la capacité des règles

-
4. Voir à ce sujet Pièces et Main d'Œuvre (2 décembre 2009), « Récit du sabotage de la réunion CNDP Nano de Grenoble et tract : "Vous aussi organisez un débat pipeau!" », <http://www.piecesetmaindoeuvre.com/spip.php?page=resume&id_article=64>.
 5. P. Deslandes (9 avril 2010), *Bilan du débat public sur le développement et la régulation des nanotechnologies*, Président de la Commission nationale du débat public, p. 2, <http://cpdp.debatpublic.fr/cpdp-nano/_script/ntsp-document-file_download7d8e.pdf?document_id=503&document_file_id=761>.
 6. C.-É. Daniel, G.A. Legault et L. Bernier (2015), « La régulation des nanotechnologies, le débat national français et le dialogue social : nanomonde, grandes attentes normatives? », *Lex Electronica*, vol. 20, n° 1, p. 93-125.

juridiques à prodiguer un encadrement normatif adéquat pour prendre en compte les impacts soulevés. Comment bénéficier des avantages que ces technologies sont ou seront à même de nous procurer, sans subir pour autant leurs conséquences adverses? Que doit-on favoriser, limiter ou interdire – et sous quelles conditions – pour que le développement technologique de la robotique et de l’IA soit socialement acceptable? Le refus des OGM en France a déjà démontré que l’opinion publique ne peut plus être ignorée ou considérée comme implicitement acquise lorsque de nouveaux produits issus du développement technologique sont commercialisés à plus large échelle, sous peine d’un rejet collectif⁷. Considérant l’ampleur des bénéfices et des risques soulevés par la robotique et l’IA, le défi d’une gouvernance de ces deux technologies se pose avec une certaine acuité.

La présente contribution vise à mieux éclairer de quelle manière le droit peut être utilisé comme outil (ou mode) de gouvernance pour encadrer les impacts soulevés par le développement de l’IA et de la robotique. Plus précisément, cette mise en lumière s’effectuera en deux temps. En premier lieu, nous examinerons dans la partie I le contexte économique et l’importance accordée à l’innovation au sein desquels s’inscrit le développement de l’IA et de la robotique. Cela nous permettra de mieux déterminer par la suite la nature des défis liés à la mise en place d’un encadrement du développement technologique. En second lieu, nous analyserons au cours de la partie II de quelle manière la gouvernance par le droit offre une prise en compte des impacts soulevés par le développement de l’IA et de la robotique. L’analyse des formes que peut adopter le droit pour encadrer le développement de l’IA nous permettra d’esquisser un bref portrait des dynamiques à l’œuvre dans la gouvernance juridique qui se profile à l’aube d’une diffusion croissante de ces technologies.

PARTIE I – L’INNOVATION DANS LE DOMAINE DE L’IA COMME MOTEUR ÉCONOMIQUE ET LES DÉFIS DE SON ENCADREMENT

Section 1. L’importance de l’innovation dans le développement économique

Les retombées économiques issues du développement de l’intelligence artificielle et de ses diverses applications sont vertigineuses. Selon un rapport d’analyse réalisé en 2017 par la firme PwC, d’ici 2030, l’intelligence artificielle pourrait contribuer à l’économie globale planétaire jusqu’à concurrence de 15

7. Voir à titre illustratif R. Jules (25 juillet 2014), « Ce que nous disent les OGM de la France », *La Tribune*, <<https://www.latribune.fr/opinions/tribunes/20140725trib000841661/ce-que-nous-disent-les-ogm-de-la-france.html>>.

700 milliards de dollars américains, ce qui représente une somme plus importante que le PIB actuel combiné de la Chine et des États-Unis⁸. Il n'est donc pas surprenant de constater que, dans les dernières années, plusieurs pays se sont lancés dans la course aux investissements publics, de manière à se positionner avantageusement pour profiter des nombreux bénéfices économiques. La Chine et l'Amérique du Nord représentent les joueurs les plus importants dans cette course, chacun d'entre eux étant respectivement susceptible de profiter de 7 000 et de 3 700 milliards de dollars américains⁹. Un rapport réalisé en 2016 par la firme Accenture et intitulé *Why Artificial Intelligence is the Future of Growth* souligne que les entreprises appliquant avec succès l'IA pourraient hausser leur rentabilité moyenne de 38 % d'ici 2035, représentant ainsi un fort potentiel de croissance¹⁰. De surcroît, l'IA est aussi susceptible de doubler les taux de croissance économique annuelle d'ici 2035 tout en haussant la productivité de la main-d'œuvre jusqu'à 40 % en changeant de manière fondamentale les méthodes de travail¹¹. Pour Mark Purdy, directeur général d'Accenture Research, « en optimisant les processus avec l'automatisation intelligente, en appuyant la main-d'œuvre humaine et le capital physique, et en propulsant l'innovation, l'IA peut entraîner une rentabilité et une croissance économique exceptionnelle¹² ».

Compte tenu de l'ampleur des bénéfices attendus, le secteur de l'intelligence artificielle représente le fer de lance de plusieurs stratégies nationales en matière d'innovation technologique afin de stimuler et de favoriser leur développement économique. Par exemple, c'est pour s'imposer comme le « plus important écosystème d'intelligence artificielle en Amérique du Nord »¹³ que le Canada et le Québec ont investi près de 2,5 milliards de dollars canadiens depuis 2017 en infrastructures techniques comme des centres de données et de réseaux sans fil plus performants, de même qu'en infrastructure

8. PricewaterhouseCoopers (2017). *Sizing the prize. What's the real value of AI for your business and how can you capitalize?*, p. 3, <<https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>>.

9. *Ibid.*

10. M. Purdy et P. Daugherty (2016), *Why Artificial Intelligence is the Future of Growth*, Accenture Institute for High Performance, p. 3, <https://www.accenture.com/t20170524T055435_w_w_/ca-en/_acnmedia/PDF-52/Accenture-Why-AI-is-the-Future-of-Growth.pdf>.

11. *Ibid.*

12. Accenture (22 juin 2017), « Selon un rapport d'Accenture, l'intelligence artificielle a le potentiel d'accroître la rentabilité des entreprises dans 16 secteurs d'une moyenne de 38 pour cent, d'ici 2035 », <<https://www.accenture.com/ca-fr/artificial-intelligence-index>>.

13. Conseil consultatif sur l'économie et l'innovation (Septembre 2017), *Agir ensemble. Pour un Québec innovant, inclusif et prospère*, p. 88, <http://s3.amazonaws.com/medias.ccei.quebec/report/CCEI_rapport_integrateur.pdf>.

« de connaissance » comme des laboratoires de recherche ou des incubateurs d'entreprises¹⁴. La raison pour laquelle de tels investissements ont été réalisés, plus particulièrement au Québec, est d'ailleurs mise de l'avant par le Conseil consultatif sur l'économie et l'innovation :

Pour la première fois dans notre histoire récente, le Québec a la chance d'être là, au moment de la naissance d'un extraordinaire filon de savoir et de développement économique. Ici, au Québec, de véritables percées ont été réalisées dans le domaine précis de l'apprentissage profond (*deep learning*). L'apprentissage profond permet aux ordinateurs d'apprendre comme des humains. Il relègue la programmation traditionnelle par éventualités (*if, then*) au passé en donnant aux ordinateurs la capacité de faire des liens non linéaires, comme le cerveau, pour saisir un contexte et améliorer de fois en fois leur « compréhension » d'une situation. Ce créneau de l'intelligence artificielle ouvre la voie à un nombre incalculable d'applications allant de la voiture autonome, au conseiller automatisé, à la reconnaissance vocale, à la lecture par ordinateur de radiographies avec plus de précision que l'œil humain, etc.¹⁵.

Ainsi, l'intelligence artificielle n'est pas simplement un secteur d'activité prometteur : il s'agit littéralement d'une nouvelle ressource susceptible d'affecter *l'ensemble des secteurs économiques*. Le financement public québécois s'inscrit par ailleurs dans un contexte où les montants investis par le secteur privé sont tout autant mirobolants. Outre Londres et la Silicon Valley, c'est à Montréal que Google a annoncé l'ouverture de son centre de recherche¹⁶. Microsoft a pour sa part acheté la société montréalaise Maluuba¹⁷ et investi 7 millions de dollars dans la recherche universitaire¹⁸, alors que Facebook a fondé le FAIR (*Facebook AI Research Lab*) en partenariat avec l'Université McGill¹⁹. La situation québécoise n'est qu'une illustration du ballet d'investissements publics et

14. K. Rettino-Parazelli (26 mars 2019), « Des investissements publics en IA sans retombées garanties », *Le Devoir*, <<https://www.ledevoir.com/economie/550654/intelligence-artificielle>>.

15. Conseil consultatif sur l'économie et l'innovation, *op. cit.*, p. 88.

16. K. Benessaïeh (2 février 2018), « Google ouvre son centre de données à Montréal », *La Presse*, <<https://www.lapresse.ca/affaires/economie/quebec/201802/02/01-5152362-google-ouvre-son-centre-de-donnees-a-montreal.php>>.

17. H. Shum (13 janvier 2017), « Microsoft acquies deep learning startup Maluuba; AI pioneer Yoshua Bengio to have advisory role », *Microsoft Blog*, <<https://blogs.microsoft.com/blog/2017/01/13/microsoft-acquires-deep-learning-startup-maluuba-ai-pioneer-yoshua-bengio-advisory-role/>>.

18. Conseil consultatif sur l'économie et l'innovation, *Agir ensemble. Pour un Québec innovant, inclusif et prospère*, *op. cit.*, p. 88.

19. K. Gombay (15 septembre 2017), « McGill researcher to head Facebook's new Montreal AI Lab », *McGill.ca*, <<https://www.mcgill.ca/newsroom/channels/news/mcgill-researcher-head-facebooks-new-montreal-ai-lab-270309>>.

privés que l'on retrouve à différents endroits et qui stimule l'engouement (*hype*) collectif relativement aux promesses de l'IA.

Ce contexte permet surtout de démontrer l'importance de l'innovation pour stimuler le développement économique. Comme d'autres technologies liées aux sciences de la vie et au génie génétique, à l'industrie chimique et pétrochimique ou encore à l'électronique, l'essor de l'IA est présenté comme un moteur incontournable contribuant à l'économie des États et, incidemment, à l'accroissement du progrès social. La diffusion des applications résultant de l'innovation technologique ne s'effectue toutefois pas sans heurts. Les crises publiques des dernières années liées à certaines conséquences néfastes entraînées par les biotechnologies (le procès de Monsanto sur le glyphosate), le nucléaire (les accidents de Fukushima et Tchernobyl) ou encore le numérique (la faille de sécurité Heartbleed et le scandale de Cambridge Analytica) illustrent bien que l'innovation comporte différents risques de natures variées. Tout l'enjeu du développement technologique comme moteur du progrès social se retrouve donc dans la capacité qu'ont les États à baliser les risques considérés comme acceptables, de manière à maximiser les retombées positives.

Section 2. Les défis de l'encadrement du développement économique et technologique de l'IA

L'encadrement des risques qu'engendre l'IA et ses différentes applications comme la robotique n'est cependant pas une tâche facile pour les autorités publiques. À quel moment doit-on intervenir et comment peut-on le faire? Comme l'a démontré David Collingridge dans son ouvrage *The Social Control of Technology*²⁰, les autorités chargées d'élaborer les réglementations se retrouvent confrontées à un dilemme. D'une part, il leur est difficile d'intervenir tôt dans le développement d'une technologie. En effet, les risques présentés par la technologie ou par un dispositif en particulier peuvent ne pas être bien compris et connus. Face à toute l'importance accordée à la liberté de commerce et d'innovation, les autorités publiques doivent se fonder sur des preuves scientifiques crédibles et solides pour adopter une réglementation limitant les risques présentés par une pratique ou par un produit commercialisé. Le fait d'amasser la documentation scientifique pertinente pour baliser le risque en le limitant ou en l'interdisant peut requérir un certain laps de temps, ce qui limite l'action précoce de l'État. D'autre part, il peut être difficile pour les autorités publiques d'intervenir une fois la pratique ou le produit répandu dans la société, puisqu'elle ou il sera plus susceptible d'être profondément ancré dans les mœurs

20. New York: St. Martin's Press, 1980.

sociales²¹. Internet représente ici une illustration du double défi exposé par Collingridge. Lorsque la technologie était à ses balbutiements, il était difficile de prévoir les différentes conséquences néfastes que le monde du numérique pouvait entraîner en matière de cyberdépendance, de discrimination, d'intimidation, etc. À ce jour, l'ampleur de la diffusion des services procurés par Internet a complètement transformé nos sociétés, et ceux-ci sont profondément ancrés dans nos pratiques sociales. Il n'est donc pas surprenant de voir hésiter les autorités publiques à réglementer l'utilisation de Facebook et des autres médias sociaux, tant ces derniers sont partie intégrante de notre quotidien²².

Cette double difficulté pour les autorités publiques d'encadrer le développement technologique a fait l'objet de différentes tentatives pour mieux assurer un encadrement juridique approprié. Parmi celles-ci, nous retrouvons l'appel au principe de précaution. Énoncé pour la première fois dans la *Déclaration de Rio sur le développement et l'environnement*²³ de 1992, ce principe stipule que « [p]our protéger l'environnement, des mesures de précaution doivent être largement appliquées par les États selon leurs capacités. En cas de risque de dommages graves ou irréversibles, l'absence de certitude scientifique absolue ne doit pas servir de prétexte pour remettre à plus tard l'adoption de mesures effectives visant à prévenir la dégradation de l'environnement ». Reformulé autrement, le principe de précaution incite les États à ne pas attendre l'obtention de preuves scientifiques absolues lorsque se profile le risque d'un dommage grave ou irréversible pour l'environnement. De cette manière, l'objectif du principe de précaution consiste surtout à interpeller les gouvernements en vue d'anticiper les dommages causés à l'environnement et à prendre les mesures qui s'imposent en vue de les prévenir.

Or, l'invocation de ce principe par une pluralité d'acteurs a révélé toute la complexité de sa mise en œuvre²⁴. Parmi les différentes limites du principe, soulignons d'entrée de jeu la géométrie variable de sa portée normative. Que vise-t-il plus précisément et comment celui-ci peut-il être mobilisé pour prévenir

-
21. *Ibid.*, p. 19. Voir également sur le sujet A. Genus et A. Stirling (2018), « Collingridge and the dilemma of control: Towards responsible and accountable innovation », *Research Policy*, vol. 47, n° 1, p. 63.
 22. C. Newton (2 avril 2019), « Facebook is making an all-out push for regulation – on its own terms », *The Verge*, <<https://www.theverge.com/interface/2019/4/2/18291413/facebook-regulation-mark-zuckerberg-european-tour>>.
 23. *Déclaration de Rio sur l'environnement et le développement*, 12 août 1992, Doc. N.U. A/ CONF.151/26, principe 15.
 24. Pour une analyse approfondie des forces et des limites du principe de précaution, voir G.A. Legault, L. Bernier, C.-É. Daniel, C. Fontaine et J. Patenaude (dir.) (2012), *Nanotechnologies et principe de précaution. Forces et limites de l'appel au principe*, Québec: Presses de l'Université Laval.

les conséquences néfastes du développement technologique? L'une des premières difficultés rencontrées lorsqu'on tente d'appliquer le principe de précaution se situe dans sa double appartenance: ce principe peut autant être rattaché à la morale ou au droit²⁵. L'un et l'autre de ces deux univers normatifs n'entraînent toutefois pas les mêmes conséquences. Sous sa forme morale, le principe de précaution est surtout invoqué relativement aux risques que pose le développement technologique dans l'ensemble de son processus, ce qui inclut les différents usages d'une technologie potentiellement néfaste. Par exemple, c'est dans ce contexte plus large que le principe moral de précaution a été soulevé lors du débat public français sur les nanotechnologies par différents acteurs sociaux, en guise de réponse aux transformations sociétales néfastes anticipées relativement à leur développement²⁶. Ainsi, certains groupes d'intérêts voués à la protection des consommateurs et de l'environnement ont prôné la précaution envers les dérives de l'utilisation des nanotechnologies, notamment lorsqu'elles sont employées dans des nanodispositifs augmentant les capacités humaines²⁷. De cette manière, la forme morale du principe de précaution interpelle surtout une prise de conscience collective, en vertu de laquelle le développement technologique doit pouvoir être soumis à des choix de société²⁸. C'est pour cette raison que l'appel du principe de précaution sous sa forme morale a amené plusieurs acteurs sociaux à justifier l'imposition de moratoires sur l'innovation technologique²⁹.

Sous sa forme juridique, les différentes interprétations du principe de précaution par les tribunaux ont toutefois restreint sa portée aux seuls risques qu'un produit issu du développement technologique peut entraîner pour l'environnement et – par extension – la santé humaine. Par exemple, c'est dans cette optique plus restreinte que la Cour suprême du Canada a validé l'interdiction prise par une municipalité d'interdire l'épandage de pesticides³⁰ pour protéger l'environnement, ou encore que la Cour d'appel de Versailles a ordonné

25. G.A. Legault (2012), « Comprendre la géométrie variable et les usages multiples du principe de précaution », dans G.A. Legault, L. Bernier, C.-É. Daniel, C. Fontaine et J. Patenaude (dir.), *Nanotechnologies et principe de précaution. Forces et limites de l'appel au principe*, Québec: Presses de l'Université Laval, p. 64.

26. C.-É. Daniel, G.A. Legault et L. Bernier (2015), « La régulation des nanotechnologies, le débat national français et le dialogue social: nanomonde, grandes attentes normatives? », *op. cit.*, p. 109-111.

27. *Ibid.*, p. 111.

28. G.A. Legault (2012), « Conclusion », dans G.A. Legault, L. Bernier, C.-É. Daniel, C. Fontaine et J. Patenaude (dir.), *Nanotechnologies et principe de précaution. Forces et limites de l'appel au principe*, Québec: Presses de l'Université Laval, p. 110-111.

29. C.-É. Daniel, G.A. Legault et L. Bernier (2015), *op. cit.*, p. 109.

30. *114957 Canada Ltée (Spraytech, Société d'arrosage) c. Hudson (Ville)*, 2001 CSC 40, para. 31 et 32.

le démantèlement d'une antenne-relais³¹ pour protéger la santé humaine. Comment expliquer une telle différence entre les deux interprétations? Dans l'univers juridique, la mise en œuvre du principe de précaution s'effectue par le biais des juges, qui examinent alors si une décision prise par les autorités publiques respecte ou non les exigences du principe³². Il ne s'agit donc pas de remettre en question les orientations du développement technologique, mais plutôt d'examiner après coup si les données scientifiques disponibles étaient de nature à justifier la prise ou l'absence de mesures relativement au risque anticipé pour l'environnement ou la santé humaine. Nous retrouvons ici une seconde difficulté dans l'application du principe juridique de précaution, cette fois liée à sa formulation. Quel niveau de risques est nécessaire pour mettre en branle ce principe? L'énoncé du principe requiert qu'en cas de dommages graves ou irréversibles, l'absence de certitude scientifique absolue ne doit pas empêcher les États à mettre en place des mesures visant à prévenir le risque anticipé. Dans ce cas, la gravité ou l'irréversibilité des risques doit-elle être possible, probable ou peut-elle être plutôt hypothétique? La preuve par analogies est-elle de nature à entraîner l'application du principe? Où se situe le seuil minimal de connaissances scientifiques à partir duquel la prise de mesures par les autorités publiques peut être justifiée?

Traditionnellement, la science et la technologie sont directement mobilisées lors de l'élaboration des règlements pour déterminer la nature et l'occurrence des risques à encadrer à partir des données scientifiques permettant d'établir les relations d'une cause et de ses effets³³. En d'autres termes, la prise de mesures destinées à limiter ou à empêcher la survenue de risques pour la santé ou l'environnement doit se justifier par la présence de « fondements scientifiques solides et crédibles »³⁴. Cette exigence s'explique du fait qu'elle

31. *S.A. Bouygues Telecom*, Cour d'appel de Versailles, 4 février 2009, 08/08775.

32. Voir à titre d'illustration les décisions suivantes liées à l'analyse de l'application du principe de précaution: *Morton c. Canada (Pêches et Océans)*, 2019 CF 143 aux para 149 et ss.; *Morton c. Canada (Pêches et Océans)*, 2015 CF 575, aux para 40-49.

33. A. Zei (2013), « Shifting the boundaries or breaking the branches? On some problems arising with the regulation of technology », dans E. Palmerini et E. Stradella (dir.), *Law and Technology: The Challenge of Regulating Technological Development*, Pise: Pisa Universitu Press, p. 170-171.

34. Voir à titre d'exemple: Secrétariat du Conseil du Trésor du Canada (Septembre 2018), *Politique sur l'élaboration de la réglementation*, Bureau du Conseil privé, <<https://www.canada.ca/fr/secretariat-conseil-tresor/services/gestion-reglementation-federale/lignes-directrices-outils/politique-elaboration-reglementation.html>>; T. Tao *et al.* (Août 2018), *Poids de la preuve: Principes généraux et applications actuelles à Santé Canada*, Santé Canada, <<https://www.canada.ca/content/dam/hc-sc/documents/services/publications/science-research-data/weight-evidence-general-principles-current-applications/poids-preuve-principes-generaux-applications-actuelles.pdf>>.

découle d'une analyse des coûts et des bénéfices réalisés par les autorités publiques lors de l'élaboration d'un règlement. Compte tenu de l'importance accordée au libre marché et au développement économique, les mesures réglementaires visant à limiter ou interdire la commercialisation d'un produit ne peuvent être adoptées que sur la base de données scientifiques établissant la présence d'un risque pour la santé et la sécurité humaines ou l'environnement³⁵. Dans ce contexte, le principe de précaution invoqué en situation d'incertitude scientifique se heurte à l'exigence d'obtenir une base minimale de connaissances permettant de déterminer la nature, la gravité et l'occurrence du risque anticipé. C'est à cet égard que se situe cette deuxième difficulté d'appliquer le principe de précaution. L'univers juridique exige une certaine causalité scientifique pour la prise de mesures réglementaires visant à restreindre le développement économique et – incidemment – l'innovation. Une fois intégré dans l'univers juridique, le principe de précaution se retrouve essentiellement mobilisé à titre de principe interprétatif pour fonder une décision judiciaire, laquelle valide ou non le choix des autorités publiques d'agir ou de s'abstenir face à un risque pour l'environnement ou la santé³⁶. En d'autres termes, la seconde difficulté de l'application juridique du principe de précaution provient du fait qu'une fois intégré au droit, il quitte le champ de la prévention pour fonder une réparation comme la remise en l'état des choses (comme dans le cas des antennes-relais) ou l'octroi de dommages-intérêts³⁷.

Ces deux difficultés liées au principe de précaution illustrent ainsi tout le défi du droit à mettre en place un encadrement normatif davantage axé sur la prévention et l'anticipation des risques soulevés par le développement technologique. En arrière-plan de ces difficultés, nous retrouvons la préoccupation constante de laisser – le plus possible – libre cours au développement économique et à l'innovation. Autrement dit, la mobilisation du principe de précaution sous sa forme juridique limite son action préventive par rapport aux risques que peut causer un produit pour la santé ou l'environnement, puisque l'adoption de mesures réglementaires ne peut restreindre indûment la liberté de commerce et l'innovation technologique sans justification scientifique minimale préalable. Que peut donc offrir le droit comme encadrement normatif du développement de l'IA et de la robotique, alors que se posent constamment

35. E. Sirsi (2013), « Agri-food technologies and the law », Dans E. Palmerini et Stradella E. (dir.), *Law and Technology: The Challenge of Regulating Technological Development*, Pise: Pisa University Press, p. 233.

36. C.-É. Daniel, G.A. Legault et L. Bernier (2015), « La régulation des nanotechnologies, le débat national français et le dialogue social: nanomonde, grandes attentes normatives? », *op. cit.*, p. 111 et 115.

37. G.A. Legault, L. Bernier et C.-É. Daniel (2014), « Démocratisation du développement technologique: trois défis pour l'éthique publique », *Ethica*, vol. 18, n° 2, p. 65-91.

en trame de fond les tensions illustrées par le dilemme de Collingridge et l'importance accordée au libre cours de l'innovation ? De quelle manière peut-il le faire ? Pour mieux esquisser une réponse à ces questions, nous effectuerons dans la prochaine section un survol des différentes possibilités qu'offre le droit lorsque celui-ci est mobilisé comme outil d'encadrement du développement technologique.

PARTIE II – LE RECOURS AU DROIT COMME MODE DE GOUVERNANCE POUR ENCADRER LE DÉVELOPPEMENT DE L'IA ET DE LA ROBOTIQUE

Tout développement technologique entraîne des conséquences pour la société dans laquelle il s'inscrit. Nous l'avons vu, le développement de l'IA et de ses applications telles que la robotique promet de nombreux bénéfices, mais pose également différents risques de natures variées. Certains craignent par exemple un accroissement des atteintes à la vie privée par le biais d'une surveillance accrue³⁸ ou soulèvent les risques engendrés par les systèmes d'IA pour la sécurité humaine³⁹. D'autres s'inquiètent autrement des biais et des discriminations pouvant être implicitement introduits dans les algorithmes décisionnels⁴⁰ ou encore de la transformation des emplois résultant d'un recours de plus en plus répandu aux machines dans les différents secteurs d'activité⁴¹. Dans ce contexte, lorsqu'on interpelle le droit pour encadrer le développement technologique, c'est plus particulièrement pour agir (limiter, permettre, interdire, etc.) sur ces différentes conséquences soulevées que l'on mobilise les différents cadres normatifs. Autrement dit, cette volonté d'agir sur le

-
38. M. Whittaker *et al.* (décembre 2018), *AI Now Report 2018*, <https://ainowinstitute.org/AI_Now_2018_Report.pdf>; Feldstein S. (22 avril 2019), «How artificial intelligence systems could threaten democracy», *The Conversation*, <<http://theconversation.com/how-artificial-intelligence-systems-could-threaten-democracy-109698>>.
39. J. Hecht (29 novembre 2018), «Managing expectations of artificial intelligence», *Nature Outlook*, vol. 563, S141, <<https://www.nature.com/articles/d41586-018-07504-9>>.
40. A. Springer, J. Garcia-Gathright et H. Cramer (2018), «Assessing and Addressing Algorithmic Bias – But Before We Get There», *2019 AAAI Spring Symposium Series*, <<https://www.aaai.org/ocs/index.php/SSS/SSS18/paper/viewFile/17542/15470>>; Y. Eisenstat (12 février 2019), «The real reason tech struggles with algorithmic bias», *Wired*, <<https://www.wired.com/story/the-real-reason-tech-struggles-with-algorithmic-bias>>; K. Hao (4 février 2019), «This is how AI bias really happens», *MIT Technology Review*, <<https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix>>.
41. C.B. Frey et M.A. Osborne (2013), «The Future of Employment: How Susceptible Are Jobs to Computerisation?», Oxford: Oxford Martin School, <https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf>; M. Segal (29 novembre 2018), «How automation is changing work», *Nature Outlook*, vol. 563, S132, <<https://www.nature.com/articles/d41586-018-07501-y>>.

développement technologique implique d'employer le droit comme un mode de gouvernance – c'est-à-dire comme un moyen normatif permettant d'encadrer les différentes conséquences soulevées, soit par une technologie en particulier ou encore par son processus de développement.

Section 1. Une méthode d'analyse pour examiner la capacité du droit comme mode de gouvernance à faire face au développement de l'IA et de la robotique

En réalité, chacune de ces conséquences – ou impacts – du développement technologique peut se rattacher à un enjeu particulier qui concerne une dimension de nature personnelle ou collective de la vie humaine. Par exemple, lorsqu'une analyse de risque est effectuée sur un nouveau produit technologique, on cherche plus précisément à déterminer les impacts négatifs que ce produit entraîne à l'égard des enjeux de santé ou de sécurité humaines. Dans la même veine, la projection des retombées économiques d'une technologie consiste en fait à déterminer ses impacts positifs sur l'enjeu de l'économie. En ce sens, les impacts qu'entraîne une technologie à l'égard d'un ou plusieurs enjeux peuvent être de nature positive ou négative. Le recours au concept d'enjeu nous permet surtout de mieux décrire de quelle manière les impacts soulevés par une technologie se retrouvent à concerner les dimensions de notre vie personnelle et collective. La grille d'analyse d'impacts et d'acceptabilité éthique développée par le groupe de recherche interdisciplinaire InternE³LS propose l'identification de dix enjeux couramment retrouvés dans le contexte du développement technologique : la santé, la sécurité, la vie/mort, l'environnement, l'économie, la recherche scientifique, la liberté de choix, la vie privée, la nature et l'identité humaine, et le vivre-ensemble de nature locale, nationale et internationale⁴².

En cherchant à encadrer l'incidence du développement technologique à titre de mode de gouvernance, le droit offre de prendre en compte plus particulièrement certains enjeux parmi l'ensemble de ceux soulevés. Autrement dit, les impacts du développement technologique ne sont pas tous pris en compte de la même manière par la gouvernance juridique. Dans cette optique, la seconde section du texte propose d'éclairer davantage comment le droit encadre les impacts du développement de l'IA et de la robotique en examinant de quelle manière celui-ci tient compte d'enjeux plus spécifiques que d'autres.

42. J. Patenaude, G.A. Legault, J. Bauvais, L. Bernier, J.-P. Béland, P. Boissy, V. Chenel, C.-É. Daniel, J. Genest, M.-S. Poirier et D. Tapin (2015), « Framework for the Analysis of Nanotechnologies' Impacts and Ethical Acceptability : Basis for an Interdisciplinary Approach to Assessing Novel Technologies », *Science and Engineering Ethics*, vol. 21, n° 2, p. 293-315.

Pour ce faire, il importe en premier lieu de préciser ce qu'est le « droit », lorsque celui-ci est mobilisé comme mode de gouvernance pour encadrer le développement technologique. Traditionnellement, le droit se conçoit comme un ensemble de règles obligatoires et mobilisant la contrainte publique dont se dote un État pour diriger les comportements dans la société⁴³. En ce sens, le qualificatif de « dur » est souvent donné aux règles juridiques retrouvées dans la législation et la réglementation (« *hard law* »), compte tenu de leur double caractéristique liée à l'obligation et à la sanction⁴⁴. Outre le droit dur, la gouvernance par le droit du développement technologique mobilise également depuis quelques années une autre forme de droit, cette fois plus indistincte à définir. Communément désignée dans la littérature juridique sous le terme de « *soft law* », plusieurs juristes, chercheurs et groupes d'experts du secteur public et privé ont régulièrement soulevé les avantages de recourir à cette forme de droit considérée comme plus souple pour mieux tenir compte du contexte rapide du développement technologique⁴⁵. Regroupant différents codes de conduite, normes techniques et normes éthiques, le *soft law* offrirait ainsi le bénéfice d'être adapté plus facilement et rapidement relativement à l'évolution rapide des technologies émergentes⁴⁶.

En 2013, un rapport annuel du Conseil d'État français est venu mieux préciser la nature du concept de *soft law*⁴⁷. Deux formes de droit peuvent ainsi être dégagées du concept de *soft law*, chacune comportant des caractéristiques leur étant propres. D'un côté, nous pouvons établir une forme « molle » du droit, qui se distingue du droit dur par une absence totale de contrainte et de sanction. Ce droit mou doit toutefois être employé pour orienter, baliser ou guider les comportements sociaux et être élaboré en suivant un certain formalisme⁴⁸. De l'autre, nous retrouvons entre le droit dur (obligatoire et contraignant) et le droit mou (non obligatoire, non contraignant) une forme

43. I. Hachez (2010), « Balises conceptuelles autour des notions de « source du droit », « force normative » et « *soft law* », *Revue interdisciplinaire d'études juridiques*, vol. 65, n° 2, p. 19.

44. I. Hachez (2012), « Le *soft law* : qui trop embrasse mal étirent ? », dans I. Hachez et al. (dir.), *Les sources du droit revisitées. Théorie des sources du droit*, vol. 4, Bruxelles : Publications des Facultés universitaires Saint-Louis, p. 51-100.

45. G.N. Mandel (2013), « Emerging technology governance », dans G.E. Marchant, K.W. Abbott et B. Allenby (dir.), *Innovative Governance Models for Emerging Technologies*, Northampton : Edward Elgar, p. 44-62 ; R. Leenes, E. Palmerini, B.-J. Koops, A. Bertolini, P. Salvini et F. Lucivero (2017), « Regulatory challenges of robotics : Some guidelines for addressing legal and ethical issues », *Law, Innovation and Technology*, vol. 9, n° 1, p. 1-44.

46. E. Stradella (2013), « Approaches for regulating robotic technologies : Lessons learned and concluding remarks », dans Palmerini E. et Stradella E. (dir.), *Law and Technology : The Challenge of Regulating Technological Development*, Pise : Pisa University Press, p. 335-338.

47. Conseil d'État (2013), *Le droit souple*, Étude annuelle 2013, Paris : La Documentation française.

48. *Ibid.*, p. 62.

«souple» du droit. Sans être complètement dénué de contrainte juridique, le droit souple incite davantage qu'il n'oblige à accomplir ou à interdire une action⁴⁹. En ce sens, le droit souple peut être utilisé par les juges pour interpréter une disposition législative ou appuyer leurs motifs dans une décision judiciaire. Par exemple, le principe de précaution invoqué par la Cour suprême du Canada dans l'affaire *Spraytech*⁵⁰ constitue une manifestation du droit souple, puisqu'il a été employé par les juges comme moyen interprétatif pour appuyer leur appréciation de la validité d'une disposition réglementaire.

Ayant ainsi d'abord présenté le portrait de la gouvernance par le droit du développement technologique, il nous faut par la suite établir la manière dont elle sera mobilisée pour encadrer les impacts de l'IA et de ses utilisations. Comment ces trois formes de nature dure, souple et molle du droit peuvent-elles être utilisées pour réguler le développement de l'IA et de la robotique? Que couvrent les différents dispositifs normatifs rattachés à ces trois formes de droit; autrement dit, quels impacts de quels enjeux sont-ils pris en compte par le droit dur, souple et mou? Deux éléments d'analyse peuvent être employés pour répondre à ces questions. En premier lieu, les trois formes de droit ne sont pas toutes invoquées de la même manière relativement aux impacts de l'IA et de la robotique. Chacune d'elles offre de couvrir de manière différente les impacts du développement de l'IA et de la robotique, en tenant compte de certains enjeux plus spécifiques que d'autres. En ce sens, *l'étendue* des dispositifs normatifs peut donc varier et constituera ainsi notre premier critère d'examen. En second lieu, les droits dur, souple et mou n'ont pas tous la même capacité à produire des effets sur les individus relativement à leur mise en œuvre. Nous avons vu que ces trois formes de droit possèdent des caractéristiques différentes, selon qu'elles sont obligatoire, incitative ou non obligatoire. Par conséquent, la nature de ces caractéristiques entraînera différents effets pour les individus, une norme obligatoire n'entraînant pas les mêmes effets dans la société que celle qui ne l'est pas. De cette manière, la *portée normative* des dispositifs normatifs liés au droit dur, souple et mou constituera notre second critère d'examen.

Cette méthode en deux temps nous permettra de mieux comprendre par un bref survol comment les formes de droit dur, souple et mou cherchent à encadrer – chacune à leur manière – les impacts du développement de l'IA et de la robotique, tout en tenant compte du dilemme de Collingridge. Les trois sections subséquentes présenteront respectivement les dispositifs de droit dur,

49. C. Thibierge (2009), «Rapport de synthèse», dans *Association Henri Capitant des amis de la culture juridique française, Le droit souple: Journées nationales Tome XIII / Boulogne-sur-Mer*, Paris: Dalloz, p. 141-161.

50. *114957 Canada Ltée (Spraytech, Société d'arrosage) c. Hudson (Ville)*, *op. cit.*

souple et mou, en examinant pour chacune d'elles (a) l'étendue des dispositifs normatifs qu'elles comportent et (b) leur portée normative.

Section 2. Le droit dur par rapport aux enjeux de santé, de sécurité et de vie privée soulevés par l'IA et la robotique

Parmi l'ensemble des différentes lois et réglementations composant le droit dur, deux perspectives orientent de manière générale le recours aux normes juridiques pour encadrer le développement de l'IA et de la robotique⁵¹. D'un côté, le droit public s'incarne dans les exigences qu'établit l'État par ses lois et règlements spécifiques à des domaines d'application particuliers, tels que l'environnement, les aliments et médicaments, la sécurité automobile, etc. En matière d'intelligence artificielle et de robotique, le droit réglementaire énoncé par les autorités publiques cible principalement trois enjeux : la santé, la sécurité et la vie privée. De l'autre côté, différents régimes juridiques retrouvés dans le *Code civil* québécois ou français composent le droit privé, ceux-ci énonçant une série d'obligations qui régissent plutôt les rapports qu'ont les personnes physiques et morales entre elles, comme le mariage, les contrats, les successions, etc. Parmi les différents régimes du *Code civil*, celui de la responsabilité civile est plus particulièrement applicable au contexte du développement de l'IA et de la robotique. Bien qu'elle n'ait pas le même mode de fonctionnement que le droit réglementaire, la responsabilité civile offre elle aussi d'encadrer les mêmes enjeux de santé, sécurité et vie privée. Cette distinction effectuée entre le droit dur de nature publique et privée offre surtout de mieux saisir la manière dont la gouvernance juridique peut être mise en œuvre pour encadrer les enjeux de santé, de sécurité et de vie privée. C'est en tenant compte de cette distinction que nous examinerons respectivement les régimes public du droit réglementaire (2.1) et privé de la responsabilité civile (2.2).

2.1 Le droit réglementaire

a) L'étendue des dispositifs juridiques retrouvés dans le droit réglementaire

Si l'intelligence artificielle à proprement parler ne fait pas actuellement l'objet d'une réglementation lui étant explicitement spécifique, son intégration dans une multitude d'applications et de dispositifs technologiques est – quant à elle – susceptible d'être encadrée par l'arsenal réglementaire déjà en place. Autrement dit, le droit réglementaire offre d'encadrer le développement de l'IA par le biais des différents instruments législatifs et réglementaires

51. A. Émond (2016), *Introduction au droit canadien*, 2^e éd., Montréal : Wilson & Lafleur, p. 62-63.

présentement applicables aux dispositifs technologiques dans lesquels seront intégrés les algorithmes de l'intelligence artificielle. Par exemple, une modification au *Règlement de l'aviation canadien*⁵² a été effectuée en 2019 pour tenir compte des situations où un drone effectuerait des vols de manière autonome (art. 901.32). Parmi les autres applications de l'IA en matière de robotique, d'autres dispositifs technologiques comme les voitures connectées et autonomes et les robots de soins personnels pourraient par exemple faire éventuellement l'objet de précisions réglementaires en vertu du *Règlement sur la sécurité des véhicules automobiles*⁵³ ou le *Règlement sur les instruments médicaux*⁵⁴.

À cet égard, la réglementation destinée aux drones, aux voitures autonomes ou aux robots de soins personnels offre principalement d'encadrer les risques que peuvent causer ces dispositifs technologiques à l'égard des enjeux de santé et de sécurité humaines. Quels paramètres de vol un individu doit-il respecter pour que l'utilisation de son drone n'entraîne pas un risque de blessures pour les autres personnes se situant à proximité? À quelles exigences de sécurité un fabricant automobile doit-il se conformer pour concevoir, vendre ou importer un véhicule connecté et semi-autonome? Quels tests doivent être effectués par le fabricant pour démontrer qu'un robot de soins personnels peut être employé dans un contexte médical de manière sécuritaire pour éviter la survenue de risques pour la santé ou la sécurité d'un patient? Toutes ces questions sollicitent l'identification préalable des dangers qu'un fabricant doit rencontrer pour obtenir un permis ou une autorisation de la part des autorités publiques. En ce sens, le droit réglementaire se concentre principalement sur les différentes exigences sanitaires et sécuritaires d'un produit devant être respectées autant par les différentes entreprises souhaitant fabriquer, vendre ou importer ce dernier que par les particuliers désirant l'utiliser.

Dans cette optique, le droit réglementaire distingue l'encadrement qu'il offre spécifiquement au produit technologique de celui plus généralement appliqué aux usages qui en sont faits. Selon cette logique, la technologie doit être considérée comme neutre, et ce sont les personnes l'utilisant qui se retrouvent en fait à déterminer si son usage est adéquat ou non. À titre illustratif, un marteau peut très bien servir comme outil pour planter un clou ou comme arme offensive pour blesser quelqu'un. Pourvu que l'utilisation du marteau soit sécuritaire – par exemple que la tête ne se détache pas du manche –, l'usage qui en est effectué n'est plus de l'apanage du droit réglementaire.

52. DORS/96-433.

53. CRC, ch. 1038.

54. DORS/1998-282.

Ainsi, un drone peut être utilisé pour mettre en valeur une propriété à vendre, tout comme il peut être utilisé pour surveiller les propriétaires voisins⁵⁵. De la même manière, les systèmes de télématique embarqués dans les voitures connectées peuvent rendre plus sécuritaire l'utilisation d'applications comme *Google Maps*, tout en recueillant une foule d'informations sur nos déplacements qui détaillent nos habitudes de consommation les plus intimes⁵⁶. Dans cet esprit, les possibles atteintes à la vie privée entraînées par le développement de l'IA sont reléguées à des considérations plus générales liées à ses usages, qui échappent à la protection conférée par la réglementation sur des applications comme les drones ou les voitures connectées. Ces usages n'étant pas du ressort exclusif des autorités réglementaires, c'est aux personnes morales et physiques qu'il convient de s'assurer que leur conduite demeure respectueuse des cadres mis en place pour protéger la vie privée, plus particulièrement en matière de protection des renseignements personnels.

En vertu de cette distinction réalisée par les cadres réglementaires, l'enjeu de la vie privée n'est donc pas spécifiquement pris en compte dans le *Règlement de l'aviation canadien* et relève plutôt des dispositifs législatifs plus généraux encadrant la protection des renseignements personnels. Les autorités fédérales canadiennes se contentent ainsi d'indiquer sur leur site Internet l'importance de protéger la vie privée, en indiquant qu'il demeure de la responsabilité des pilotes de drones de s'assurer que les renseignements recueillis par un drone et permettant d'identifier un individu respectent la législation fédérale en matière de protection des renseignements personnels⁵⁷. Autrement dit, l'étendue de la protection juridique du droit réglementaire par rapport au développement de

55. S. Gawdin (27 juillet 2018), «Residents in B.C. city complain about drones spying in backyards», *The Colombia Valley Pioneer*, <<https://www.columbiavalleypioneer.com/news/residents-in-b-c-city-complain-about-drones-spying-on-backyards>>. Dans l'affaire *Pilote c. Nolet*, le défendeur a, durant une partie de soccer à laquelle participait son fils, sérieusement abîmé à l'aide d'un bâton le drone du demandeur qui le faisait voler à proximité. Les faits rapportés dans la décision ne précisent pas les motifs du geste du défendeur, le juge se contentant de le décrire comme étant «excédé par la présence du drone». Voir *Pilote c. Nolet*, 2017 QCCQ 14302.

56. British Columbia Freedom of Information and Privacy Association (8 juillet 2019), *The Connected Car: Who is in the driver's seat?*, 2019 Update, <https://fipa.bc.ca/wordpress/wp-content/uploads/2019/07/2019_CCReport_Update-FINAL.pdf>.

57. Transport Canada (2019), «Utiliser votre drone de façon sécuritaire et légale», <<https://www.tc.gc.ca/fr/services/aviation/securite-drones/utiliser-drone-facon-securitaire-legale.html>>. Il est intéressant de constater que Transport Canada a également mis en ligne des lignes directrices, qui se retrouvent essentiellement à vulgariser de manière générale le contenu des lois canadiennes en matière de protection des renseignements personnels. Voir à ce sujet Transport Canada (2019), *Lignes directrices sur la protection de la vie privée à l'intention des utilisateurs de drones*, <<https://www.tc.gc.ca/fr/services/aviation/securite-drones/lignes-directrice-protection-vie-privee-utilisateurs-drones.html>>.

l'IA et de la robotique permet de cibler les impacts entraînés par les dispositifs sur les enjeux de santé et de sécurité.

b) La portée normative des normes réglementaires

Comment le droit réglementaire peut-il assurer le respect des exigences qu'il énonce pour protéger la santé, la sécurité et la vie privée relativement aux impacts du développement de l'IA et de la robotique? Cette question est importante, parce qu'un règlement qui ne produit aucun effet réellement constaté peut voir son utilité remise en cause et être éventuellement abrogé. Toute la protection garantie par le droit réglementaire à l'égard des enjeux de santé, de sécurité et – incidemment – de la vie privée repose sur sa capacité à être mis en œuvre par les personnes physiques ou morales auxquelles il s'adresse. Par exemple, un drone ne sera utilisé de manière sécuritaire que si son pilote respecte les exigences prescrites à chaque vol effectué. Dans la même veine, une voiture autonome ne sera sécuritaire et ne respectera la vie privée de ses utilisateurs que dans la mesure où le fabricant et les autres intermédiaires ayant participé à sa conception se conforment aux exigences énoncées dans les règlements pertinents.

Pour ce faire, les législations et les règlements adoptés par les autorités publiques reposent sur l'imposition de sanctions, telles que des amendes, afin de dissuader et de punir tout écart de conduite qui dérogerait aux prescriptions réglementaires imposées. S'il peut être difficile de sanctionner *tout écart de conduite identifié*, la présence d'un certain degré de surveillance peut quand même entraîner le phénomène de « l'ombre portée de la sanction »⁵⁸, c'est-à-dire la considération que son comportement non conforme aux exigences soit possiblement sanctionné. Ce mode de fonctionnement du droit réglementaire comporte toutefois une importante limite: ses exigences ne sont susceptibles d'être respectées que dans la mesure où une forme de contrôle et de surveillance *est réellement effectuée* par des inspecteurs ou d'autres fonctionnaires sur le terrain.

Outre le respect des exigences réglementaires établies par l'État, les impacts du développement de l'IA peuvent également être pris en compte par les différents régimes de responsabilité civile retrouvés dans le *Code civil*. Le recours à cet autre outil du droit dur entraîne cependant un changement de perspective. Si – dans l'optique réglementaire – l'encadrement des risques repose sur des règles contraignantes établies et appliquées par l'État, la responsabilité civile encadre pour sa part les droits et obligations que doivent respecter les personnes physiques et morales entre elles-mêmes. Ce changement de perspective du

58. Conseil d'État (2013), *Le droit souple*, *op. cit.*, p. 11.

droit public au droit privé entraîne toutefois un changement relativement à l'étendue et à la portée normative des règles de la responsabilité civile, qu'il convient d'examiner dans la prochaine sous-section.

2.2 La responsabilité civile

a) L'étendue des régimes de responsabilité civile

C'est en raison des inquiétudes manifestées notamment à l'égard de la sécurité du développement de l'IA et de la robotique que la responsabilité civile a alimenté plusieurs réflexions et discussions. Tout comme le droit réglementaire, la responsabilité civile offre aussi la possibilité d'encadrer les impacts soulevés par le développement de l'IA à l'égard des enjeux de santé et de vie privée. En revanche, la mise en œuvre de l'encadrement proposé par la responsabilité civile s'effectue d'une tout autre manière. Plutôt que d'imposer une pluralité d'exigences principalement destinées à identifier et à circonscrire les risques des produits issus du développement de l'IA, les règles de la responsabilité civile offrent à l'inverse de compenser les dommages subis par les applications de l'intelligence artificielle. Autrement dit, la responsabilité civile impose aux entreprises et aux individus un devoir général de prudence et de bonne conduite par rapport aux autres personnes de la société. Si la personne physique ou morale commet une faute en ne se conformant pas aux différentes obligations découlant de ce devoir général de prudence et que cela cause un préjudice, cette personne doit alors indemniser la victime ayant subi un tel préjudice.

C'est en vertu de sa fonction de compensation des dommages subis que le droit de la responsabilité civile offre d'encadrer les impacts du développement de l'IA et de la robotique sur les enjeux de santé, de sécurité et de vie privée. Ainsi, ces impacts prendront la forme de différentes atteintes de nature physique (telles que des blessures occasionnées par un robot) ou encore de nature psychologique (telles qu'un stress vécu à la suite d'une surveillance exercée par un drone). Toute personne ayant subi un préjudice lié, par ces atteintes, à la santé, à la sécurité ou à la vie privée pourra, à la suite d'une preuve déposée devant un tribunal, requérir une indemnisation de la part du fautif. Ce n'est donc qu'après coup, une fois le dommage causé, que le droit de la responsabilité civile permet à la victime d'obtenir une réparation du tort subi. Pour ce faire, la preuve requiert la présence des trois éléments constitutifs de la responsabilité civile: la commission d'une faute par une personne (qu'il s'agisse d'un individu ou d'une entreprise), un préjudice sous forme d'un dommage subi et un lien de causalité reliant directement ces deux éléments.

C'est en raison de l'exigence de ces trois éléments constitutifs que l'application des règles actuelles de la responsabilité civile au développement de l'IA et de la robotique suscite plusieurs débats au sein de la communauté juridique⁵⁹. Qui demeure responsable du préjudice commis par une IA lorsque celle-ci prend une décision imprévisible, et ce, de manière tout à fait autonome? Jusqu'à où un fabricant de robots ou son utilisateur peut-il voir sa responsabilité civile engagée par un geste commis par un robot? Dans son état actuel, la responsabilité civile ne peut s'appliquer qu'à l'égard d'une personne physique ou morale, excluant la possibilité qu'une IA soit tenue de réparer les dommages qu'elle a elle-même causés⁶⁰. En effet, seuls les humains et les entreprises sont juridiquement reconnus comme des personnes, ce qui leur confère différents droits et obligations – dont la constitution d'un patrimoine. Considérée comme un bien meuble à l'instar d'un animal domestique ou d'un outil, l'intelligence artificielle ne dispose donc pas d'un statut juridique spécifique. Tout l'enjeu de l'application de la responsabilité civile repose ainsi sur la capacité du droit à déterminer jusqu'à quel point un individu ou une entreprise sera tenu d'indemniser la victime d'un dommage commis par une IA⁶¹. Dans cette optique, le nœud du problème lié au recours à la responsabilité civile pour encadrer le développement de l'IA et de la robotique se situe dans la considération qu'accorde le droit aux capacités d'autonomie décisionnelle conférée à l'intelligence artificielle⁶². Plus précisément, le développement de l'IA et de la robotique vient mettre en question le degré de contrôle nécessaire que doit exercer le fabricant ou l'utilisateur d'un robot à l'égard des décisions prises par la machine⁶³.

59. Voir notamment à ce sujet I. Cofone (2018), « Servers and Waiters: What Matters in the Law of A.I. », *Stanford Law Review*, vol. 21, n° 2, p. 167-197; A. Bertolini (2013), « Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules », *Law, Innovation and Technology*, vol. 5, n° 2, p. 214-247.

60. I. Giuffrida, F. Lederer et F. Vermeys (2018), « A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law », *Case Western Reserve Law Review*, vol. 68, n° 3, p. 763.

61. S. Beck (2016), « The problem of ascribing legal responsibility in the case of robotics », *AI & Society*, vol. 31, n° 4, p. 473-481; A. Matthias (2004), « The responsibility gap: Ascribing responsibility for the actions of learning automata », *Ethics and Information Technology*, vol. 6, n° 4, p. 175-183.

62. P. Asaro (21-23 mars 2016), « The Liability Problem for Autonomous Artificial Agents », *Proceedings of the AAAI Symposium on Ethical and Moral Considerations in Non-Human Agents*, Stanford: Stanford University, p. 191.

63. C.J. Karnow (2016), « The application of traditional tort theory to embodied machine intelligence », dans R. Calo, A.M. Froomkin et I. Kerr (dir.), *Robot Law*, Northampton (MA): Edward Elgar, p. 73.

En droit civil québécois comme en droit civil français, certains régimes spécifiques de responsabilité civile ont été adoptés pour faciliter la preuve de la victime et favoriser son indemnisation lors d'un recours judiciaire en matière d'atteinte à la sécurité. Par exemple, le gardien d'un bien est présumé avoir eu une conduite fautive si ce dernier cause un préjudice par son fait autonome. Autrement dit, celui qui contrôle, détient ou utilise un bien est présumé avoir commis une faute dans sa garde si ce bien – par l'effet d'un dynamisme qui lui est propre – cause un préjudice à une autre personne. En vertu de l'article 1465 du *Code civil du Québec*, le gardien sera tenu de réparer le préjudice causé par le fait autonome du robot, à moins qu'il prouve n'avoir commis aucune faute dans la garde qu'il a exercée. Dans l'éventualité où une machine cause un préjudice résultant de son propre fait, le problème que devra résoudre un juge consiste à déterminer dans quelle mesure le gardien pourra voir sa responsabilité engagée sur la base de la présomption de faute que lui impose le droit civil. Dans la mesure où le gardien réussit à démontrer au juge qu'il n'a commis aucune faute dans la garde du robot, celui-ci pourra s'exonérer de sa responsabilité et ne sera pas tenu de réparer le préjudice causé par le robot. Ainsi, le préjudice occasionné par un robot doté d'une autonomie décisionnelle totale et susceptible d'échapper au contrôle et à la surveillance exercés par le gardien complexifie l'application de ce régime particulier de responsabilité civile⁶⁴. Comment accorder une réparation dans une telle situation si le gardien démontre qu'il n'a commis aucune faute dans sa garde du robot autonome ?

b) La portée normative des normes de responsabilité civile

De cette manière, toute la mise en œuvre de la responsabilité civile pour encadrer le développement de l'IA et de la robotique s'effectue par l'entremise du juge, qui déterminera au cas par cas l'application des différentes règles juridiques. Compte tenu du rôle qu'occupe le processus judiciaire pour déterminer si une indemnisation peut être accordée à la victime, le fardeau de la preuve acquiert une importance capitale. En effet, c'est en se basant sur les différentes expertises et les témoignages déposés en preuve devant lui que le juge pourra notamment apprécier – selon les circonstances ayant mené la victime à subir un préjudice – dans quelle mesure la décision prise par l'IA pourrait entraîner ou non la responsabilité de son fabricant ou de son utilisateur. C'est pour cette raison que le poids de la preuve est aussi important, car il incombera à la victime de démontrer que l'utilisateur ou le fabricant pouvait raisonnablement prévoir les risques de préjudice en vertu des différentes obligations leur incombant.

64. D.C. Vladeck (2014), « Machines Without Principals: Liability Rules and Artificial Intelligence », *Washington Law Review*, vol. 89, p. 145.

À l'image des sanctions prévues par le droit réglementaire, le risque pour le fabricant ou l'utilisateur d'un robot de voir sa responsabilité engagée à la suite d'une décision judiciaire peut fortement l'inciter à prendre certaines précautions relativement à la fabrication ou à l'utilisation d'un dispositif incorporant une IA. Ce n'est cependant qu'une fois le jugement rendu par une cour de justice qu'il devra verser une compensation financière à la victime sous forme de dommages-intérêts. Outre la difficulté liée au fardeau de la preuve, le processus judiciaire est souvent plombé par de nombreux délais et coûts, qui peuvent autrement limiter pour la victime l'obtention d'une compensation pour le préjudice subi⁶⁵.

En définitive, que peut-on retenir du recours au droit dur pour encadrer le développement de l'IA et de la robotique ? Dans la mesure où les normes du droit réglementaire et de la responsabilité civile sont invoquées pour faire face aux impacts de l'IA et de la robotique, le droit dur offre une protection majoritairement axée sur les enjeux de santé et de sécurité lorsqu'il prend en considération le produit – tel qu'un système décisionnel intégrant une IA ou un robot interactif. Pour sa part, l'enjeu de la vie privée demeure cantonné dans une considération principalement accordée aux usages, laquelle interpelle la fonction réparatrice qu'offre la responsabilité civile. Dans les deux cas, tant l'encadrement sur le produit que ses différents usages mobilisent une gouvernance du développement de l'IA et de la robotique essentiellement réactive. Ce n'est qu'une fois le dispositif technologique conçu et commercialisé que le droit dur se met en œuvre, en développant des normes réglementaires lui étant plus particulièrement dédiées ou encore en imposant une compensation à la suite de préjudices qu'il a occasionnés.

Compte tenu de l'étendue et de la portée normative des règles juridiques qu'il propose, le droit dur ne semble pas offrir actuellement un encadrement lui permettant d'anticiper de manière préventive l'ensemble des impacts soulevés par le développement de l'IA et de la robotique. Bien au contraire, seuls quelques enjeux bien précis sont directement ciblés par les normes du droit dur. Dans cette perspective, l'encadrement du droit dur permet de mobiliser la force publique pour assurer le respect des normes établies, mais demeure limité à réagir en amont aux impacts de l'IA et de la robotique. Doit-on conclure que la gouvernance par le droit demeurera continuellement en retrait relativement aux avancées du développement de l'IA et de la robotique ?

Les deux autres formes souple et molle du droit proposent pour leur part de répondre à certaines limites du droit dur. La prochaine section présentera

65. P.-C. Lafond (2008), « Le consommateur et le procès – Rapport général », *Les Cahiers de Droit*, vol. 49, n° 1, p. 131-157.

plus amplement le rôle que celles-ci pourraient jouer dans la gouvernance du développement de l'IA et de la robotique.

Section 3. Le recours aux autres formes du droit pour encadrer le développement de l'IA et de la robotique

Tout comme nous l'avons fait pour le droit dur, nous reprendrons pour la prochaine section la même méthode d'analyse pour présenter l'étendue et la portée normative des dispositifs du droit souple (3.1) et mou (3.2). De cette manière, cette seconde section offrira une brève vue d'ensemble du portrait qu'offre la gouvernance par le droit pour encadrer le développement de l'IA et de la robotique.

3.1 Le droit souple et la normalisation technique

a) L'étendue des dispositifs du droit souple

Comme l'a souligné le Conseil d'État français dans son avis annuel de 2013, le droit souple peut parfaitement s'imbriquer directement dans le droit dur. Autrement dit, si les normes souples ne sont pas *a priori* des normes juridiques comme telles, celles-ci peuvent toutefois être utilisées par le droit dur pour préciser le contenu de certaines exigences. À cet égard, les normes techniques représentent une excellente illustration. En effet, la normalisation technique « a pour objet de fournir des documents de référence comportant des solutions à des problèmes techniques et commerciaux concernant les produits, biens et services qui se posent de façon répétée dans des relations entre partenaires économiques, scientifiques, techniques et sociaux⁶⁶ ». Ces documents de référence – ou normes techniques – comportent ainsi différentes spécifications techniques permettant de décrire, selon l'état actuel des connaissances scientifiques, les caractéristiques de différents biens et services. Située à la jonction des domaines techniques, scientifiques et économiques, la normalisation permet surtout de représenter une référence commune sur laquelle s'entend l'industrie ou les différents acteurs d'un domaine d'activité⁶⁷. Conséquemment, les normes techniques sont d'application volontaire. Les normes techniques telles que les normes ISO, développées à la suite de travaux résultant d'un processus consultatif et participatif par des groupes d'experts internationaux, sont périodiquement révisées. Ce processus assure ainsi la mise à jour

66. Centre de ressources et d'information sur l'intelligence économique, « Normalisation », <<https://portail-ie.fr/ressource/glossary/8/normalisation>>.

67. F.M. Lanord (2005), « La norme technique: une source de droit légitime? », *Revue française de droit administratif*, n° 4, juillet 2005, p. 738-751.

de l'état des connaissances techniques et scientifiques pour les différents secteurs d'activité commerciale visés par les normes pertinentes.

De cette manière, la normalisation cible principalement les enjeux de santé et de sécurité du développement de l'IA et de la robotique. Par exemple, la norme *ISO 13482:2014 – Robots et composants robotiques – Exigences de sécurité pour les robots de soins personnels*⁶⁸ énonce certaines exigences techniques pour la conception de robots de soins personnels permettant de réduire les risques et les dangers inhérents liés à l'utilisation de ce type de robots d'assistance. Parmi les différentes exigences de sécurité retrouvées dans la norme, mentionnons celles relatives aux systèmes de commande du robot pour : son arrêt (d'urgence ou de protection), sa vitesse limite, sa stabilité et sa force relative pour saisir ou déplacer un corps humain ou d'autres objets⁶⁹. En l'absence de normes réglementaires spécifiques à la robotique d'assistance, la normalisation technique établit néanmoins des balises pour les fabricants, leur permettant de fabriquer – dès la phase de conception – des robots sécuritaires et destinés à offrir des soins personnels. En fournissant ainsi des repères sur les dangers possibles que peut présenter l'utilisation d'un robot d'assistance, la normalisation technique offre une prise en compte plus précoce des impacts du développement de l'IA et de la robotique sur les enjeux de santé et de sécurité.

Grâce au mécanisme d'incorporation par renvoi, les autorités publiques peuvent directement intégrer les normes techniques dans le droit réglementaire. En effet, ce mécanisme permet d'incorporer dans un règlement une référence directe à une ou plusieurs normes techniques, intégrant du même coup tout le contenu qui y est énoncé à même la législation ou le règlement. De cette manière, la normalisation peut contribuer à rendre le droit dur plus flexible et préventif, en évitant les délais parfois très longs des processus usuels de modification législative ou réglementaire. Dans un contexte où certaines technologies comme l'IA et la robotique se développent à grande vitesse, les normes techniques comme la norme ISO 13482 pour la robotique d'assistance offrent aux autorités publiques l'opportunité de mieux ajuster le contenu des règlements – à vocation plus générale – aux spécificités présentées par certains dispositifs particuliers comme les robots de soins personnels.

68. Organisation internationale de normalisation, « ISO 13482:2014 », <<https://www.iso.org/fr/standard/53820.html>>. Comme le décrit l'ISO, la « présente Norme internationale décrit les phénomènes dangereux associés à l'utilisation de ces [robots d'assistance à la personne mobiles, robots d'assistance physique et robots de transport de personne] et fournit des exigences pour éliminer ou réduire les risques associés à ces phénomènes dangereux. La présente Norme internationale traite des applications avec contact physique homme-robot. »

69. S. Oliveira (2016), *La responsabilité civile dans les cas de dommages causés par les robots d'assistance*, mémoire de maîtrise, Montréal : Université de Montréal, p. 103-104.

La responsabilité civile peut elle aussi recourir au droit souple pour orienter la conduite des acteurs du domaine de la robotique et de l'IA. Dans le cadre d'un litige en matière de défaut de sécurité, un juge peut ainsi se référer au contenu d'une norme technique pour vérifier si la conception ou la fabrication d'un robot a bel et bien respecté les exigences de sécurité qui auraient dû l'être⁷⁰. La norme ISO 13482 pourrait être invoquée dans le cadre d'un procès en responsabilité civile, en vertu duquel le juge se retrouverait à vérifier si le préjudice causé par un robot d'assistance provient d'une conception ou d'une fabrication défectueuse.

b) La portée normative des normes du droit souple

A priori, les normes de droit souple n'ont aucune valeur contraignante. Néanmoins, malgré leur absence d'obligation et de sanction juridiques, celles-ci peuvent éventuellement faire l'objet d'une forme de reconnaissance par le droit dur. Dès lors que le droit dur leur accorde une forme d'attention, les normes du droit souple voient leur portée normative susceptible d'être modulée en fonction du degré de reconnaissance⁷¹. Autrement dit, tout dépend de la manière dont le droit dur intègre le droit souple dans ses différents instruments. La reconnaissance la plus formelle est celle de nature *directe*, employée par le procédé réglementaire d'incorporation par renvoi⁷². En faisant explicitement référence dans un règlement à l'obligation de respecter une norme technique particulière, le droit souple devient dès sa reconnaissance pleinement obligatoire.

La reconnaissance du droit souple par le droit dur peut aussi être de nature *indirecte*, lorsqu'un juge recourt par exemple à une norme technique pour déterminer la présence d'un défaut de sécurité⁷³. Dans d'autres cas, le juge peut également se baser sur des lignes directrices ou des codes de conduite pour déterminer si un comportement fautif et répréhensible respectait les règles de l'art en matière de responsabilité civile ou professionnelle⁷⁴. Ces deux types d'exemples illustrent toute la subtilité de la portée normative du droit souple. À quel point une norme technique ou une ligne directrice invoquée par le juge pour apprécier un défaut de sécurité ou interpréter une conduite doivent-elles

70. Pour une illustration jurisprudentielle, voir à ce sujet l'arrêt de la Cour d'appel du Québec *Manac inc./Nortex c. The Boiler Inspection and Insurance Company of Canada*, 2006 QCCA 1395.

71. Conseil d'État (2013), *Le droit souple*, *op. cit.*, p. 70-71.

72. J. Desjardins et J. Legault (1991), «L'incorporation par renvoi dans l'exercice du pouvoir réglementaire à l'échelon fédéral», *Revue du Barreau canadien*, vol. 70, n° 2, p. 245.

73. Voir par exemple *Camirand c. Baldor Electric Company*, 2010 QCCS 2621.

74. *Imbert de Friberg c. Bouchard* 2010 QCCS 1703; *Audet c. Landry* 2009 QCCS 3312.

être respectées, une fois le jugement rendu ? Celles-ci n'acquièrent pas une force obligatoire comme telle, mais demeurent toutefois susceptibles d'être invoquées de nouveau dans une situation similaire. Doivent-elles faire l'objet d'une attention plus importante de la part de ceux à qui elles s'adressent ? La reconnaissance indirecte du droit souple par un juge mobilise en quelque sorte une certaine *force normative*⁷⁵, de manière à inciter – plutôt qu'à obliger – les destinataires visés à se conformer aux normes retrouvées dans l'instrument pris en compte.

Cet effet incitatif du droit souple n'est toutefois effectif que dans la mesure où les parties prenantes adhèrent au contenu des normes qu'il propose. Tel est le cas des normes techniques. La mise en œuvre de la normalisation, d'application volontaire, dépend principalement de son processus collaboratif d'élaboration : les normes techniques sont élaborées par et pour les parties prenantes auxquelles elles se destinent. Leur légitimité s'incarne ainsi dans les différents avantages qu'en retirent leurs destinataires, tels que l'harmonisation des marchés internationaux et l'accroissement de la compétitivité économique des entreprises les adoptant⁷⁶. À l'inverse, l'adhésion des parties prenantes peut être limitée notamment par un processus collaboratif déficient ou des conflits d'intérêts. L'accessibilité aux normes techniques peut également constituer un obstacle important, compte tenu des coûts exigés pour obtenir la documentation pertinente. De cette manière, la reconnaissance de l'importance d'adhérer au contenu proposé par les normes de droit souple représente le principal critère permettant d'assurer sa mise en œuvre.

3.2 *Le droit mou et les enjeux de nature éthique, légale et sociale (ELS)*

a) L'étendue des dispositifs du droit mou

Nous l'avons évoqué plus haut, le droit mou regroupe sous sa houlette une diversité d'instruments normatifs qui ne comportent aucune obligation ni sanction juridiques. Le droit mou n'est toutefois pas une catégorie fourre-tout pour autant : deux autres éléments caractérisent plus précisément les normes qu'il contient. D'une part, le droit mou doit comporter une fonction « normatrice »⁷⁷, qui le destine à orienter, baliser ou guider le comportement des personnes auxquelles il s'adresse. D'autre part, les normes de droit mou doivent avoir été élaborées suivant un certain degré de formalisme, c'est-à-dire

75. C. Thibierge (2009), « Conclusion. Le concept de "force normative" », dans C. Thibierge (dir.), *La force normative. Naissance d'un concept*, Paris : LGDJ, p. 816.

76. V. Lasserre (2015), *Le nouvel ordre juridique. Le droit de la gouvernance*, Paris : LexisNexis, p. 240 ; N. Nevejans (2017), *Traité de droit et d'éthique de la robotique civile*, Bordeaux : LEH Édition, p. 285.

77. Conseil d'État (2013), *Le droit souple, op. cit.*, p. 56.

en vertu d'un processus minimal d'élaboration normative. Le langage employé dans l'instrument normatif doit pouvoir s'apparenter au langage juridique : l'exhortation d'un ministre à suivre une conduite donnée ne relève pas du droit mou, mais bien de la politique⁷⁸. En revanche, cette exhortation prendra la forme du droit mou si elle adopte la forme d'une recommandation de bonne pratique⁷⁹. Bien qu'il n'existe pas de délimitation claire distinguant le non-droit du droit mou, l'apparence et la formulation du contenu énoncé représentent des critères utiles pour déterminer si un document se rattache au droit mou.

Si les normes techniques de droit souple intégrées dans la réglementation ciblent principalement les enjeux de santé et de sécurité, les normes de droit mou visent pour leur part un spectre d'enjeux beaucoup plus large. La *Déclaration de Montréal sur le développement responsable de l'intelligence artificielle*⁸⁰ constitue à cet égard une excellente illustration. Lancée en 2018 à l'initiative de l'Université de Montréal et élaborée dans un contexte de coconstruction regroupant des chercheurs universitaires et des membres du grand public, cette déclaration est une « œuvre collective qui a pour objectif de mettre le développement de l'IA au service du bien-être de tout un chacun, et d'orienter le changement social en élaborant des recommandations ayant une forte légitimité démocratique⁸¹ ». Pour ce faire, la *Déclaration de Montréal* présente dix principes articulés autour des valeurs du bien-être, de l'autonomie, de l'intimité et de la vie privée, de la solidarité, de la démocratie, de l'équité, de l'inclusion, de la prudence, de la responsabilité et de la soutenabilité environnementale⁸². Ainsi, ces valeurs se rattachent aussi bien aux enjeux de sécurité et d'environnement qu'à ceux de nature éthique, légale et sociale (enjeux ELS), notamment la liberté de choix, la vie privée ou le vivre-ensemble. Pour mieux illustrer l'ampleur des enjeux que vise la perspective collective de la *Déclaration de Montréal*, voici les dix principes qu'elle énonce :

- 1) **Principe de bien-être** : Le développement et l'utilisation des systèmes d'intelligence artificielle (SIA) doivent permettre d'accroître le bien-être de tous les êtres sensibles.

78. *Ibid.*, p. 62-63.

79. *Ibid.*

80. *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, <<https://www.declarationmontreal-iaresponsable.com/la-declaration>>.

81. Déclaration de Montréal IA responsable (2018), *Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, p. 25, <https://docs.wixstatic.com/ugd/ebc3a3_d806f109c4104c91a2e719a7bef77ce6.pdf>.

82. *Id.*

- 2) **Principe d'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 solidarité:** Le développement de SIA doit être compatible avec le maintien de liens de solidarité entre les personnes et les générations.
- 5) **Principe de participation démocratique:** Les SIA doivent satisfaire les critères d'intelligibilité, de justifiabilité et d'accessibilité, et doivent pouvoir être soumis à un examen, un débat et un contrôle démocratiques.
- 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 liées à 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 de SIA doivent se réaliser de manière à assurer une soutenabilité écologique forte de la planète.

Dans la même optique, le Parlement européen a suggéré l'élaboration d'une Charte éthique de la robotique dédiée aux ingénieurs, aux concepteurs et aux chercheurs dans la Résolution 2015/2103(INL) adressée en février 2017 à la Commission européenne sur la création de nouvelles règles de droit civil pour la robotique. Ce code de conduite jetterait ainsi les « bases de l'identification, de la surveillance et du respect de principes éthiques fonda-

mentaux dès la phase de conception et de développement⁸³». Le document aurait comme objectif principal de « répondre à la nécessité de veiller au respect des normes éthiques », tout en introduisant « un processus de recherche de solution pour résoudre les dilemmes éthiques et permettre à ces systèmes [robotisés] de fonctionner de manière responsable d'un point de vue éthique »⁸⁴. Parmi les principes que les ingénieurs, chercheurs et concepteurs de robots devraient respecter, l'ébauche proposée souligne l'importance devant être accordée aux droits fondamentaux, aux principes de bienfaisance, non-malfaisance, autonomie et justice, aux principes de précaution, de transparence et d'accès à l'information, à la considération et au respect de la santé, la sécurité, du bien-être et des droits de la personne, de même qu'à la protection de la vie privée⁸⁵.

b) La portée normative des normes du droit mou

À l'instar de la normalisation technique, toute la mise en œuvre du droit mou dépend strictement de l'adhésion des parties prenantes, étant donné qu'il ne comporte aucune obligation ni sanction juridiques. La reconnaissance qu'accordent les destinataires à l'importance de respecter le contenu des normes de droit mou est donc primordiale. Deux éléments peuvent nous aider à mieux apprécier cette importance qu'occupe l'adhésion dans la mise en œuvre du droit mou. Le premier élément consiste à examiner de quelle manière s'est déroulé le processus de création de l'instrument de droit mou. A-t-il fait l'objet d'un certain consensus? Permet-il de rejoindre les intérêts de l'ensemble des parties prenantes visées par l'instrument? Les détails fournis dans le rapport sur la *Déclaration de Montréal* permettent de mieux examiner de quelle manière s'est déroulé son processus d'élaboration. Qualifiée de « coconstruction inclusive », la démarche a préconisé une participation publique, délibérative et itérative de l'ensemble de la société et la mise en place d'un dialogue entre des citoyens, des professionnels et des experts pour favoriser sa légitimité démocratique⁸⁶. En effet, le processus consultatif s'est effectué sur la base d'une série

83. *Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL))*, <http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_FR.html>.

84. *Ibid.*

85. *Ibid.*

86. Dans son rapport, la Déclaration de Montréal indique qu'elle « est adressée à toute personne, toute organisation de la société civile et toute compagnie désireuses de participer au développement de l'intelligence artificielle de manière responsable, que ce soit pour y contribuer scientifiquement et technologiquement, pour développer des projets sociaux, pour élaborer des règles (règlements, codes) qui s'y appliquent, pour pouvoir en contester les orientations mauvaises ou imprudentes, ou encore pour être en mesure de lancer des alertes à l'opinion

de consultations publiques, tables ou cafés de citoyens et journées de coconstruction, où tous étaient invités à réfléchir et à débattre au sujet d'enjeux prospectifs liés à des scénarios hypothétiques basés sur l'intégration de systèmes d'IA dans différents secteurs de la société : voitures autonomes et transport intelligent, emplois et monde du travail, santé prédictive, ville intelligente, police prédictive et système judiciaire et éducation prédictive⁸⁷. Une telle démarche participative offre ainsi d'accroître la légitimité de la *Déclaration de Montréal*, ce qui pourrait se traduire par une adhésion plus forte par les parties prenantes qu'elle vise relativement à sa mise en œuvre.

Le second élément rattaché à l'importance de l'adhésion du droit mou concerne la mise en place de mécanismes assurant un suivi ou une coordination des parties prenantes visés par l'instrument normatif. En ce sens, la présence de comités ou d'autres dispositifs chargés d'assurer une bonne coordination entre les différents acteurs peut jouer un rôle-clé dans la mise en œuvre du droit mou. À titre d'exemple, les comités d'éthique clinique et organisationnelle (CÉCO) occupent un rôle comme dispositif institutionnel pour sensibiliser ou conseiller les intervenants œuvrant dans un établissement de santé et de services sociaux par rapport à différents dilemmes éthiques pouvant se poser dans le cadre des services rendus⁸⁸. Sans détenir de pouvoir décisionnel, ce comité émet des avis qui servent uniquement à soutenir la réflexion éthique entreprise tandis que la décision finale revient toujours au demandeur ayant soumis le problème à résoudre.

Sans instituer explicitement un tel dispositif de mise en œuvre dans son énoncé normatif, la *Déclaration de Montréal* établit toutefois dans sa liste de recommandations destinées au développement de politiques publiques la mise en place d'un organisme de veille et de recherche sur les usages et les impacts sociétaux du numérique et de l'IA⁸⁹. Plutôt que de veiller spécifiquement à l'application de la *Déclaration de Montréal*, cet organisme de veille se verrait attribuer un rôle beaucoup plus large, en instituant un processus de veille

publique quand cela est nécessaire. Elle s'adresse également aux responsables politiques, élus ou nommés, dont les citoyens attendent qu'ils prennent la mesure des changements sociaux en gestation, qu'ils mettent en place rapidement les cadres permettant la transition numérique pour le bien de tous, et qu'ils anticipent les risques sérieux que présente le développement de l'IA.» Voir *Déclaration de Montréal IA responsable* (2018), *Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, *op. cit.*, p. 6.

87. *Ibid.*, p. 76.

88. Centre intégré universitaire de santé et de services sociaux de la Mauricie-et-du-Centre-du-Québec, « Comité d'éthique clinique et organisationnelle », <<https://ciussmq.ca/a-propos-de-nous/comites-et-instances/ethique-clinique-et-organisationnelle>>.

89. *Déclaration de Montréal IA responsable* (2018), *Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, *op. cit.*, p. 311.

continue, en favorisant la création de nouvelles connaissances sur les aspects techniques, éthiques, juridiques et sociaux des systèmes d'IA, en mobilisant les citoyens et les parties prenantes pour préciser les attentes sociales relatives au développement de l'IA et en éclairant la décision publique. Si l'élaboration de la *Déclaration de Montréal* a bénéficié d'un processus de coconstruction inclusive en impliquant l'ensemble des acteurs sociaux, les détails de sa mise en œuvre demeurent pour l'instant moins explicites.

Précisons qu'avec la grande diversité d'instruments pouvant se retrouver dans la catégorie du droit mou, le rôle et la nature des dispositifs d'accompagnement sont susceptibles de varier considérablement. Par ailleurs, le droit mou peut toujours bénéficier d'une forme de reconnaissance par le droit dur, de façon à ce qu'il se transforme en droit souple. Par exemple, la proposition de Charte éthique de la robotique mentionnée dans l'Annexe de la Résolution 2015/2103(INL) du Parlement européen ne comporte aucune information relative à un dispositif de mise en œuvre. Cependant, la Résolution 2015/2103(INL) mentionne la possibilité que la Charte éthique de la robotique puisse être mobilisée dans le contexte de la responsabilité civile. En effet, dans le paragraphe 51 de sa Résolution, le Parlement européen «demande à la Commission de présenter, sur la base de l'article 114 du traité sur le fonctionnement de l'Union européenne, une proposition d'instrument législatif sur les aspects juridiques du développement et de l'utilisation de la robotique et de l'intelligence artificielle à un horizon de 10 ou 15 ans, combinée à des instruments non législatifs, tels que des lignes directrices et des codes de conduite, comme mentionné dans les recommandations figurant en annexe». En ce sens, la possibilité pour la Charte éthique de la robotique de se voir reconnaître comme instrument de droit souple par le droit dur pourrait pallier le manque d'adhésion des destinataires qu'elle vise, dans l'éventualité où une telle situation est constatée.

CONCLUSION

Quel portrait pouvons-nous dresser de la gouvernance qui se profile à l'aube d'un développement croissant de l'intelligence artificielle et de la robotique? Que peut-on attendre du droit lorsque celui-ci est mobilisé comme outil sociétal pour encadrer l'essor de l'IA et son intégration dans une foule de secteurs d'activité économique? Comme d'autres technologies avant elles, l'IA et la robotique fondent bien des espoirs à titre de moteurs du développement économique et technologique. Face à l'ampleur des conséquences que le développement de ces deux technologies entraîne, la gouvernance par le droit est confrontée au dilemme de Collingridge dans sa prise d'action. Peut-il intervenir en amont d'une diffusion massive de systèmes intégrant l'IA pour prévenir les

impacts néfastes anticipés, ou bien est-il condamné à attendre que ces impacts se réalisent avant de mettre en place des mesures les encadrant ? L'illustration du principe de précaution démontre bien l'ampleur des défis guettant la gouvernance juridique pour encadrer l'ampleur des impacts entraînés par le développement technologique, tout en tenant compte de l'importance accordée aux bénéfices économiques.

C'est dans ce contexte que cette contribution a proposé de déterminer et d'examiner les dynamiques des trois formes de droit à l'œuvre dans la gouvernance juridique du développement de l'IA et de la robotique, pour mieux saisir la nature de l'encadrement qu'offre le droit à l'égard de ces technologies. Au-delà de la simple considération d'un droit dur axé sur le recours à la contrainte et à la sanction, la gouvernance par le droit révèle plutôt un éventail de nuances dans sa manière d'orienter et de baliser la conduite des différents acteurs impliqués dans le développement de l'IA et de la robotique. Face aux multiples impacts qu'entraînent l'IA et la robotique, les trois formes de droit ciblent chacune différents enjeux affectés par leur développement.

D'un côté, les formes dure et souple du droit entendent encadrer les impacts négatifs soulevés par l'IA et la robotique à l'égard de la santé et de la sécurité, par le biais du droit réglementaire et de la responsabilité civile. Si les mécanismes propres au droit dur lui imposent d'agir une fois la preuve des risques présentés ou des dommages causés par ces technologies suffisamment étouffée, le droit souple propose pour sa part d'adopter une posture un peu plus préventive. En ce sens, la normalisation technique offre ainsi la possibilité de définir plus rapidement – et par le biais d'un consensus – les exigences sanitaires et sécuritaires nécessaires pour protéger le consommateur. La possible reconnaissance des normes techniques par le droit dur offre par ailleurs une certaine garantie du respect des exigences énoncées. De cette manière, le droit dur et le droit souple offrent tous deux une prise en compte majoritairement axée sur les impacts du dispositif ou du système intégrant une IA.

De l'autre côté, le droit mou est interpellé d'une manière beaucoup plus large à l'égard des enjeux de santé et de sécurité, mais également ceux de nature éthique, légale et sociale, comme la liberté de choix ou le vivre-ensemble. Les principes énoncés dans la *Déclaration de Montréal sur le développement responsable de l'intelligence artificielle* illustrent d'ailleurs la diversité des impacts qu'elle entend encadrer. Bien que la question des usages soit notamment encadrée par le droit dur relativement à l'enjeu de la vie privée, le droit mou offre une prise en compte plus affirmée des impacts découlant des différentes utilisations de l'IA dans la société. Dans l'univers normatif du droit mou, cette prise en compte des usages de l'IA et de la robotique dépend directement de l'adhésion que peut susciter un instrument comme la *Déclaration de Montréal*. Compte tenu

de l'absence d'obligation et de sanctions juridiques, une attention particulière doit ainsi être accordée au processus d'élaboration normative, de même qu'à la mise en place d'un dispositif permettant d'assurer minimalement une forme de suivi de ses énoncés.

Vu l'ampleur de la diversité d'usages qu'offre actuellement ou promet l'essor des systèmes d'IA, il n'est pas surprenant de constater un appel grandissant à la mise en place d'un accompagnement visant à favoriser un développement responsable de l'innovation en matière d'intelligence artificielle et de robotique⁹⁰. Dans ce contexte, de plus en plus d'instruments normatifs destinés à prendre en compte les usages de l'IA et de la robotique émergent d'initiatives de nature publique et/ou privée, tels le *Code of conduct for data-driven health and care technology* adopté par le gouvernement du Royaume-Uni⁹¹ ou le guide *Ethically Aligned Design* proposé par l'IEEE⁹². Face au risque d'empiéter les uns sur les autres et de voir ainsi leur portée normative affaiblie, l'émergence de tels instruments normatifs interpelle la poursuite de la réflexion amorcée par cette contribution au sujet du rôle, des forces et des limites que peut jouer le droit mou dans la gouvernance juridique du développement de l'IA et de la robotique.

90. Voir par exemple Marchant, G. (25 janvier 2019), «“Soft Law” Governance of Artificial Intelligence», *AI PULSE*, <<https://aipulse.org/soft-law-governance-of-artificial-intelligence>>.

91. Department of Health and Social Care (18 juillet 2019), *Code of conduct for data-driven health and care technology*, Gov.UK, <<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>>.

92. Institute of Electric and Electronic Engineers (2018), *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, version 2, <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf>.

CHAPITRE 8

La personne transformée : comment peut-on mettre une limite ?

Jean-Pierre Béland

Au cours des dernières décennies, des domaines de recherche impliquant les nanotechnologies, les biotechnologies, les technologies de l'information et les sciences cognitives (NBIC) ont vu le jour, donnant naissance à des produits, tel l'interface cerveau-machine, visant l'humain augmenté¹. N'y a-t-il pas là un appel à la vigilance ? Le droit se doit-il d'assurer une forme de protection de la personne humaine ? Plusieurs questions se posent.

C'est au cœur de cette problématique, lors du colloque international sur *La personne transformée* tenu à l'Université catholique de Lyon en juin 2018, que s'est ainsi logée la question philosophique de l'évaluation de la transformation humaine par le développement des technologies convergentes « NBIC »². Cette question m'a été adressée par un professeur en droit et organisateur du colloque, pour savoir au nom de quels grands principes et valeurs relevant de l'éthique il serait possible d'interdire, de limiter ou de favoriser ce développement :

-
- 1 « Nano Brain Implant », Mindtech Research, <<https://nanobrainimplant.com/2018/02>>.
 2. Le terme « NBIC » signifie que certains scientifiques étudient les possibilités de la convergence des nanotechnologies, des biotechnologies, des technologies de l'informatique et des sciences cognitives pour l'augmentation des performances physiques ou cognitives de l'être humain. Voir Roco, M., Bainbridge, W.S. (2003), *Converging Technologies for Improving Human Performance: Nanotechnology, biotechnology, information technology and cognitive science*. Dordrecht: Kluwer Academic Publishers.

La transformation humaine annoncée par le développement des technologies convergentes, telles que les *nanotechnologies*, les *biotechnologies*, les technologies de l'informatique et les sciences cognitives (*NBIC*), celle qui se produit progressivement devant nous pourrait-elle, doit-elle être remise en cause, limitée, freinée ou bien, pourquoi pas, favorisée, ceci au prisme de grands principes et des valeurs relevant de l'éthique? (Franck Violet)

Cette question me semble des plus pertinentes, puisqu'elle se pose partout, incluant différentes instances gouvernementales, comme la CEST (Commission de l'éthique de la science et de la technologie) et l'INESSS (Institut national d'excellence en santé et en services sociaux) au Québec; le CCNE (Comité consultatif national d'éthique) et le HAS (Haute Autorité de Santé) en France. Cette question est aussi au cœur des approches des entreprises visant le développement des usages responsables des technologies dans la société. Cette question, telle que formulée, présuppose :

- I. que la transformation humaine est l'impact produit par la convergence des NBIC ;
- II. que cet impact peut être évalué à la lumière de grands principes et de valeurs, de sorte que cette évaluation conduit à la décision d'interdire, de limiter ou de favoriser la transformation humaine par le développement des NBIC.

En quoi cette question est-elle difficile à répondre? Cette question suppose un processus d'évaluation qui conduira à une décision. Le problème théorique à surmonter est de savoir comment parvenir à une telle décision : par quel processus d'évaluation peut-on arriver à une décision d'interdire, de freiner ou encore de favoriser la transformation humaine? Et quel type d'évaluation choisir : l'évaluation morale, l'évaluation juridique ou l'évaluation éthique? Chacun de ces types d'évaluation nous renvoie à des conclusions différentes : la décision morale est de l'ordre de l'interdiction ; la décision juridique est de l'ordre de la régulation (interdire, limiter) ; et la décision éthique est de l'ordre de la meilleure chose à faire (interdire, limiter ou favoriser). À cela s'ajoute la difficulté d'identifier les principes et les valeurs ainsi que leurs rôles dans l'évaluation (morale, juridique ou éthique). En résumé, tous ces problèmes convergent vers une double difficulté : 1) comprendre la complexité du processus d'évaluation d'une technologie qui conduit à la conclusion d'interdire, de limiter ou de favoriser la transformation humaine, et 2) comprendre les difficultés d'opérationnaliser ce processus pour chacun des cas de transformation humaine dont il est question.

À partir de quel cadre théorique pouvons-nous résoudre ces deux problèmes de compréhension? Depuis 2012, notre groupe de recherche interdisciplinaire (Groupe de recherche InternE³LS) se penche sur les enjeux

E³LS (enjeux économiques, environnementaux, éthiques, légaux et sociaux) et utilise une grille conceptuelle sous forme de « Processus réflexif d'analyse d'impact et d'acceptabilité »³ qui est applicable dans l'évaluation de tous les développements technologiques. Ce processus d'évaluation d'une technologie renvoie à trois moments subséquents que nous présentons dans le tableau suivant :

Processus d'analyse d'impact et d'acceptabilité :

Moment 1 : La détermination des impacts sur des enjeux

Étape 1 : Identification de la source technologique pouvant avoir un impact sur un enjeu

Étape 2 : Identification d'un enjeu pouvant subir un impact de la source

Étape 3 : Détermination de l'impact réel ou négatif de la source sur l'enjeu

Moment 2 : L'évaluation des impacts à partir des valeurs retenues

Étape 1 : Qualification des impacts sur des enjeux en termes de valeur

Étape 2 : Jugement final d'évaluation d'un impact positif ou négatif

Moment 3 : La pondération des jugements finaux d'évaluation en vue de la décision

Étape 1 : Détermination du type de pondération retenue : acceptabilité des risques ou acceptabilité globale des impacts ?

Étape 2 : Processus de pondération a) selon l'acceptabilité des risques ou b) selon l'acceptabilité globale des impacts

Voyons, en résumé, comment dégager un processus d'évaluation implicite dans ce cadre de référence. Le Moment 1 vise à déterminer l'impact du développement technologique. Par exemple, pour l'implant cochléaire (implant électronique), qui vise à fournir un certain niveau d'audition à des personnes atteintes de surdité profonde ou sévère et des personnes souffrant d'acouphènes, le premier impact est sur la capacité d'audition. Et, puisque cet impact physiologique modifie la façon de vivre des personnes qui bénéficient de l'implant, il peut y avoir plusieurs conséquences sur la vie de la personne et de celles qui l'entourent. C'est l'impact sur un enjeu. Chaque enjeu soulève déjà un critère

3. J.-P. Béland et G.A. Legault (2012), *Asimov et l'acceptabilité des robots*, Québec : Presses de l'Université Laval ; J. Patenaude *et al.* (2014), « Framework for the Analysis of Nanotechnologies' Impacts and Ethical Acceptability: Basis of an Interdisciplinary Approach to Assessing Novel Technologies », *Sci Eng Ethics*, DOI : 10.1007/s11948-014-9543-y ; J.-P. Béland, G.A. Legault et J. Patenaude (2014), « Les enjeux identitaires de l'humain dans le débat philosophique sur la robotique humanoïde et l'amélioration humaine », *BioéthiqueOnline*, p. 3-14, <<http://bioethiqueonline.ca/3/14>> ; J.-P. Béland (2016), « Enjeux éthiques du développement de la robotique », *Ethica*, 20, n° 1, p. 75-95.

pour évaluer l'impact. Ainsi, dans notre exemple, la santé, la communication, le bien-être de la personne, la qualité de ses rapports aux autres sont autant d'enjeux soulevés par l'implant cochléaire et indiquent la valeur qui servira de base à l'évaluation. L'évaluation (Moment 2) précisera jusqu'où l'impact diminue ou augmente la valeur de référence. Ainsi, l'implant peut augmenter de beaucoup la santé de la personne, mais diminuer la qualité de communication et de ses relations avec les personnes au sein de sa communauté de sourds ou de malentendants. La décision d'interdire, de limiter ou de favoriser l'implantation (Moment 3) devra trancher entre les différentes évaluations afin de prioriser certaines valeurs plutôt que d'autres. Cette priorisation d'une valeur plus qu'une autre dépend du principe qui fonde l'évaluation.

Selon ce cadre simplifié de référence, le processus d'évaluation conduisant à la décision d'interdire, de limiter ou de favoriser l'implantation débute par l'analyse d'impact qui permet de déterminer le critère servant à l'évaluation ainsi que la valeur correspondante qui permet de faire l'évaluation des impacts. Cette évaluation d'impact est gouvernée par un principe sur lequel sera fondé la décision finale.

Dans ce chapitre, le but est de montrer qu'il y a plusieurs principes et valeurs mobilisés dans l'évaluation de produits issus des NBIC. Il s'agira de démontrer ensuite comment les principes et valeurs s'articulent dans les différents discours normatifs : l'évaluation morale, l'évaluation juridique et l'évaluation éthique. Il s'agit enfin de montrer les difficultés liées à l'opérationnalisation de ces évaluations.

Pour atteindre ces trois objectifs, j'ai retenu quatre exemples de produits qui sont en voie de commercialisation ou en développement, issus de la convergence des NBIC :

Exemple 1 : L'implant de mémoire dans le domaine des BCI (Brain-Computer-Interfaces) en « neurotechnologie » ;

Exemple 2 : Les systèmes d'aide à la décision (Intelligence algorithmique) ;

Exemple 3 : Les implants RFID (Radio Frequency Identification Device) ;

Exemple 4 : Le robot d'assistance aux personnes âgées.

Les exemples 1 et 2 serviront à opérationnaliser les évaluations morales qui mobilisent les obligations de la représentation humaine (nature, intégrité et dignité) dans la décision d'interdire. L'exemple 3 servira à opérationnaliser l'évaluation juridique qui se rapproche du langage moral parce qu'elle mobilise des obligations, mais qui se distingue de la morale par sa capacité de contraindre (interdire, limiter) les personnes à exécuter leurs obligations selon une règle précise dans la décision. L'exemple 4 servira à l'évaluation éthique qui, contrairement à la morale et au droit, ne guide pas la décision par le biais d'obligations,

mais par le biais de valeurs partagées. Ces exemples serviront à illustrer la différence entre chacune de ces évaluations (évaluation morale, évaluation juridique et évaluation éthique) et la difficulté qu'on rencontre lorsqu'il faut les opérationnaliser.

1. IMPLANT DE MÉMOIRE DANS LE DOMAINE DES BCI EN « NEUROTECHNOLOGIE »

Quand on veut prendre une décision d'interdire, de limiter ou de favoriser la transformation humaine par le développement des NBIC, il convient de choisir un produit issu de ce développement qui est le plus susceptible de permettre cette transformation. Dans ce cas-ci, nous examinerons l'implant de mémoire issu du domaine des BCI (« Brain-Computer-Interfaces ») en « neurotechnologie ». Cet implant de mémoire servira d'exemple pour illustrer l'évaluation morale à partir de notre processus simplifié d'évaluation. Ce processus nous permettra de préciser les éléments de l'évaluation – i) l'impact, ii) le critère de l'évaluation, iii) la valeur correspondante et iv) le principe qui fonde la décision d'interdire, de freiner ou de favoriser – ainsi que les difficultés à opérationnaliser cette évaluation morale en fonction de l'interdiction du tout ou d'une partie de ce développement.

i) Impact. La première étape de l'évaluation vise à déterminer l'impact réel et potentiel du développement technologique des NBIC en déterminant d'abord quel est le type d'implant BCI que nous devons évaluer et comment il fonctionne pour la mémoire. Car les interfaces cerveau-ordinateur (ou BCI) ne sont pas toutes pareilles et évoluent depuis plusieurs années. Différentes formes de BCI sont déjà disponibles, allant de celles qui reposent sur la tête et mesurent les signaux du cerveau jusqu'à l'implant de mémoire BCI incorporé dans le tissu cérébral (*Nano Brain Implant*⁴).

Impact réel. Selon Mohseni, l'un des meilleurs exemples actuels d'innovations attendues pour une telle amélioration cognitive dans le domaine des BCI en « neurotechnologie » découle du travail du professeur Theodore Berger, de l'Université de Californie du Sud⁵. Celui-ci développe une prothèse de mémoire pour remplacer les parties endommagées de l'hippocampe chez des patients ayant perdu la mémoire comme ceux qui, par exemple, souffrent de la maladie d'Alzheimer : « Dans ce cas, un ordinateur doit être implanté dans le cerveau et agir de la même manière que l'hippocampe biologique, du point

4. Voir « Nano Brain Implant », *Mindtech Research*, <<https://nanobrainimplant.com/2018/02>>.

5. *Ibid.*

de vue des entrées et des sorties. Berger a des résultats à la fois sur des modèles de rongeurs et de primates non humains⁶. »

Est-ce que ce dispositif – une puce de silicium servant à imiter le traitement du signal que ces neurones effectuent quand ils fonctionnent correctement – peut réellement produire les résultats escomptés ? Les résultats montrent que la puce implantée dans le cerveau des singes « permet de mémoriser des expériences et informations plus d'une minute⁷ ». Tel est son impact réel. « En fin de compte, Berger veut rétablir la possibilité de créer des souvenirs à long terme par l'implantation de puces de ce type dans le cerveau [humain]⁸. » Mais on n'a pas encore pu démontrer de résultats provenant d'une puce de mémoire implantée dans le cerveau humain.

Impact potentiel. La *start-up* américaine KERNEL s'est ainsi lancée, en 2016, dans la fabrication « d'implants de mémoire » assistés par ordinateur. Musk (2018) nous permet ainsi de déterminer l'impact potentiel des recherches en BCI⁹. Ces recherches favoriseraient l'utilisation des implants BCI de manière bidirectionnelle, de sorte que le branchement puisse nous rendre plus intelligent que les humains qui les ont fabriqués. Même si, dans cette optique, la commercialisation attendue de l'implant ne vise pas à augmenter la capacité cognitive de la mémoire, mais bien de la maintenir dans son état malgré la maladie dégénérative, tout en conservant les nouveaux apprentissages, l'implant en développement dans le domaine des recherches en BCI pourrait être, selon Valeriani, un outil fondamental pour dépasser les limites humaines¹⁰. Nous serions alors en route vers l'« *amélioration cognitive humaine* ». Par exemple, en améliorant les capacités cognitives du cerveau humain par l'utilisation de machines nanotechnologiques encore plus poussées qui augmentent la vitesse des interfaces cerveau-ordinateur jusqu'à dépasser les limites biologiques humaines. L'impact positif potentiel retenu en fonction des auteurs recensés (Marsh, Mohseni et Valeriani en 2018) est donc :

L'amélioration de la mémoire chez les personnes atteintes de détérioration de la mémoire.

6. *Ibid.*

7. J. Cohen (2013), « Theodore Berger et les implants de mémoire. Vers des neuroprothèses ? », *MIT Technology Review*, <<https://transhumanistes.com/theodore-berger-et-les-implants-memoire-vers-des-neuroprotheses>>.

8. *Ibid.*

9. S. Marsh (2018), « Neurotechnology, Elon Musk and the goal of human enhancement », <<https://www.theguardian.com/technology/2018/jan/01/elon-musk-neurotechnology-human-enhancement-brain-computer-interfaces>>.

10. Voir « Nano Brain Implant », *Mindtech Research*, *op. cit.*

Difficulté. Aujourd'hui, nous ne disposons pas de preuves solides au sujet de l'amélioration de la mémoire chez les personnes atteintes de détérioration de la mémoire. Or l'évaluation d'une technologie repose sur la nature des études de mesure d'impact et la qualité de la preuve qu'elles fournissent. C'est ici qu'apparaît une première difficulté, celle d'opérationnaliser l'analyse d'impact : le débat sur la qualité des études et des preuves fournies. Il existe, ensuite, une seconde difficulté. Est-ce que nous devrions attendre des preuves certaines d'analyse d'impact avant de poursuivre l'évaluation ? Procéder ainsi risque de nous mettre devant un fait accompli et il serait impossible d'interdire ou de limiter le développement. C'est pourquoi l'évaluation d'une technologie repose sur les impacts réels lorsqu'ils ont été étudiés, mais aussi potentiels lorsqu'ils n'ont pas encore été étudiés, parce qu'ils sont probables en raison du dispositif implanté sur la personne.

ii. Critère. Une fois que nous avons isolé l'impact réel et potentiel dans le contexte du développement de l'implant BCI en neurobiologie, nous devons déterminer le critère qui va nous servir à évaluer l'impact (faire un jugement de valeur). Mais, avant de choisir un critère de jugement de valeur dans ce contexte, il faut identifier de quelles manières l'impact réel ou potentiel peut créer différentes conséquences positives ou négatives dans la vie de la personne, de son entourage et de la société. C'est ce que nous nommons l'enjeu :

Le terme « enjeu » [...] renvoie, selon les différentes définitions, à ce que l'on peut gagner ou perdre dans un jeu. [...] Autrement dit, nos actions provoquent des conséquences qui peuvent être positives ou négatives, et c'est ce qui est en jeu dans le contexte. [...] Ainsi, pour déterminer un enjeu éthique, il faut une analyse des conséquences qui découlent de l'activité et une évaluation de celle-ci¹¹.

Donc, quels sont les enjeux associés à l'implant de mémoire qui pourrait avoir des conséquences positives et négatives dans la vie de la personne, de son entourage et de la société ?

Enjeu majeur en médecine (M) : l'implant peut améliorer la condition clinique (rétablir la mémoire de la personne en ayant comme risque des effets biologiques ou des risques de complications possibles sur la constitution biologique du corps (enjeu sur le processus biologique fonctionnel et la constitution humaine).

Enjeu majeur en psychologie cognitive (PC) : l'implant, en améliorant la mémoire, risque de modifier la capacité d'oubli, qui a une fonction indispensable dans l'équilibre psychologique de la personne (enjeu sur le bien-être psychologique de la personne humaine).

11. G.A. Legault (2015), « Enjeux éthiques du développement technologique », *Éthica*, vol. 20, n° 1, Québec : Université du Québec à Rimouski, p. 54-55.

Enjeu majeur pour la qualité de vie des personnes (QV) : l'implant, en améliorant la mémoire chez les personnes souffrant de la maladie d'Alzheimer, favorise la qualité de vie de chacune de ces personnes (enjeu sur la qualité de vie comme capacité de penser, de prendre des décisions et de garder le contrôle sur sa vie quotidienne).

Enjeu majeur pour les activités économiques (AE) : l'implant serait nécessaire pour conserver une qualité de vie économique d'une population (enjeu sur l'utilité et le bien-être économique nécessaire au développement humain).

Enjeu majeur pour la vie en société (VS) : si l'implant, en augmentant la capacité de mémoire, donne un avantage à la personne sur les autres (proches), alors se pose la question de la qualité des relations aux autres (enjeu sur le vivre-ensemble avec les autres)

Enjeu majeur en politique (P) : l'implant peut réduire les coûts en santé (enjeu sur l'efficacité et l'efficience des soins)

Enjeu majeur en morale (M) : si l'implant, en améliorant la mémoire, provoque la disparition des frontières entre l'humain et l'ordinateur (la machine), de sorte qu'il n'y a plus qu'une différence de degrés entre la vie artificielle et la vie réelle du corps biologique, alors se pose la question de la transformation de la condition biologique humaine (enjeu sur la transformation de la condition biologique – identité – nature – de la personne).

Chacun de ces enjeux propose un critère permettant de faire un jugement de valeur sur l'impact positif ou négatif potentiel de l'implant de mémoire. Ainsi, en retenant l'impact sur la condition clinique, nous privilégions la santé comme valeur de référence. Choisir l'impact sur la façon d'être de la personne dans sa vie de tous les jours comme critère renvoie à la valeur de la qualité de vie. Si on prend l'impact sur la transformation de la condition biologique comme critère, l'évaluation se fera à partir de la valeur de l'intégrité de la personne humaine.

Comment s'opérationnalise le choix du critère d'évaluation ? Le choix peut sembler arbitraire, mais il est en fait guidé par un « Principe » qui fonde l'évaluation faite par la personne. Par exemple, dans l'« Évaluation des technologies en santé » (HTA), les évaluations sont surtout fondées sur la santé et la sécurité des dispositifs. En morale, le critère retenu est celui de la transformation de la condition humaine biologique de la personne, qui est guidée par le principe de la « nature humaine ». Dans le *Vocabulaire technique et critique de la philosophie*, ce mot « nature¹² » est défini ainsi :

12. A. Lalande (2002), « Nature », *Vocabulaire technique et critique de la philosophie*, Paris : Presses universitaires de France/Quadrige, p. 667-671.

- a) la « nature [substance] d'un être qui se développe lui-même » ;
- b) l'« état où naissent les hommes », « par opposition à ce qui est artificiel et voulu » ;
- c) le « principe fondamental de tout jugement normatif » ;
- d) les « "lois de la nature", prises dans ce sens, sont alors les règles idéales et parfaites, dont les morales ou les législations humaines ne sont qu'une imitation imparfaite ».

Difficulté. La difficulté du choix du critère repose sur le fait que si certains enjeux sont privilégiés et non l'ensemble, la justification se trouve dans le principe : c'est-à-dire la nature du discours normatif qui guide l'évaluation. Ainsi, parmi tous les enjeux identifiés (ci-dessus), la morale tend à absolutiser le privilège de l'espèce humaine dans sa biologie (nature biologique). Elle prend racine dans l'état biologique (le naturel) par opposition à la technique (l'artificiel). Toute la difficulté d'opérationnaliser le choix de ce critère se pose dans le débat qu'il suscite autour du principe de la nature humaine.

iii) Valeur correspondante. L'évaluation morale à cette étape demande d'établir la valeur correspondante au critère. Comme cela a été signifié ci-dessus, le choix du critère de la condition biologique de vie de la personne pointe déjà vers une valeur considérée comme importante en morale pour réaliser le jugement de valeur : l'intégrité de la personne, qui signifie l'état d'un tout d'une personne qui est dans son entièreté à l'opposé de l'être désintégré.

Mais nous vivons dans un contexte où la violation de l'intégrité avec consentement est considérée comme normale, particulièrement en médecine, qui englobe différents types de chirurgie, parce qu'une société qui en serait exempte serait tout à fait impossible. Qu'en serait-il de ce cas (prévu, mais non pas interdit, sans nécessairement avoir été autorisé) où l'on remplacerait les parties endommagées de l'hippocampe chez des patients souffrant de la maladie d'Alzheimer par un implant BCI qui les rendrait plus intelligents que les humains qui l'ont fabriqué ?

Difficulté. Pour bien cerner la difficulté d'une évaluation morale, il faut la faire à partir de ses composantes. L'impact considéré est l'amélioration de la mémoire humaine et on peut se demander si cette amélioration viole l'intégrité du corps. La difficulté est donc celle de déterminer si l'amélioration en cause transforme suffisamment la personne en tant qu'être humain pour que l'on puisse considérer qu'il s'agit d'une violation de l'intégrité de la personne. Cette violation est-elle suffisante pour qu'elle soit considérée ?

Quelles seraient les raisons objectives qui pourraient expliquer la difficulté du jugement ? Le jugement de valeur est attributif¹³, tout comme l'illustre le **paradoxe du tas de sable** qui apparaît dans l'« éthique de l'amélioration humaine¹⁴ ». Selon une logique quantitative en mathématique, on est ramené à la question suivante : Combien de grains faut-il pour faire un tas ? Combien d'impacts de l'implant BCI faudrait-il pour attribuer le jugement qu'il y a un manquement à l'intégrité corporelle ? Appliquée au jugement de la personne, la difficulté est donc de déterminer combien de transformations sont nécessaires pour juger du manquement ou encore quel type de transformation est considéré comme suffisant pour déterminer le manquement. La logique quantitative décrit un raisonnement qui porte à conclusion, l'impossibilité de fixer une frontière (ligne de démarcation) entre le fait d'avoir ou de ne plus avoir un tas de sable. Comment sortir de l'impasse épistémologique ? Le paradoxe du tas de sable a ceci de précieux qu'il permet de montrer que toute notre tradition casuistique morale ne consiste pas à appliquer des critères du jugement qui sont préétablis dans ce contexte. La détermination de la violation de l'intégrité corporelle repose sur des données qualitatives et quantitatives, mais elle ne repose pas sur un jugement déductif. C'est la décision morale qui établit *a posteriori* le degré de suffisance de l'impact. D'ailleurs, Lin et Allhoff le reconnaissent eux-mêmes, par suite de leur explication du paradoxe tas de sable : « it could still be that some enhancements (and therapies) are morally problematic in certain applications or areas of life, i. e. contexts may matter¹⁵. » Mais la difficulté à déterminer s'il y a violation de l'intégrité morale de la personne ne change pas pour autant dans ce contexte.

iv) Principe. La décision d'interdire, de limiter ou de favoriser l'implémentation de l'implant BCI dépend du principe de la nature humaine sur lequel est fondé le jugement de valeur. Le premier chapitre de ce livre nous réfère donc à la « recherche de la définition substantive » de la nature humaine pour la rendre « objective » avec ses implications morales par la suite. Est-ce que la transformation de l'humain par l'implant de mémoire BCI s'accorde avec une telle définition substantive de notre nature ? Si tel est le cas, jusqu'à quel point augmente-t-elle ou diminue-t-elle notre intégrité ?

Difficulté. Il n'y a pas d'accord universel sur la vérité des définitions substantives de la nature humaine proposées en morale. Or, en morale, cette

13. Voir J.-P. Béland, G.A. Legault et J. Patenaude (2014), « Les enjeux identitaires de l'humain dans le débat philosophique sur la robotique humanoïde et l'amélioration humaine », *BioéthiqueOnline*, p. 3-14, <<http://bioethiqueonline.ca/3/>>.

14. Voir F. Allhoff, P. Lin, J. Moore et J. Weckert (2009), *Ethics of Human Enhancement : 25 Questions & Answers*, US National Science Foundation Report, p. 13.

15. *Ibid.*

conception est le fondement de l'acceptation ou de l'interdiction. C'est ici que l'on retrouve le débat philosophique contemporain sur les fondements objectifs de la morale, comme celui soulevé par le transhumanisme contre l'humanisme : le transhumanisme (tel que décrit dans la *Déclaration transhumaniste* en 2012) prêche le droit à la « liberté morphologique »¹⁶ du corps dans son intégralité selon l'usage des technologies, alors que l'humanisme commande la « fixité » intégrale du corps. Ce débat international provoque l'ambiguïté du sens de cette nature humaine comme principe qui fonde les obligations morales pour décider s'il y a violation ou pas de l'intégrité. Ce principe de la nature humaine est ambigu, car il est susceptible de recevoir au moins trois définitions substantives qui fondent des décisions divergentes :

A. Définition substantive de la nature selon l'humanisme radical pour

faire interdire la modification de la condition humaine : « Fukuyama soutient que la nature biologique humaine doit être considérée comme fixe même si elle ne l'est pas, car la conséquence d'une malléabilité extrême de l'humanité (en tant qu'individu, espèce biologique) serait la disparition des valeurs démocratiques¹⁷ » comme l'égalité (égalité biologique) et l'autonomie en raison de l'« obligation sociale de s'améliorer sous peine d'être hors jeu¹⁸ ». Les démocraties peuvent et doivent restreindre ces conséquences sur la *nature humaine*¹⁹.

B. Définition substantive de la nature selon le transhumanisme pour

favoriser la liberté de la modification de la condition humaine : Les transhumanistes comme Kurzweil répliquent que la nature de l'être humain n'est pas dans ses limitations biologiques, mais dans notre capacité à les dépasser : « Then perhaps our basic disagreement is over the nature of being human. To me, the essence of being human is not our limitations – although we do have many – it's our ability to reach beyond our limitations. We didn't stay on the ground. We didn't even stay on the

-
16. M. More et N. Vita-More (2013), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, Singapour : Wiley-Blackwell, p. 54.
 17. C. Bégorre-Bret (2004), « Bioéthique et posthumanité. F. Fukuyama : La fin de l'homme. Les conséquences de la révolution biotechnique / J. Habermas L'avenir de la nature humaine. Vers un eugénisme libéral / D. Lecourt : Humain, post-humain », *Les études philosophiques* 2, n° 69, p. 263. Cet article est disponible en ligne : <http://www.cairn.info/article.php?ID_REVUE=LEPH&ID_NUMPUBLIE=LEPH_042&ID_ARTICLE=LEPH_042_0253>.
 18. Voir N. Nevejans (2017), *Traité de droit et d'éthique de la robotique civile*, Bordeaux : LEH Édition, p. 1144.
 19. F. Fukuyama (2002), *Our Posthuman Future. Consequences of the Biotechnology Revolution*, New York : Farrar, Straus and Giroux, p. 218.

planet. And we are already not settling for the limitations of our biology²⁰.» C'est pourquoi l'interdiction des impacts potentiels du développement des nano-implants (nanorobots) pour favoriser l'augmentation des performances humaines lui apparaît comme illégitime.

C. Définition substantive de la nature selon l'humanisme modéré pour que la modification de la condition humaine se freine d'elle-même selon ses propres normes : Cette technique est plutôt *constitutive* de la nature humaine qui possède un « pouvoir d'autoconstruction ». [...] Qu'on le veuille ou non, la vie même de l'être humain dépend de la technique par laquelle il s'autoconstruit. Et si l'on inscrit l'homme dans le « flux du vivant », comme il le convient, la réalité technique ne peut être pensée sans la considérer comme une dimension essentielle des êtres humains dont le caractère propre est de se montrer en perpétuel devenir, animés d'une dynamique constructive et destructive permanente. Il en découle que la spécificité de la vie n'est pas dans la structure de l'organisme envisagé de façon statique [...]. Alors ce qu'il y a de plus naturel, c'est que « la technique n'est pas extérieur [*sic*] à la vie humaine. » Issue de la vie, elle y trouve sa place, y insère et y compose ses normes²¹. »

Le problème de l'évaluation morale provient du fait que l'on peut y donner trois définitions qui se prétendent substantives et donc réelles. Mais il n'existe aucune méthode reconnue pour établir qu'une seule d'entre elles soit vraie.

2. SYSTÈMES D'AIDE À LA DÉCISION (INTELLIGENCE ALGORITHMIQUE)

Les systèmes d'aide à la décision, grâce à l'intelligence algorithmique (algorithme de l'intelligence artificielle²²), constituent un autre cas de développement issu de la convergence des NBIC pour l'augmentation humaine. Cette intelligence algorithmique (en réseaux de neurones profonds) est mise à l'essai dans des domaines critiques comme la santé (pour assister les médecins dans leurs décisions cliniques), la justice (pour guider les avocats dans l'analyse de la jurisprudence, le risque de récidive à la suite de l'attribution d'une peine en droit criminel, la prédiction de l'issue des procès), les banques (pour donner

20. R. Kurzweil (2005), *The Singularity Is Near. When Humans Transcend Biology*, New York: Viking Penguin, p. 311.

21. D. Lecourt (2003), *Humain post humain*, Paris: Presses universitaires de France, p. 85-86.

22. Voir Commission nationale de l'informatique et des libertés (Décembre 2017), *Comment permettre à l'homme de garder la main?*, p. 16-18, <<https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de>>.

les conseils de placements financiers, décider les échanges boursiers), les transports (pour rendre les voitures autonomes), et l'armée (pour assister les soldats, les pilotes d'avion, les stratèges militaires)²³. Le fait qu'il y a toujours un risque à se fier aux algorithmes et à l'IA (algorithme sophistiqué d'apprentissage) de ces systèmes d'aide à la décision (puisque'il s'agit là de déléguer des tâches et des décisions à des algorithmes recourant à des techniques d'apprentissage machine de plus en plus autonomes), cela soulève le débat sur la question de l'évaluation des conséquences (avantages et risques) de ces systèmes d'aide à la décision dans la vie des personnes (individus, institutions et société) et sur la question de leur autorisation, leur limitation ou leur interdiction. L'évaluation de ces systèmes d'aide à la décision demeure un problème difficile. Notre processus simplifié d'évaluation nous permettra ici de préciser les éléments de l'évaluation – i) l'impact, ii) le critère de l'évaluation, iii) la valeur correspondante et iv) le principe qui fonde la décision d'interdire, de freiner ou de favoriser – de même que les difficultés à opérationnaliser cette évaluation en fonction de l'interdiction du tout ou d'une partie de ce développement (évaluation morale).

i) Impact. La première étape de l'évaluation vise à déterminer l'impact réel et potentiel du développement de tels systèmes d'aide à la décision en indiquant d'abord de quel type d'intelligence algorithmique il s'agit et comment celle-ci fonctionne aujourd'hui.

Impact réel. L'impact réel est la prise de décision fiable, valide et rapide, lorsqu'il est possible de tester des systèmes d'aide à la décision efficaces :

Apprendre à un système à reconnaître, sur une photographie aérienne d'une zone maritime, les différents navires présents peut servir à confronter l'emplacement des embarcations ainsi repérées aux signaux émis par les balises et à identifier des navires en perdition ou qui cherchent à se soustraire à la surveillance maritime. L'intérêt réside dans la rapidité d'une opération qui, si elle n'est pas automatisée, réclame un temps et des moyens considérables. Depuis quelques années, les progrès de ces techniques sont tels que la machine surpasse désormais l'humain pour la fiabilité de l'identification de navires parfois difficilement distinguables de nuages²⁴.

L'impact réel de ces systèmes d'aide à la décision est en évolution constante dans le domaine de l'intelligence artificielle (algorithme de l'intelligence

23. Voir V. Nouyrgat (2017), « Algorithmes : une nouvelle intelligence est née », *Science et vie*, n° 1198, p. 47-62.

24. Voir Commission nationale de l'informatique et des libertés (Décembre 2017), *Comment permettre à l'homme de garder la main?*, op. cit., p. 17.

artificielle), et il est tout aussi difficile à comprendre que la technologie qui en est le socle :

Le deep learning, technique parmi d'autres en intelligence artificielle, est à la fois celle qui présente le plus d'applications spectaculaires et qui présente un inconvénient majeur. On ne sait pas en expliquer les résultats. Ce sont des réseaux de neurones qui fonctionnent comme des boîtes noires²⁵.

« Allez-vous prendre le risque d'utiliser des outils qui sont des **black box** sans aucune compréhension du fonctionnement et de la pertinence des résultats²⁶ ? » C'est le nœud du problème que la relation à l'impact réel de la machine d'aide à la décision met en cause :

Le cas de la médecine est particulièrement critique non seulement en raison de l'impact des décisions et recommandations sur les personnes mais aussi en raison du fait que la discussion implique ici des systèmes fondés sur la technologie du machine learning. Ceci implique que les logiques sous-jacentes des systèmes d'intelligence artificielle sont potentiellement incompréhensibles pour celui à qui ils sont proposés, autant d'ailleurs que pour les concepteurs du système²⁷.

C'est ainsi dire que les concepteurs ont de la difficulté à comprendre comment leur propre machine d'apprentissage automatique (en anglais *machine learning* littéralement) fonctionne, étant donné que celle-ci se développe par « apprentissage » (phase 1) afin de pouvoir donner des « réponses intelligentes » et des résultats produits (en phase 2), qui défient l'entendement :

Le débat public organisé par la CNIL a d'ailleurs été l'occasion de constater une controverse sur ce point, à propos notamment du logiciel Watson d'IBM. Le discours d'IBM souligne que Watson fonctionne sur le mode de l'« apprentissage supervisé ». Autrement dit, le système est accompagné pas à pas dans son apprentissage, ce qui permettrait d'en contrôler la logique, par opposition à un apprentissage non supervisé qui reviendrait effectivement à laisser une pleine et entière autonomie à la machine pour déterminer ses critères de fonctionnement. IBM indique également contrôler le fonctionnement des systèmes avant de décider de conserver

25. C. Zeitoun (2018), CNRS le Journal, « Peut-on faire confiance à l'intelligence artificielle? », *Dialoguea*, Tiré de *Carnet de Science*, n° 3, <<https://dialoguea.fr/ia/debat/5acc86817999044dc378f3e7>>.

26. Voir P. Dambly (2018), Éthique et systèmes algorithmiques & A1, Bruxelles, p. 5, <<http://www.legaltechcongres2018.be/wp-content/uploads/2018/10/E-P-Dambly-ETHIQUE-ET-SYSTEMES-ALGORITHMIQUES-V-FINAL-SITE.pdf>>.

27. Commission nationale de l'informatique et des libertés (Décembre 2017), *Comment permettre à l'homme de garder la main?*, op. cit., p. 28.

l'apprentissage réalisé. Au contraire, les chercheurs experts de ce domaine qui ont eu l'occasion de s'exprimer lors des différents débats organisés (et notamment la CERNA) ont régulièrement rappelé qu'en l'état actuel de la recherche les résultats fournis par les algorithmes de machine learning les plus récents n'étaient pas explicables. Cette explicabilité constitue d'ailleurs l'objet de recherches en cours. Ils insistent également sur le fait qu'il est très difficile de contrôler effectivement un système de machine learning²⁸.

Mais quels sont ces impacts réels qu'on qualifie de réponses intelligentes ? On peut commencer à utiliser la définition classique de l'algorithme pour préciser comment la machine permet de résoudre un problème de façon systématique et en arriver à ce résultat :

Classiquement, l'algorithme se définit ainsi comme une suite finie et non ambiguë d'instructions permettant d'aboutir à un résultat à partir de données fournies en entrée. Cette définition rend compte des multiples applications numériques qui, exécutant des programmes traduisant eux-mêmes en langage informatique un algorithme, remplissent des fonctions aussi diverses que fournir des résultats sur un moteur de recherche, proposer un diagnostic médical, conduire une voiture d'un point à un autre, détecter des suspects de fraude parmi les allocataires de prestations sociales, etc.²⁹.

Mais ce sont les « réseaux de neurones » dans le domaine de l'intelligence artificielle (IA) qui rendent les nouveaux algorithmes intelligents³⁰. L'humain serait-il dépassé par les résultats de cette « technique d'apprentissage machine » (« machine learning ») ultra rapide qui fait compétition à notre cerveau ? Prenons en exemple la victoire d'AphaGo (Google) contre le champion du monde du jeu de go, Lee Sedol, en 2016. Contrairement au jeu d'échecs, le jeu de go, du fait de la multiplicité innombrable des combinaisons qu'il permet, ne se prête pas à la mémorisation d'un grand nombre de parties que la machine pourrait se contenter de reproduire :

La victoire d'Alpha Go illustre le fait que les développements récents de l'IA sont notamment liés au perfectionnement de la technique du machine learning (apprentissage automatique), qui en constitue l'une des branches. Alors que le programmeur doit traditionnellement décomposer en de multiples instructions la tâche qu'il s'agit d'automatiser de façon à en

28. *Ibid.*

29. *Ibid.*, p. 5.

30. Voir V. Nouyrigat (2017), « Algorithmes : une nouvelle intelligence est née », *Science et vie*, n° 1198, p. 51.

explicitement toutes les étapes, l'apprentissage automatique consiste à alimenter la machine avec des exemples de la tâche que l'on se propose de lui faire accomplir. L'homme entraîne ainsi le système en lui fournissant des données à partir desquelles celui-ci va apprendre et déterminer lui-même les opérations à effectuer pour accomplir la tâche en question. Cette technique permet de réaliser des tâches hautement plus complexes qu'un algorithme classique. Andrew Ng, de l'Université Stanford, définit ainsi le machine learning comme « la science permettant de faire agir les ordinateurs sans qu'ils aient à être explicitement programmés ». Cela recouvre la conception, l'analyse, le développement et la mise en œuvre de méthodes permettant à une machine d'évoluer par un processus systématique, et de remplir des tâches difficiles. L'intelligence artificielle qui repose sur le machine learning concerne donc des algorithmes dont la particularité est d'être conçus de sorte que leur comportement évolue dans le temps, en fonction des données qui leur sont fournies³¹.

Les multiples applications de ces nouveaux algorithmes d'apprentissage démontrent qu'ils peuvent accomplir des tâches dont sont incapables les algorithmes classiques (la reconnaissance d'un visage sur de très vastes corpus d'images, par exemple³²). Les réponses intelligentes (impacts réels) surprennent les chercheurs au point qu'ils parlent d'une « forme d'intelligence radicalement différente de la nôtre » :

C'est une autre manière de voir le monde. Une autre manière de percevoir les couleurs d'un coucher de soleil, la forme d'une roue, les traits d'un visage, les premières notes de la musique d'une chanson populaire ; une autre manière de comprendre ; une autre manière d'imaginer de nouveaux objets ou de trouver des solutions inédites à un problème³³.

Un algorithme de détection des typologies de fraudes peut utiliser cette technique : on fournit à l'algorithme toutes les données relatives à des fraudes avérées, et l'algorithme serait capable de dégager des similitudes entre ces fraudes, et de dégager des typologies de fraudes. L'apprentissage non supervisé peut aussi servir à identifier, sur la bande sonore d'une émission de radio, les séquences de parole de différents locuteurs³⁴.

Autre exploit intellectuel : l'apprentissage profond a révolutionné le domaine de la traduction automatique :

31. Commission nationale de l'informatique et des libertés (décembre 2017), *Comment permettre à l'homme de garder la main?*, *op. cit.*, p. 16.

32. Voir V. Nouyrigat (2017), « Algorithmes : une nouvelle intelligence est née », *op. cit.*, p. 52-55.

33. *Ibid.*, p. 52.

34. Voir Commission nationale de l'informatique et des libertés (décembre 2017), *Comment permettre à l'homme de garder la main?*, *op. cit.*, p. 16-17.

En novembre [2016], une équipe de Google a publié en effet une étude sur un modèle de traduction automatique capable de traduire correctement une paire de langue pour laquelle il n'avait reçu aucun apprentissage... Entraîné à passer du portugais à l'anglais et de l'anglais à l'espagnol, il s'est avéré capable de traduire correctement du portugais en espagnol sans passer par l'anglais. Comment est-ce possible³⁵ ?

Les exemples d'utilisation de ce système algorithmique ultra rapide sont déjà très nombreux. « Les algorithmes de l'IA et les *datas* (*big data*) constituent le moteur et le carburant de ces systèmes. Ces systèmes, dans la mesure où ils reposent sur des techniques d'apprentissage, nécessitent d'énormes quantités de données³⁶. » Quel est le secret de cette technique des « réseaux de neurones profonds » qui, en quelques années, a supplanté la plupart des autres méthodes de développement de l'intelligence artificielle ? Une architecture complexe, inspirée du cortex des mammifères, qui lui permet d'apprendre aisément et de traiter toutes les subtilités du monde environnant, mais selon des critères qui lui sont propres :

Un réseau de neurones profond est un vaste programme informatique constitué de centaines ou de milliers de fonctions mathématiques (les neurones) organisées en plusieurs couches logicielles successives (le réseau le plus imposant à ce jour en compte 152). Les neurones d'une couche sont reliés à ceux des couches suivantes et précédentes. Chacun d'eux traite, selon des transformations hautement non linéaires (sigmoïde, tangentes hyperboliques), les signaux qu'il reçoit de la couche précédente et propage le résultat en sortie vers les neurones de la couche suivante. Le niveau d'information devient de plus en plus sophistiqué à mesure que l'on s'enfonce dans le réseau³⁷.

Les chercheurs ne cessent d'être surpris par cette nouvelle intelligence qui ressemble un peu à la nôtre, mais qui a des facultés insaisissables. Elle invente son propre langage. Mais il s'agit d'une « intelligence qui n'est pas consciente d'elle-même³⁸ ». Nous sommes ainsi dans la « boîte noire » (« Black box ») : données, apprentissage et un résultat produit, qui n'est pas explicable. Il y a forcément des erreurs à corriger. Il y a des dérapages possibles : par exemple, le 7 mai 2016, en Floride, une voiture Tesla en pilotage semi-automatique a percuté un camion. Cela prouve que l'impact réel (résultat produit) demeure toujours incertain. Rien n'empêche, toutefois, pour ces réseaux de neurones

35. Voir V. Nouyrigat (2017), « Algorithmes : une nouvelle intelligence est née », *op. cit.*, p. 56.

36. Commission nationale de l'informatique et des libertés (décembre 2017), *Comment permettre à l'homme de garder la main ?*, *op. cit.*, p. 5.

37. V. Nouyrigat (2017), « Algorithmes : une nouvelle intelligence est née », *op. cit.*, p. 58.

38. *Ibid.*, p. 56.

de faire irruption dans nos vies quotidiennes : « Les réseaux de neurones profonds débarquent de plus en plus dans notre vie quotidienne. Mais confrontés à des situations critiques, sauront-ils prendre les bonnes décisions ? Rien n'est moins sûr...³⁹ »

Si les décisions de ces réseaux de neurones ne sont pas toujours explicables, il y a un besoin d'explicabilité qui constitue d'ailleurs l'objet de recherches en cours⁴⁰. Et s'il est très difficile pour les chercheurs de contrôler effectivement un système recourant aux techniques d'apprentissage profond (*deep learning*), il s'ajoute alors une exigence d'intelligibilité pour ne pas en perdre le contrôle.

Impact potentiel. « Aider la prise de décision⁴¹ » dans tous les secteurs (éducation, justice, santé, sécurité, travail, culture, autre) est sans doute le type d'impact positif potentiel qui est le plus souhaitable pour justifier l'acte de déléguer une tâche décisionnelle à ces algorithmes d'intelligence artificielle qui pourraient être définis comme des « assistants personnels intelligents⁴² ». Ces assistants peuvent être purement logiciels, agissant dans le monde comme des agents vocaux, moteurs de recherche ou systèmes de reconnaissance, mais ils peuvent aussi être intégrés dans le cerveau de robots évolués. Le rapport de la CNIL (2017) mentionne des exemples de grandes applications de ces algorithmes d'apprentissage qui pourraient remplir des fonctions critiques aussi diverses que : générer de la connaissance en fournissant des résultats sur un moteur de recherche, suggérer au juge la solution jurisprudentielle la plus adéquate pour un cas donné, élaborer automatiquement un diagnostic médical personnalisé, détecter des suspects de fraude parmi les allocataires de prestations sociales, etc.⁴³.

Un avocat robot ? Un groupe de recherche à l'Université de Toronto a développé un programme informatique capable d'effectuer des recherches juridiques pour préparer les procès. Cet algorithme scanne en détail tous les procès terminés afin de trouver la meilleure défense. Il remplace le travail que fait typiquement le dernier avocat embauché dans un cabinet⁴⁴.

39. *Ibid.*, p. 58.

40. Depuis 2006, par exemple, on parle du besoin de la recherche sur l'XAI (Explainable Artificial Intelligence) dans une agence du Département de la défense (DARPA) américaine, <<https://www.darpa.mil/attachments/XAIProgramUpdate.pdf>>.

41. Voir Commission nationale de l'informatique et des libertés (décembre 2017), *Comment permettre à l'homme de garder la main?*, *op. cit.*, p. 20.

42. Voir P. Dambly (2018), *Éthique et systèmes algorithmiques & A1*, Bruxelles, p. 13, <<http://www.legaltechcongres2018.be/wp-content/uploads/2018/10/E-P-Dambly-ETHIQUE-ET-SYSTEMES-ALGORITHMIQUES-V-FINAL-SITE.pdf>>.

43. Commission nationale de l'informatique et des libertés (décembre 2017), *Comment permettre à l'homme de garder la main?*, *op. cit.*, p. 21-22.

44. A. Riopel (2018), « La révolution des machines » *Curium Le Magazine Science, Techno et Société*, <<https://curiummag.com/la-revolution-des-machines/?fbclid=IwAR1Ngxe7GvKQrordZAPdCDYeI7leI2qAwpknGSNtMH-YwuPh7ttyBoPQjNg>>.

Et le médecin ? Des compagnies comme Google, IBM et d'innombrables *start-up* multiplient les partenariats avec des hôpitaux. Peut-il être rationnel de ne pas suivre ces algorithmes qui promettent d'exceller dans la prescription de médicaments, des examens dermatologiques et ophtalmologiques et dans la détection de cellules cancéreuses ? « Déjà les algorithmes identifient des cellules cancéreuses dans le foie plus facilement et plus précisément que les médecins⁴⁵ ! »

Un autre exemple est celui de systèmes d'expertise médicale et juridique, où il y a également délégation de la fonction consistant à trouver des outils normatifs.

Le développement de ces systèmes pourrait évoluer vers des machines morales. Wallach et Allen ont proposé le projet de créer des « machines morales » (« agents moraux artificiels » ou « AMAs »⁴⁶) pour aider à prendre des décisions morales dans tous les secteurs : robots d'assistance, robots militaires, robots compagnons, robots éducatifs, robots de service dans le monde médical, etc. Ces robots de service devenant de plus en plus autonomes du point de vue moral (autonomie morale) pourraient nous aider à améliorer nos décisions morales en fixant les objectifs moraux et les valeurs morales à atteindre. Mais ils pourraient aussi faire des choix éthiques (solutionner des dilemmes) mieux que les humains dans le but de leur rendre service en accomplissant ces tâches, ce qui soulève le « dilemme éthique du tramway⁴⁷ ».

« Mais que se passerait-il si on délèguait à des IA et autres algorithmes des décisions importantes⁴⁸ ? »

Difficulté. La difficulté de la mesure d'impact pour ces systèmes d'aide à la décision est sensiblement la même d'un cas à l'autre. L'enjeu particulier dans le cas de l'IA est la complexité de la technique du « *deep learning* » qui, échappant à nos contrôles, ne permettrait pas d'avoir la preuve de la validité des résultats.

ii) Critère. Comme nous l'avons vu dans le premier exemple, le critère utilisé pour faire le jugement de valeur est directement associé aux enjeux soulevés par les impacts dans la vie de la personne, des institutions et de la société.

45. *Ibid.*

46. Voir W. Wallach et C. Allen (2009), *Moral Machines. Teaching Robots Right from Wrong*, New York : Oxford University Press.

47. « Dilemme du tramway », Wikipédia, <https://fr.wikipedia.org/wiki/Dilemme_du_tramway>.

48. C. Zeitoun (2018), CNRS le Journal, « Peut-on faire confiance à l'intelligence artificielle ? », *Dialoguea*, Tiré de *Carnet de Science*, n° 3, <<https://dialoguea.fr/ia/debat/5acc86817999044dc378f3e7>>.

Enjeu de la sécurité des personnes : la question est de savoir comment faire de la programmation (c'est-à-dire de l'intelligence algorithmique) en s'assurant de rendre la machine sécuritaire, alors qu'il a été démontré que cette machine n'est pas sans risque réel d'erreur (risque de biais décisionnel selon le programmeur de la machine ; risque d'erreur selon les circonstances particulières), et que la moindre erreur pourrait impliquer un risque grave, comme le signifie Lin dans *Robot Ethics* : « – even a tiny software flaw in machinery, such as a car or a robot, could lead to fatal results⁴⁹ ».

Enjeu de la responsabilité professionnelle : cependant, même si ce développement de la machine morale, en tant qu'agent moral capable de prendre les meilleures décisions pour solutionner un dilemme, devient sécuritaire, cela veut-il dire que l'humain ne risque pas de se décharger de sa responsabilité professionnelle sur des machines ?

Le prestige et la confiance accordés à des machines jugées souvent infaillibles et « neutres » ne risquent-ils pas de générer la tentation de se décharger sur les machines de la fatigue d'exercer des responsabilités, de juger, de prendre des décisions ? Comment appréhender les formes de dilution de la responsabilité que sont susceptibles de susciter les systèmes algorithmiques, complexes et très segmentés⁵⁰ ?

Enjeu du droit de la responsabilité : la délégation de la capacité de décisions critiques risque de conduire à une « forme de dilution de figures d'autorité traditionnelles, de décideurs, de responsables, voire de l'autorité même de la règle de droit » qui est considérée comme une déresponsabilisation malgré qu'aucune loi n'interdit le recours à ces processus décisionnels en France :

Les décisions les plus cruciales (diagnostics médicaux, décisions judiciaires, décision d'ouvrir le feu dans un contexte de conflit armé, etc.) qui pourraient être, voire commencent à être (à l'étranger notamment) déléguées à des systèmes automatisés sont – au moins dans certains cas – déjà clairement thématiques par la tradition juridique, en France. Seul un médecin est ainsi habilité à établir un diagnostic qui, autrement, relèverait de l'exercice illégal de la médecine. Il en va de même de la décision du juge, qui ne saurait en toute rigueur être déléguée à un système automatisé. Dans cette perspective, ce type de système est présenté dans ces domaines comme une « aide » à la prise de décision⁵¹.

49. P. Lin (2012), « Introduction to Robot Ethics », dans P. Lin, K. Abney et G.A. Bekey (dir.), *Robot Ethics. The Ethical and Social Implications of Robotics*, Cambridge : The MIT Press, p. 7.

50. *Ibid.*, p. 5.

51. *Ibid.*, p. 27.

Enjeu de la représentation de la personne (dignité) : il y a toujours le risque, soit par ignorance ou par facilité, que nous perdions notre dignité humaine en déléguant les décisions et notre autonomie morale à la machine⁵².

Le perfectionnement et l'autonomie croissante des artefacts techniques permettent des formes de délégations de tâches, de raisonnements et de décisions de plus en plus complexes et critiques à des machines. Dans ces conditions, à côté de l'augmentation de sa puissance d'agir permise par la technique, n'est-ce pas aussi son autonomie, son libre arbitre, qui peut se trouver érodé ? Comment donc s'assurer que la prédiction et la recommandation fournies par les machines ne soient effectivement qu'une aide à la prise de décision et à l'action humaine sans aboutir à une déresponsabilisation de l'homme, à une perte d'autonomie⁵³ ?

Enjeu économique : Cette délégation de nos décisions et de notre autonomie morale à la machine qui doit nous rendre service ne risque-t-elle pas en même temps d'être encouragée par le « discours marchand » qui vante les avantages de cette machine pour en assurer le développement économique et la vente ? Ce discours marchand tend en même temps à faire miroiter une réglementation algorithmique de la machine qui évolue au point qu'elle devient plus morale que l'humain faillible :

Cette évolution est parfois explicitement souhaitée. [...] Sans doute faut-il remarquer ici que la tentation [...] de se passer d'une normativité humaine et de préférer une normativité algorithmique est favorisée par les discours marchands. Ces derniers vantent l'« objectivité » supposée des systèmes automatiques (par opposition à un jugement humain toujours faillible). Ils influent donc sur la tendance des utilisateurs à prendre le résultat produit par une machine pour une vérité incontestable, alors même qu'il est de part en part déterminé par des choix (de critères, de types de données fournies au système) humains⁵⁴.

Enjeu de gouvernance : L'humain ne risque-t-il pas ainsi d'être gouverné par la machine ?

Certains, comme Tim O'Reilly, imaginent d'ores et déjà l'avènement d'une « réglementation algorithmique » qui verrait la « gouvernance » de la cité confiée aux algorithmes : grâce aux capteurs connectés, lieux, infrastructures et citoyens communiqueraient en permanence des données traitées en vue de rationaliser et d'optimiser la vie collective selon des lois considérées comme

52. Voir C. Zeitoun (2018), CNRS le Journal, « Peut-on faire confiance à l'intelligence artificielle ? », *op. cit.* ; J.-G. Ganascia (2017), *Le mythe de la singularité. Faut-il craindre l'intelligence artificielle ?*, Paris : Seuil.

53. Commission nationale de l'informatique et des libertés (Décembre 2017), *Comment permettre à l'homme de garder la main ?*, *op. cit.*, p. 5.

54. *Ibid.*, p. 28.

« naturelles », émanant des choses mêmes, une « normativité immanente », comme l'expliquent Thomas Berns et Antoinette Rouvroy⁵⁵.

Difficulté. La difficulté ici est de pouvoir justifier le choix du critère en privilégiant certains enjeux et non l'ensemble. C'est la même difficulté que celle mentionnée dans le cas précédent concernant l'implant de mémoire. La justification se trouve dans le principe qui tend à absolutiser le respect de la représentation de la personne. L'évaluation morale privilégie donc l'impact sur la représentation de la personne (dignité).

iii) Valeur correspondante. La valeur qui correspond à la représentation de la personne libre est la dignité : la valeur propre de la dignité de la personne humaine en tant qu'être raisonnable et libre constitue l'identité de la nature humaine qu'il faut protéger contre l'instrumentalisation afin de ne pas traiter la personne humaine simplement comme un moyen au service d'une technologie plutôt qu'une fin.

Il faut se rappeler que la liberté de l'être raisonnable en philosophie morale n'est pas un fruit de culture. Cette condition *sine qua non* de possibilité d'une action (d'une décision, par exemple) est un fait de la nature. Elle est aussi naturelle que notre condition biologique. Le sens primitif qualifie la liberté d'état de celui qui fait ce qu'il veut et non ce que veut un autre que lui : elle représente l'absence de contrainte étrangère. Kant est le premier grand défenseur de cette liberté en philosophie morale. Selon Kant, cette liberté est la clé de voûte de l'autonomie morale : elle est d'une nature spontanée par opposition à une nature cultivée (culture artificielle). Elle commence à agir par elle-même. Elle est la faculté d'agir librement. Et puis, c'est elle qui cause l'effet (elle n'est pas causée par une autre chose qu'elle-même). En troisième lieu, elle s'exerce selon une exigence naturelle, « c'est-à-dire une loi de sa causalité sans laquelle elle ne serait nullement cause ». Finalement, une dernière caractéristique en philosophie morale : « la liberté est, comme la chose en soi, inconnue et inconnaissable de manière directe et immédiate⁵⁶ ».

Ce qu'il faut alors retenir de la morale de Kant, lorsqu'il s'agit d'intervenir sur le plan de l'évaluation morale des systèmes d'aide à la décision, c'est que l'humain a un devoir de respecter sa liberté et son autonomie, sa faculté d'agir librement et son droit de se gouverner par ses propres lois à l'intérieur d'un État. L'évaluation morale est, par définition, le rapport que l'on entretient avec cette valeur de la liberté (que cherche à protéger notre tradition morale humaniste contre la machine inconsciente d'elle-même). L'humain en tant qu'agent

55. *Ibid.*

56. Voir G.V. Riet (1980), « Liberté et espérance chez Kant », *Revue philosophique de Louvain* 38, p. 189, <https://www.persee.fr/doc/phlou_0035-3841_1980_num_78_38_6081>.

moral applique cette valeur de la liberté afin de mesurer, juger, et de s'assurer que toute délégation de sa capacité décisionnelle à des systèmes automatisés à son service ne l'empêche pas de prendre les décisions les plus cruciales (diagnostics médicaux, décisions judiciaires, décision d'ouvrir le feu dans un contexte de conflit armé, etc.) et ne viole pas la valeur de la liberté de la personne humaine (identité – nature – dignité de la représentation de la personne humaine) en tant qu'agent moral.

Difficulté. Il est toutefois difficile, dans l'évaluation morale, de juger jusqu'à quel point l'acte de délégation peut violer la liberté (identité-dignité humaine). Jusqu'où les faits peuvent-ils déterminer que ce viol par la délégation de notre capacité décisionnelle à l'IA d'une machine morale est majeur ou mineur ?

Cette difficulté à faire un jugement de valeur et d'attribuer à l'analyse d'impacts la confirmation que ces faits sont tels qu'il s'agit là d'un manquement à la dignité humaine est la même que celle soulevée dans l'analyse sur le cas de l'intégrité (exemple 1).

La question qui se pose lorsqu'il s'agit d'intervenir à partir de l'évaluation morale en est toujours une de degré : jusqu'à quel point la délégation d'une tâche ou d'une décision critique accordée à un robot décisionnel en développement peut-elle violer la valeur de l'identité-dignité humaine ? On revient ainsi sur un problème : jusqu'où faut-il approfondir les difficultés de chaque cas de délégation de la liberté et de l'autonomie à une machine, ou bien simplement signifier qu'elles sont de même nature ?

Le *premier type* (cas) de « délégation » de la liberté et de l'autonomie à des robots implique que l'on est enclin à confier aux IA des machines plus rapides que nous pour traiter des données. Il y a un risque d'externalisation de la capacité décisionnelle de l'humain dans des conceptualisations techniques (algorithmes) qui implique une perte d'apprentissage de l'autonomie morale.

Le deuxième type de délégation semble être plus de l'ordre de la pression sociale exercée sur la décision professionnelle et le paradoxe décisionnel auquel pourrait être confronté un médecin : s'il décide, par exemple, d'aller contre l'avis des systèmes médicaux APACHE⁵⁷ (qui sont des systèmes de classification de la gravité de la maladie) au nom de la décision professionnelle, et que la décision d'APACHE s'avérerait la meilleure en fin de compte, cette décision professionnelle pourrait être jugée moralement irresponsable.

57. Voir Open clinical (2002-2011), *AI Systems in Clinical Practice*, <http://www.openclinical.org/aisp_apache.html>.

L'enjeu moral est celui de la valeur qu'on accorde à une machine plus rapide que nous pour traiter plusieurs données et en arriver à une conclusion comme un diagnostic dans le domaine médical. Face à ce diagnostic, le médecin pourrait ainsi se servir des données de celui-ci pour confirmer ce qu'il pensait déjà afin de pouvoir établir son diagnostic final.

Si ce même robot en développement devient capable de traiter les données personnelles du patient (son historique médical, résultats de tests, etc.) et de proposer une voie thérapeutique, nous sommes alors confrontés à un autre niveau de délégation (comme dans le cas du paradoxe décisionnel exprimé ci-dessus) : si le médecin décide de ne pas suivre la voie thérapeutique suggérée par la machine au nom de l'autonomie professionnelle, ce médecin pourrait être moralement blâmé pour ne pas avoir choisi la bonne thérapie.

Dans cette délégation progressive (comme nous l'avons vu ci-dessus), la question demeure : comment justifier les peurs ou les raisons objectives qui influencent notre jugement sur l'enjeu moral ? Cette question est tout particulièrement pertinente, quand il est question du risque, soit par ignorance ou par facilité, que nous perdions notre dignité humaine en déléguant les décisions et notre autonomie morale à la machine. Il y a une double peur de délégation de la liberté morale qui s'opère dans la vie réelle, au sein du monde actuel, par les algorithmes des machines :

- La peur de la relativisation progressive de la capacité décisionnelle humaine si la machine fonctionne (trop) bien.
- La peur de la diminution progressive de l'autonomie morale de l'humain et de l'accroissement constant de l'autonomie morale des ordinateurs.

Même si elle peut s'avérer infime, la délégation pourrait-elle être suffisante pour considérer que nous perdons notre dignité-identité ?

iv) Principe. Il s'agit encore ici de l'application du discours moral, interdisant ou non au nom de la valeur fondée sur la représentation de la personne humaine en tant qu'un être raisonnable et libre (identité – nature – dignité humaine) qui sert de principe fondamental à l'ensemble de l'évaluation morale dans son sens humaniste. L'exemple de l'intégrité (cas 1) nous démontre qu'il n'y a pas d'accord universel sur la valeur de la définition substantive de la nature humaine nous permettant de décider si la modification technologique par l'implant BCI se doit d'être interdite, parce que violant l'intégrité de la personne humaine. Il en est de même pour le cas présent de la délégation de notre capacité décisionnelle et de notre autonomie morale à une machine d'aide à la décision : ce deuxième cas s'accorde-t-il plus avec une définition universelle de l'identité – nature – dignité humaine en tant qu'être rationnel et libre ?

Difficulté. Il n'y a pas d'accord universel sur la définition de cette liberté substantive caractérisant l'identité-dignité de l'être humain. Dès que nous déléguons des décisions et notre autonomie morale à des machines, sommes-nous conscients que nous attribuons ainsi la valeur substantielle de la dignité humaine à des machines ? Le discours transhumaniste soutient une telle attribution en déléguant la dignité humaine à des machines morales qui pourraient devenir plus humaines que les humains. Cela soulève deux problèmes : l'identification des critères essentiels à la dignité de la personne humaine et la preuve que cette conception est vraie. C'est ici que l'on retrouve le débat philosophique sur les fondements objectifs de la dignité humaine.

Mais qu'est-ce qui fonde la nécessité d'un accord universel sur le principe de la dignité de la personne humaine en tant que valeur suprême pour fixer des interdictions à la délégation de notre dignité à une machine morale ? Le débat entre le transhumanisme et l'humanisme remet en question la légitimité de ce principe en provoquant l'équivocité de la dignité permettant d'interdire ou non :

- A. Dignité selon le transhumanisme** – la dignité au sens de Pic de la Mirandole qui préconise la liberté sans contrainte de changer la fin (la machine morale devenant une fin) *versus*
- B. Dignité selon l'humanisme** – la dignité au sens kantien comme contrainte à la liberté pour préserver l'humain (l'humain comme fin plutôt que l'instrumentalisation de l'humain pouvant être traité comme un moyen d'une technologie)⁵⁸.

Nous sommes ici en présence de deux définitions substantielles de l'être humain et aucune d'elles ne peut être prouvée vraie. À défaut de vérité, les deux jugements moraux sont équivalents.

3. LES IMPLANTS RFID

Des puces émettrices, nommées Radio Frequency Identification Device (RFID) (comme la puce VerichipTM en forme de grain de riz ou son successeur qui sont liés au développement des technologies convergentes des NBIC⁵⁹), pourraient éventuellement être implantées dans notre corps pour différents usages :

58. Voir J.-P. Béland, G.A. Legault, P. Boissy, M. Parent et J. Patenaude (2011), « The Social and Ethical Acceptability of NBICs for Purposes of Human Enhancement. Why Does the Debate Remain Mired in Impasse? », *op. cit.*, p. 297-299.

59. Voir Centre de recherche Travail et technologies (2006), « Les enjeux sociétaux des technologies convergentes », *La lettre EMERIT*, n° 48, Belgique : Namur, p. 1-8.

- l'identifiant personnel ;
- l'interaction avec notre environnement intelligent ou des réseaux ambiants diversifiés (Wi-Fi, Bluetooth, radiofréquences) ;
- le traçage ;
- le capteur d'informations.

Les expériences médiatisées de Kevin Warwick (« Cyborg 1 » en 1998 ; « Cyborg 2 » en 2002 et « projet Cyborg » par la suite⁶⁰) illustrent la faisabilité de telles implantations dans le corps humain.

La remise en cause des implications de ces implants sur la liberté des personnes a suscité un questionnement éthique dans *The Guardian* (3 septembre 2002) et par la suite « critique » de la *Community Care* (5 septembre 2002). Lorsque Warwick et ses collègues proclamaient que le « projet Cyborg » pourrait être utilisé non seulement à des fins de diagnostic pour traiter des patients, mais aussi à des fins de contrôle ou de surveillance à distance des enfants si ceux-ci étaient enlevés, des parents ont voulu contraindre leur fille de 11 ans à se faire implanter une telle puce comme mesure de sécurité pour la retracer au cas où elle serait enlevée.

Le bien-être des enfants semble être un enjeu primordial, mais il est aussi question d'implanter la puce chez des personnes atteintes de la maladie d'Alzheimer et chez des prisonniers en libération conditionnelle⁶¹.

Ces puces deviendront-elles un jour indispensables ? Pensons-nous que les gens accepteront de se faire implanter une puce en ayant consenti à cette option en toute liberté de choix ? Allons-nous « vers une société de contrôle⁶² » ? « Vers un monde sous surveillance⁶³ ? » L'accepterons-nous ? Nathalie Nevejans rappelle que les droits fondamentaux de la personne s'opposent à une telle implantation sans qu'un consentement soit préalablement exprimé⁶⁴. Avant que les limites techniques qui freinent le développement de la puce ne soient

60. S. Lavelle (03/2010), « Implants, puces et transhumains », <<http://base.d-p-h.info/fr/fiches/dph/fiche-dph-8223.html>>.

61. A.L. Madinier (2011), « La surveillance de demain : puces RFID et implants sous-cutanés », dans E. Labrot et P. Ségur (dir.), *Un monde sous surveillance?*, Presses universitaires de Perpignan, voir le paragraphe numéroté 25 en marge, <<https://books.openedition.org/pupvd/3969?lang=fr>>.

62. D. Cerqui et B. Müller (2010), « La fusion de la chair et du métal : entre science-fiction et expérimentation scientifique », *Sociologie et sociétés*, vol. 42, n° 2, <<https://www.erudit.org/fr/revues/socsoc/2010-v42-n2-socsoc3977/045355ar.pdf>>.

63. Voir E. Labrot et P. Ségur (dir.), *Un monde sous surveillance?*, Presses universitaires de Perpignan, <<https://books.openedition.org/pupvd/3969?lang=fr>>.

64. H. Molineri (24 avril 2017), « Faut-il avoir peur des puces RFID? », Numerama, <<https://www.numerama.com/tech/251447-faut-il-avoir-peur-des-puces-rfid.html>>.

dépassées, n'avons-nous pas besoin d'une réglementation précise servant à faire une évaluation pour interdire, limiter ou autoriser ?

Ces implants RFID serviront ici d'exemple pour illustrer l'évaluation juridique qui ressemble partiellement au langage moral parce qu'elle mobilise des obligations, mais qui se distingue de la morale par sa capacité à contraindre les personnes à exécuter leurs obligations selon une législation précise comme les droits fondamentaux, et plus particulièrement le principe du respect de la vie privée et de l'intégrité des données à caractère personnel qui fait partie de notre dignité. Notre processus simplifié d'évaluation permettrait non seulement de préciser les éléments de l'évaluation : l'impact réel et potentiel de l'implant, le critère sur lequel porte le jugement, la valeur correspondante qui permet de faire le jugement et le principe fondamental qui justifie le jugement, mais également de mettre en lumière les difficultés auxquelles ces éléments nous confrontent.

i) Impact. La première étape de l'évaluation vise à déterminer l'impact réel et potentiel du développement technologique des implants RFID en montrant tout d'abord plus précisément de quel type de puce il s'agit et comment elle fonctionne. Le traçage des objets, des animaux et des personnes par les puces RFID n'est pas une nouvelle idée. Différentes formes de puces sont déjà disponibles, allant des *puces passives* ayant une mémoire qu'il suffit de scanner pour avoir de l'information, jusqu'aux *puces actives* et *préventives*, pour lesquelles les informations qu'elles contiennent peuvent être captées à distance (par satellite), « créant ainsi un formidable outil de localisation et d'identification des biens et des personnes, dans l'espace et dans le temps⁶⁵. »

Impact réel. Selon des experts⁶⁶, la puce RFID a connu un fort développement depuis les années 2000. Cette puce peut contenir des informations sur un produit, un véhicule ou le corps de la personne sur lesquels l'étiquette est collée ou implantée. Elle sert donc à stocker des données et à les transmettre au lecteur RFID par des ondes radio. Deux types de puces se distinguent selon l'utilité à laquelle elles sont destinées :

- Les puces à usage unique (*lecture seule*) : la puce contient des données qui sont lues par le lecteur RFID sans possibilité de les modifier.

65. Voir J. Miller (22 novembre 2017), « La RFID et la différence entre les étiquettes passives, semi-passives et actives », <<https://www.computype.com/fr/blog/rfid-and-the-difference-in-passive-semi-passive-and-active-tags>>; Voir M. Arnaud (2009), « La RFID : un risque de traçage généralisé », Hermès, La revue, 1, n° 53, p. 109-112.

66. Ooreka Entreprise, *Gérer la traçabilité des produits RFID. Puce RFID*, <<https://rfid.ooreka.fr/comprendre/puce-rfid>>.

- Les puces réinscriptibles (*lecture/écriture*) : les données inscrites sur la puce peuvent être modifiées par le lecteur RFID. Ce lecteur peut se connecter à des serveurs informatiques comme celui d'un système d'information d'un hôpital, par exemple, afin de récupérer ou de modifier des informations pointues sur la personne implantée : groupe sanguin, allergies, maladies, nom, numéro de dossier médical, etc.

Trois autres types de puce RFID peuvent aussi être distingués :

- Les *puces passives* : elles fonctionnent sans batterie et sont activées au moyen d'un lecteur émetteur-récepteur qui leur transmet des ondes magnétiques (ex. : badges RFID).
- Les *puces actives* : elles possèdent leur propre batterie et transmettent de façon autonome des informations qu'elles ont enregistrées au capteur.
- Les *puces es* : elles sont munies d'un système de sécurité qui permet de crypter les informations qu'elles contiennent. Les données qui y sont accessibles nécessitent une identification de son porteur (ex. : carte bancaire).

La radio identification, plus communément appelée RFID, est le principe de fonctionnement de ces puces. La radio identification est une technique de traçabilité qui a déjà largement fait ses preuves dans de nombreux domaines tel que les télépéages, le suivi des livres dans les bibliothèques, les services bancaires, les passeports biométriques, les accès aux transports. C'est une technique de traçabilité qui permet d'identifier rapidement des personnes, particulièrement lorsque l'information est incorporée dans des cartes bancaires, des passeports... Tel est l'impact réel de la puce à usage unique quand la lecture (lecture seule) des informations de la puce vers le lecteur se fait par radiofréquence et non par lecture optique, comme c'est le cas pour les codes à barres.

Dans le cas de la puce implantable chez les personnes, l'impact réel de la radio identification (RFID) se perçoit dans différents usages, pourvu que la puce fonctionne normalement :

- ↳ L'identification rapide et automatique par lecteur optique des personnes, de manière à pouvoir établir (contrôler) si leurs identités sont conformes à ce qu'elles doivent être

L'exemple le plus frappant est certainement celui du Baja Beach Club situé à Barcelone. Cette discothèque a offert à ses clients VIP la possibilité de se faire implanter un dispositif RFID sous la peau. Cette puce contient le nom, la photo et le numéro matricule du client. Pour payer ses consommations, ce

dernier approche son bras d'un *scanner* qui affiche l'état du compte qu'il a ouvert et qui sera débité automatiquement⁶⁷.

↳ L'interaction rapide et efficace de la communication avec notre environnement intelligent ou réseaux ambiants

L'exemple le plus frappant est celui de l'entreprise New Fusion, de Malines, en Belgique :

Fin 2016, l'entreprise New Fusion de Malines annonçait qu'elle allait placer une puce électronique RFID (identification par radio fréquence) dans les mains de huit travailleurs volontaires [afin de remplacer les badges]. C'est désormais chose faite. L'objet, pas plus grand qu'un grain de riz (1 cm de long pour 1 millimètre d'épaisseur), leur permet d'ouvrir les portiques de la société, mais également d'activer leur ordinateur⁶⁸.

↳ Le *retraçage* (géolocalisation) rapide et efficace à distance des personnes

L'exemple le plus frappant du *retraçage* est celui des personnes ayant un GPS qui possède une énergie propre lui permettant d'émettre un signal de manière autonome⁶⁹. Un suivi continu des objets, des animaux ou des personnes le possédant devient alors possible par simple appel du bon signal. Il permet de localiser en temps réel, et à distance, des personnes ainsi que des véhicules. C'est le cas pour les contrôles d'accès dans les parkings, la traçabilité des soins pour le suivi des patients dans les hôpitaux, le suivi de flotte de véhicules, etc.

Impact potentiel. Le « projet RFID »⁷⁰ de recherche et de développement, qui est régi selon un protocole bien précis, pourrait envisager d'utiliser cette technologie (ou des solutions proches) pour permettre, par exemple, à des policiers de surveiller et de suivre les déplacements d'anciens condamnés considérés comme dangereux (avec l'objectif de réduire les risques de récidive) ou encore à des professionnels de la santé de prendre soin à distance de personnes atteintes, entre autres, de la maladie d'Alzheimer.

67. A.L. Madinier (2011), « La surveillance de demain : puces RFID et implants sous-cutanés », Dans E. Labrot et P. Ségur (dir.), *Un monde sous surveillance?*, op. cit., voir le paragraphe numéroté 24 en marge.

68. « Travailleurs pucés : Peeters s'inquiète du respect de la vie privée », <<https://www.lalibre.be/actu/belgique/travailleurs-puces-peeters-s-inquiete-du-respect-de-la-vie-privée-58d2c749cd705cd98e199570>>.

69. Voir L.-A. Topfer (2016), « GPS Locator Devices for People with Dementia », *CADTH Issues in emerging health technologies Informing Decisions About New Health Technologies*, 147, <[https://owa.uqac.ca/owa/#viewmodel=IAttachmentViewModelPopoutFactory&AttachmentJson=%7B"type"%3](https://owa.uqac.ca/owa/#viewmodel=IAttachmentViewModelPopoutFactory&AttachmentJson=%7B)>.

70. Ooreka Entreprise, *Gérer la traçabilité des produits RFID active*, <<https://rfid.ooreka.fr/comprendre/rfid-active>>.

Différents projets proposés par la société VeriChip Corporation, basée à Delray en Floride et filiale de Applied Digital Solution (<<http://www.4verichip.com/index.htm>>), visent à développer un système sous-cutané de positionnement par satellite qui permettrait effectivement de localiser et d'identifier partout dans le monde des personnes à risque. En connexion avec un système d'information, la puce VeriChip (ou son successeur) permettra alors d'assister les équipes médicales au cas où le porteur se retrouverait, par exemple, seul dans le coma ou encore qu'il manifesterait des troubles de la mémoire⁷¹. Le 13 octobre 2004, la puissante et intransigeante Food and Drug Administration (FDA) américaine a autorisé la mise en marché de la première puce VeriChip à implanter dans le corps humain (en faisant de celle-ci le seul dispositif de ce type homologué à ce jour⁷²). Les projets d'implantation de Verichip se multiplient et se médiatisent, ce qui nous oblige à considérer au mieux les besoins et les éventuels systèmes RFID qui pourraient être mis en place. L'impact potentiel de ces projets est double :

- L'identification et la traçabilité permanente à distance des personnes, comme des personnes occupant un emploi à haut risque, des enfants susceptibles d'être kidnappés, des psychotiques, des personnes atteintes de démence ou des terroristes présumés, pour assurer un suivi continu plus efficace⁷³.
- L'information personnelle sur son dossier médical en temps réel permettant une meilleure prise en charge en cas d'urgence. Les services de secours seraient équipés de *scanners* pouvant détecter la puce RFID. En cas d'accident, les secouristes auraient directement accès aux données sur la puce.

Difficulté. Il serait difficile d'opérationnaliser l'analyse d'impact réel et potentiel pour éclairer les décisions à cette fin, étant donné que nous ne disposons pas de preuves solides concernant l'impact réel de la puce dont l'utilisation est peu connue chez l'humain. Jusqu'à quelle distance (ou degré) cette puce peut-elle retracer ? Il y a encore des limites techniques à surmonter :

Pour le moment, la puce Verichip n'est pas un système implantable de positionnement par satellite. Cependant, des recherches sont en cours visant à créer un

71. « Verichip » (10 février 2019), Wikipédia, <<https://fr.wikipedia.org/wiki/VeriChip>>.

72. Voir Prospective (Archive 2004), « Verichip : la première puce à être implantée dans le corps humain homologuée par la Food and Drug Administration américaine... ou « Big Brother inside »..., <<https://atelier.bnpparibas/prospective/article/verichip-premiere-puce-implanter-corps-humain-homologuee-food-drug-administration-americaine-big-brother-inside>>; Lavelle, S. (03/2010), « Implants, puces et transhumains », *op. cit.*

73. Groupe européen d'éthique des sciences et des nouvelles technologies (2005), *Aspects éthiques des implants TIC dans le corps humain, op. cit.*, p. 13.

système sous-cutané de positionnement par satellite qui permettrait effectivement de localiser et d'identifier partout dans le monde des enfants enlevés ou encore des terroristes présumés⁷⁴.

Il n'y a pas de données solides, présentement disponibles, provenant des expériences d'intervention de Verichip qui pourraient servir comme preuve de l'impact réel et potentiel :

VeriChip a récemment été rebaptisée positiveID. Si le nom de cette succursale laisse entendre [des bénéfices potentiels d'] un puçage RFID à ses clients, aucune information fiable ni solide n'a pu être récoltée sur ses expérimentations. Son site web ne présente aucun des produits évoqués [...]. Les seuls échos médiatiques de cette entreprise sont des reportages succincts de Fox News, ce qui laisse planer le doute quant à l'existence réelle de VeriChip et de ses VeriOffres⁷⁵.

Seriez-vous donc volontaires si on vous demandait de participer à de tels projets RFID qui pourraient avoir comme impact réel et potentiel que l'on vous identifie et trace à distance, ou encore que l'on obtienne des informations personnelles sur votre dossier médical en temps réel, permettant une meilleure prise en charge de votre situation en cas d'urgence ?

ii) Critère. Une fois que nous avons isolé l'impact réel et potentiel dans le contexte du développement du projet RFID, nous devons déterminer le critère qui va nous servir à évaluer l'impact (faire un jugement de valeur). Mais, avant de choisir un critère du jugement de valeur dans ce contexte, il faut identifier l'impact réel ou potentiel qui pourrait avoir différentes conséquences positives ou négatives dans la vie de la personne, de son entourage et de la société. Quels sont les principaux enjeux dans la vie de la personne, la société et les institutions, qui peuvent subir des impacts ? Le but, ici, est de spécifier les enjeux importants et retenus par le droit.

Enjeu majeur pour la santé (S) : les médecins, grâce à l'implantation d'une possible puce médicale dans le corps de tout individu, pourraient bénéficier d'un accès rapide aux informations personnelles de leurs patients en temps réel afin d'assurer une meilleure prise en charge, comme améliorer la surveillance médicale des patients à risque. Cette amélioration technologique implique, cependant, des risques pour la santé : des risques sanitaires (des réactions tissulaires, une migration du transpondeur implanté, des perturbations électromagnétiques ou encore des risques électriques (tous soulignés par la Food and Drug Administration), des risques de cancers (une augmentation de

74. A.L. Madinier (2011), « La surveillance de demain : puces RFID et implants sous-cutanés », Dans E. Labrot et P. Ségur (dir.), *Un monde sous surveillance?*, op. cit., voir le paragraphe numéroté 36 en marge.

75. Voir J.-P. Salaün (Publié le 20 novembre 2014 – Mis à jour le 8 novembre 2017), « Humain pucé, santé protégée? », <<https://archinfo01.hypotheses.org/1005>>.

10 % de cancer chez les rats selon des études américaines⁷⁶), et des risques de santé psychique (santé mentale ou psychologique) pour les patients se retrouvant avec un suivi médical moins humain. « Toute surveillance médicale ne peut pas être remplacée par un suivi connecté, notamment en ce qui concerne la psyché humaine⁷⁷. »

Enjeu majeur pour la sécurité des personnes (S) : malgré l'avantage potentiellement immense qu'offre la recherche et le développement d'une puce RFID implantée dans le corps à des fins de surveillance et de contrôle, qui devient plus intelligente et qui pourrait garantir la sécurité des personnes et leur traçabilité en cas de démence, d'enlèvement ou de catastrophe naturelle, plusieurs craignent que ce développement technologique implique le risque d'une société surveillée par « Big Brother » : « The most frightening implication of this technology is the grave possibility that it would facilitate totalitarian control of humans. [...] Not all the countries of the world prioritize autonomy, and the potential for sinister invasions of liberty and privacy are alarming⁷⁸. » Cependant, Molineri, dans « Faut-il avoir peur des puces RFID? », recommande plutôt d'éviter la « Paranoïa démesurée vs utilité relative » : « *ça reste une authentification de la personne et nous avons pour cela d'autres moyens sans que ce soit lié à notre corps*⁷⁹ ». Nous n'en sommes encore qu'à une *étape expérimentale où les chercheurs tentent d'améliorer de nouvelles versions de puces intelligentes* afin de pouvoir offrir une sécurité accrue dans la vie privée de chacun.

Enjeu majeur pour la vie privée (VP) : alors que la puce RFID offre l'avantage potentiel de fonctionner de la même façon qu'un téléphone portable et un « système GPS » (satellite) qui permettrait de retrouver rapidement des personnes aux fins médicales ou de sécurité, on craint aussi que son porteur risque de devenir totalement repérable depuis l'espace. C'est pourquoi le président de la Ligue des droits de l'Homme, Alexis Deswaef, est dubitatif concernant cette méthode d'implantation de la puce dans un cadre professionnel ou médical :

C'est un réel danger, on « flique » dorénavant les employés au plus profond de leur chair. C'est un outil de contrôle total. On peut savoir à quelle heure l'employé a commencé son service, quand celui-ci a pris sa pause cigarette. On analysera

76. Voir A.L. Madinier (2011), « La surveillance de demain : puces RFID et implants sous-cutanés », dans E. Labrot et P. Ségur (dir.), *Un monde sous surveillance?*, *op. cit.*

77. Voir J.-P. Salaiün (Publié le 20 novembre 2014 – Mis à jour le 8 novembre 2017), « Humain pucé, santé protégée? », *op. cit.*

78. E.M. McGee et G.Q. Maguire, « Ethical Assessment of Implantable Brain Chips », *Bioethics and Medical Ethics*.

79. H. Molineri (24 avril 2017), « Faut-il avoir peur des puces RFID? » Numerama, *op. cit.*

ensuite s'il est assez productif? [...] Dans le futur, braderons-nous un peu plus nos droits à la vie privée pour plus de sécurité ou de confort⁸⁰ ?

Il s'agit d'un enjeu important dans le domaine du droit étant donné qu'il est évident que nous concédons déjà, tous les jours, du terrain sur notre intimité en échange de bénéfices réels ou supposés des technologies RFID : paiement par carte, accès à un réseau sans fil, accès sécurisé (biométrie) et services de localisation par GPS. Ces pratiques se traduisent concrètement par un contrôle croissant de nos faits et gestes⁸¹. De nombreuses études démontrent aussi que les puces RFID actuelles peuvent être piratées, voire clonées : « Il ressort de cette réflexion que la généralisation des puces RFID porte atteinte à ce droit défendu par l'article 8 de la Convention Européenne de Sauvegarde des Droits de l'Homme et du Citoyen (Article 8 : "Toute personne a droit au respect de sa vie privée et familiale et de son domicile et d' [...]")⁸². »

Comme la morale, le droit pourrait donc alléguer que le développement de la puce intelligente à des fins de retraçage à distance risque éventuellement de violer nos droits fondamentaux, notamment le droit fondamental au respect de la vie privée et des données à caractère personnel.

Enjeu majeur pour l'environnement (E) : si les puces en développement offrent l'avantage de favoriser la santé et la sécurité de la personne, n'y a-t-il pas un risque pour l'environnement ? Déjà, les produits commerciaux dans leur cycle de vie, qui est la succession d'étapes de commercialisation que traverse un produit dans le temps, impliquent des risques de pollution écologique et de toxicité humaine en raison de leur faible biodégradabilité (puisque leur producteur en est responsable juridiquement jusqu'à son élimination⁸³). Cette responsabilité juridique du producteur est une préoccupation croissante pour le consommateur-citoyen qui veut, entre autres, être sûr que le produit ne fasse pas intercéder des éléments contraires à son éthique environnementale ou à sa morale, comme la destruction de ressources non renouvelables, une pollution non traitée, etc. :

Ces puces sont composées d'une puce en silicium et d'une antenne en cuivre ; elles sont souvent inséparables des objets sur lesquels elles sont apposées, sans

80. « Une entreprise belge utilise une puce électronique sous la peau de ses employés en guise de badge, <<https://www.lci.fr/sciences/une-entreprise-belge-utilise-une-puce-electronique-sous-la-peau-de-ses-employes-en-guise-de-badge-2025043.html>>.

81. D. Cerqui et B. Müller (2010), « La fusion de la chair et du métal : entre science-fiction et expérimentation scientifique », *Sociologie et sociétés*, *op. cit.*, p. 60.

82. Voir A.L. Madinier (2011), « La surveillance de demain : puces RFID et implants sous-cutanés », dans E. Labrot et P. Ségur (dir.), *Un monde sous surveillance?*, *op. cit.*

83. Voir C. London (mars 2001), La responsabilité juridique du producteur des déchets, Rapport final, ETUDE N° 00-0802/1A, <https://www.record-net.org/storage/etudes/00-0802-1A/rapport/Rapport_record00-0802_1A.pdf>.

compter le fait qu'elles sont presque invisibles. Cette inséparabilité des objets implique un risque environnemental en termes de recyclages des déchets. Or, dans la perspective de « l'Internet des objets », lorsque tout produit sera tagué par une étiquette RFID, les étiquettes risqueront d'être éliminées avec les produits qu'elles identifient, et ce avec le cuivre (et – s'il s'agit d'étiquettes actives – les piles) qu'elles contiennent. Issues de l'industrie de la microélectronique, ces puces RFID sont ainsi destinées à devenir à la fin de leur cycle de vie des déchets⁸⁴.

La question de la traçabilité des puces implantables chez l'humain, qui sont presque invisibles, devient alors aussi importante que celle de la traçabilité des nanoparticules au moyen de détecteurs habituelles : « Comment tracer des nanoproducts ou produits issus des nanotechnologies⁸⁵ ? » Devant l'importance de ces questions de la traçabilité des produits dans leur cycle de vie, le Comité consultatif national d'éthique en France déplore le manque de recherche fondamentale, ou encore de recherche non rendue publique⁸⁶. Ce manque de publications serait-il le fruit d'un manque de transparence ? D'un manque de dépenses sur le plan mondial, de ce qui a été consacré à la recherche pour servir à éliminer ce risque secondaire afin de favoriser l'avantage premier du Projet RFID ?

Difficulté. Pourquoi le droit choisit-il de privilégier ces enjeux plutôt que d'autres ? Encore une fois, la raison se trouve dans les caractéristiques de l'évaluation juridique, qui privilégie les impacts sur la personne, sa santé, sa sécurité et l'environnement. Il est donc normal que le droit accorde une importance particulière à ces enjeux selon son principe.

iii) Valeur correspondante. Contrairement à la morale, le droit ne traite pas seulement un enjeu, mais plusieurs. Et chaque fois que le droit traitera d'un enjeu, il devra faire une évaluation en fonction d'une valeur correspondante.

En santé, par exemple, la question se posera en fonction de la qualité de la santé de la personne : jusqu'où les puces peuvent-elles être un danger pour la santé de la personne et jusqu'où y a-t-il bienfait et nocivité ?

Pour l'environnement, la question se posera en fonction de la qualité de l'environnement : jusqu'où les puces peuvent-elles être un danger pour l'environnement ? On se demandera, alors, si la menace environnementale des puces est suffisamment grande pour justifier une réglementation.

84. C. Gossart (2010), Les enjeux environnementaux des technologies RFID, Entretien avec Laura Draetta, <<https://journals.openedition.org/terminal/1801>>.

85. Voir « Traçabilité », Wikipédia, <<https://fr.wikipedia.org/wiki/Tra%C3%A7abilit%C3%A9>>.

86. Comité consultatif national d'éthique pour les sciences de la vie et de la santé (janvier 2007), Avis n° 96, *Questions éthiques posées par les nanosciences, les nanotechnologies et la santé*, p. 12.

L'implantation de la puce (et non la puce elle-même) à des fins de retraçage peut aussi soulever des enjeux de consentement (liberté de choix) éclairé, notamment chez les mineurs et les personnes en perte d'autonomie. La personne a le droit de consentir ou non, en toute liberté de choix, plutôt que d'être assujettie à la dépendance et à l'instrumentalisation de l'homme, par exemple en étant considérée comme un instrument et uniquement pour son aspect utilitaire dans un projet de recherche RFID. Le choix éclairé repose sur une compréhension, aussi complète et raisonnable que possible, des buts de la recherche ; de ce qu'elle suppose, de ses avantages éventuels et des risques prévisibles, tant pour le participant que pour les autres personnes impliquées.

Mais le débat le plus éloquent, du moins celui qui a fait couler le plus d'encre, concerne la question de la violation du droit fondamental à la vie privée : jusqu'où la puce viole-t-elle cette liberté de la vie privée ? Le moindre viol de la vie privée est-il alors suffisant pour considérer que le droit se doit d'intervenir pour interdire ou limiter ?

Difficulté. La difficulté d'opérationnalisation d'une telle valeur de la vie privée pour faire le jugement servant à déterminer si le droit réglementaire doit intervenir ou non par une régulation, comme une réglementation adaptée, lorsqu'il y a un risque minimal d'atteinte aux libertés et droits de l'homme, peut s'exprimer par la question suivante : jusqu'à quel point (peu, moyennement, beaucoup) y a-t-il violation de la vie privée ? Le jugement de valeur est attributif (comme nous l'avons considéré dans le cas du tas de sable). La détermination d'une violation de la vie privée repose sur des données qualitatives et quantitatives, quand il convient de promouvoir une liberté, dont celle du consentement éclairé basé sur une information suffisante et adéquate des impacts et des risques potentiels, comme ceux relatifs aux recherches RFID, sur ces libertés et droit fondamentaux.

iv) Principe. Quel est le principe en cause dans l'évaluation juridique ? Le droit n'a pas un fondement absolu sur la vie privée. Son principe repose sur ce qui est raisonnable dans la société. Donc, qu'est-ce qui est raisonnablement acceptable pour les personnes ? Jusqu'où le partage de données privées peut-il être toléré en prenant en considération des bienfaits obtenus par les personnes ? Et jusqu'où le partage de ces données privées peut-il être considéré comme dangereux pour les personnes et la société avant d'intervenir ? De plus, quand on intervient, peut-on trouver des moyens pour éliminer les risques ? Et à quel moment, lorsqu'il est possible, intervient-on au nom du bien commun ou du respect de la personne ?

Difficulté. Comment déterminer ce qui est raisonnable au nom du bien commun, notamment de la vie privée ? La difficulté pour le droit est double (comme en éthique).

- 1) Déterminer le degré de violation de la vie privée (peu, beaucoup, énormément).
- 2) Prioriser entre l'autonomie (liberté de choix éclairée) de la personne et le degré de violation de la vie privée afin d'appliquer la règle de droit ou, le cas échéant, en statuer une nouvelle.
- 3) Pour l'autonomie (« condition d'une personne ou d'une collectivité autonome, c'est-à-dire qui détermine elle-même la loi à laquelle elle se soumet⁸⁷ »), la difficulté est double : liberté de contracter et liberté de commerce.

Le paradoxe est qu'il n'y a pas d'accord universel sur ce principe de l'autonomie. Le débat international actuel entre le transhumanisme et l'humanisme crée une ambiguïté à propos de l'autonomie permettant d'interdire ou non :

- A. **Autonomie selon le transhumanisme :** le transhumanisme ne tolère comme normatif que la conception libertarienne (issue du libéralisme ayant la liberté comme valeur fondamentale et naturelle) en support à la dignité-autonomie sans contrainte sociale (donc aucune réglementation qui entraverait la liberté personnelle) *versus*
- B. **Autonomie selon l'humanisme :** l'humanisme de type kantien ne tolère comme normatif que la liberté sous conditions (comme l'ensemble des droits et des devoirs qui constituerait une réglementation forte) en support à la morale commune pour interdire ou limiter⁸⁸.

Nous sommes ici en présence de deux définitions de l'autonomie, et aucune ne peut être prouvée vraie. À défaut de vérité, il est difficile de justifier ce qui est raisonnable dans l'évaluation juridique.

4. LE ROBOT D'ASSISTANCE AUX PERSONNES ÂGÉES

Le marché des robots d'assistance pour prendre soin des personnes âgées (« Assistive robot interacting with a physical therapy patient⁸⁹ ») évolue de plus en plus. Par exemple, l'Union européenne offre GiraffPlus, un projet de robot

87. A. Lalande (2002), « Autonomie », *Vocabulaire technique et critique de la philosophie*, *op. cit.*, p. 101.

88. Voir J.-P. Béland, G.A. Legault et J. Patenaude (2014), « Les enjeux identitaires de l'humain dans le débat philosophique sur la robotique humanoïde et l'amélioration humaine », *BioéthiqueOnline*, p. 9, <<http://bioethiqueonline.ca/3>>.

89. P. Lin (2012), « Introduction to Robot Ethics », dans P. Lin, K. Abney et G.A. Bekey (dir.), *Robots Ethics*, *op. cit.*, p. 23.

assistant combinant la robotique et la domotique, dont un certain nombre de personnes à travers l'Europe ont fait l'essai, notamment une Italienne nonagénénaire (Lea Mina Rally) qui tient même un blogue pour en parler⁹⁰. Est-ce que ces robots, en cours de développement, sont nos aides médicales du futur ? Est-ce grâce à eux que nous sommes en mesure de bien-vieillir à domicile ?

Ce projet de robot d'assistance servira ici d'exemple pour illustrer l'évaluation éthique qui ne guide pas la conduite par le biais de règles et d'obligations, mais par le biais de valeurs partagées, qu'il s'agit de pondérer en cherchant à déterminer la valeur prioritaire et à montrer dans quelle mesure les impacts augmentent ou diminuent cette valeur. Nous tenterons encore une fois de présenter les éléments de l'évaluation éthique (impact, critère, valeur correspondante et principe) et la difficulté de les opérationnaliser.

i) Impact. Essayons de voir de quel type de robot il s'agit dans le contexte du projet GiraffPlus pour en déterminer identifier l'impact réel et potentiel.

Impact réel. Le projet GiraffPlus, financé par l'Union européenne, parle d'un robot de téléprésence au sein d'un système complexe qui permet de surveiller les activités personnelles et domestiques d'un utilisateur à l'aide d'un réseau de capteurs, installés dans la maison et à proximité de son corps, permettant ainsi de révolutionner les soins à distance :

Les capteurs peuvent mesurer, par exemple, la pression artérielle ou détecter, par exemple, si une personne tombe. Différents services, selon les besoins de l'individu, peuvent être présélectionnés et adaptés aux besoins des personnes âgées et des professionnels de la santé. Au cœur du système se trouve un robot de téléprésence unique, Giraff, qui donne son nom au projet. Le robot utilise une interface de type Skype pour permettre, par exemple, à des proches ou à des personnes qui en prennent soin de rendre visite à une personne âgée à la maison⁹¹.

Quel est l'impact réel produit dans le contexte de l'essai du robot ? Une étude comme celle de Coradeschi *et al.* (2014)⁹², menée dans le contexte du projet de l'EU nommé GiraffPlus et de l'expérimentation de ce système déployé au sein de six foyers réels en Suède, en Italie et en Espagne, mesure les facteurs

90. GiraffPlus, <http://www.giraffplus.eu/index_option_com_content_view_frontpage_Itemid_54_lang_en.html>.

91. GiraffPlus, « 2015-04-17 Excellents résultats pour GiraffPlus lors de l'examen final », <http://www.giraffplus.eu/index_option_com_content_view_frontpage_Itemid_54_lang_en.html>.

92. Voir Coradeschi *et al.* (2014), « GiraffPlus : A System for Monitoring Activities and Physiological Parameters and Promoting Social Interaction Elderly », dans Hippe, Z.S., Kulikowsky, J.L., Mroczek, T., Wtorek (dir.) *Human-Computer System Interaction: Backgrounds and Applications 3*, Springer International Publishing, p. 261-271, <https://www.researchgate.net/publication/263849499_GiraffPlus_A_System_for_Monitoring_Activities_and_Physiological_Parameters_and_Promoting_Social_Interaction_for_Elderly>.

(besoins et intérêts des utilisateurs potentiels âgés et des professionnels de la santé) qui ont un impact révélateur sur sa commercialisation. De ce point de vue, l'étude présente les fonctionnalités d'un système complexe capable de surveiller les activités et les paramètres physiologiques de l'utilisateur, via un réseau de capteurs distribués dans la maison. Ce réseau de capteurs (capteur de porte et de fenêtre, détecteur de mouvement, station de tension artérielle, etc.) vise à augmenter la sécurité et la facilité de surveillance selon les besoins des utilisateurs potentiels âgés et des professionnels de la santé. Le robot télé-opéré (semi-autonome) de téléprésence permet à l'utilisateur âgé de communiquer facilement avec les professionnels de la santé et se déplace sous le contrôle de cet utilisateur au sein du réseau configuré dans la maison. L'impact réel recherché de la fonctionnalité du système du robot peut s'entendre alors comme « l'acceptabilité sociale du robot⁹³ » qui renvoie à l'examen des conditions (besoins et intérêts des utilisateurs) qui rendent ce système du robot acceptable selon ces utilisateurs (hôpitaux, médecins, professionnels de la santé, personnes âgées).

Ce type d'études sur l'acceptabilité sociale cherche à prédire et à faire valoir les conditions d'acceptation afin de mettre en place des stratégies de communication favorisant les avantages souhaités pour inciter les gens à acheter et à utiliser le système de robot Giraff (semblable à d'autres robots produits par des sociétés concurrentes). Un site de vente comme <telepresencerobots.com> se fonde sur ce genre d'étude pour vanter les avantages du robot Giraff sur les autres robots de surveillance médicale conçus aux fins d'amélioration des soins médicaux :

Les hôpitaux et les médecins du monde entier commencent à utiliser des robots de téléprésence pour améliorer les soins médicaux prodigués. Bien que de nombreux modèles de robots soient utilisés dans le domaine médical, le Giraff Plus semble avoir un avantage net sur les autres, après avoir été conçu spécifiquement pour des fonctions d'observation et de surveillance médicales. Soutenu par des subventions de recherche de l'Union européenne et animé par une équipe de scientifiques et de chercheurs principalement originaires de Suède et d'Italie, Giraff propose un robot de téléprésence dont les accessoires permettent aux médecins de fournir une profondeur sans précédent aux soins des clients⁹⁴.

Dans ce contexte de la mise à l'essai et de la vente de ce genre de système technologique, le discours de l'expérimentation cherche à rendre acceptable le

93. Voir Terrade *et al.* (2009), « L'acceptabilité sociale : la prise en compte des déterminants sociaux dans l'analyse de l'acceptabilité des systèmes technologiques », *Le travail humain*, Presses universitaires de France, 4, vol. 72, p. 384, <<http://www.cairn.info/revue-le-travail-humain-2009-4-page-383.htm>>.

94. <<https://telepresencerobots.com/reviews/giraff/giraff-plus-revolutionizing-remote-patient-care>>.

système du robot en développement qui prétend rendre meilleur (changer en mieux) le prendre soin pour le prolongement de la vie autonome de la personne âgée à la maison : « The GiraffPlus system aims at facilitate a prolonged independent living for elderly in their own home⁹⁵. » L'impact réel recherché est le prolongement de la vie autonome de la personne âgée qui se sent moins inquiète pour l'avenir et faire en sorte que les proches se sentent rassurés. Est-ce que ce robot apporte réellement cette tranquillité d'esprit à la personne âgée ainsi qu'à ses proches ? Tel est l'effet réel produit par le système du robot (tant qu'il fonctionne bien), comme en témoigne Lea Mina Rally (cette *mammy* italienne de 94 ans qui, en 2014, utilisait le système GiraffPlus depuis 5 mois et qui parle volontiers du robot qu'elle surnomme « Robin » sur son blogue). Ce genre de témoignage fait ainsi rendre compte de l'adéquation entre discours et réalité : « Les gens se demandent pourquoi je n'habite pas chez ma fille, mais elle a aussi des petits-enfants et beaucoup de nouvelles responsabilités. Avec ce précieux assistant, que j'appelle "Robin", je suis moins inquiète pour l'avenir, et mes enfants et petits-enfants sont rassurés⁹⁶. »

Impact potentiel. « Les examinateurs de l'Union européenne estiment que la réussite du projet contribuera à la création de solutions intégrées de supervision à distance pour une prochaine génération et pourra avoir un impact important sur le système européen de protection sociale⁹⁷ », qui pourrait ainsi contribuer à relever des défis communs (« vieillissement de la population, montée des inégalités et de l'exclusion, mutation du marché du travail, problème de financement de la protection sociale suite à l'augmentation des dépenses⁹⁸ »). Est-ce que les systèmes de soins par robot, « robot care systems », sont dorénavant nos aides de demain qui nous permettra de trouver des solutions à ces nouveaux défis de protection sociale ? Compte tenu de l'augmentation croissante de la population de personnes âgées dans la plupart des pays industrialisés du monde et de la réussite de la fonctionnalité du système de type GiraffPlus, ceci pourrait représenter un premier pas vers le développement d'autres systèmes robotiques encore plus sophistiqués permettant de révolutionner les soins à distance. Nous pouvons alors considérer que l'impact potentiel du développement de tels systèmes technologiques s'annonce comme une triple solution à

95. Coradeschi *et al.* (2014), « GiraffPlus: A System for Monitoring Activities and Physiological Parameters and Promoting Social Interaction Elderly », *op. cit.*, p. 270.

96. Commission européenne, Base de données des communiqués de presse, « Robot prend soin de Lea, une mammy italienne de 94 ans (Bruxelles, le 5 mai 2014), <http://europa.eu/rapid/press-release_IP-14-515_fr.htm>.

97. Les Services publics et l'Union européenne, « La protection sociale en Europe », *Europa*, p. 1, <<http://europaong.org/wp-content/uploads/2016/06/Protection-sociale.pdf>>.

98. *Ibid.*

ces défis (mentionnés ci-dessus) auxquels sont confrontés les institutions et les citoyens concernés par les systèmes de protection sociale à travers le monde :

- aider la personne âgée à demeurer le plus longtemps possible en sécurité dans sa maison ;
- améliorer le suivi du traitement médical avec l'aide d'une personne qualifiée à distance ;
- répondre à la pénurie de main-d'œuvre.

Difficulté. Il n'y a pas de données suffisantes pour prouver un tel impact réel et potentiel. C'est ici qu'apparaissent les difficultés d'opérationnaliser l'analyse d'impact. D'une part, le débat qui se préoccupe de la qualité des études d'acceptabilité sociale des impacts de tels systèmes robotiques et des preuves fournies ; et d'autre part, une seconde difficulté : est-ce que l'on doit attendre des preuves certaines d'analyse d'impact avant de poursuivre l'évaluation ? Procéder ainsi risque de nous mettre devant un fait accompli, c'est-à-dire l'état de fait actuel ou projeté que la grande majorité de la population accepte ou accepterait, ou refuse ou refuserait d'assumer le risque. C'est pourquoi l'évaluation d'une technologie repose sur les impacts réels (lorsqu'il y a des études), mais aussi sur les impacts potentiels (même s'ils ne sont pas encore étudiés) qui sont probables, compte tenu du système déjà mis en cause dans la société.

ii. Critère. Une fois que nous avons isolé l'impact réel et potentiel (le triple impact mentionné ci-dessus) dans ce contexte du développement et de la commercialisation du robot de téléprésence, nous devons déterminer quels enjeux sont suffisamment importants pour faire le choix du critère qui va nous servir à évaluer l'impact (faire un jugement de valeur). Car chaque enjeu soulève déjà un critère pour évaluer l'impact. Puisque cet impact modifie le mode de vie personnel et collectif de la personne âgée, de son entourage et de la société, il y a plusieurs enjeux. Voyons plus précisément quels sont ces enjeux les plus importants à analyser en éthique afin de déterminer le critère qui servira de base à l'évaluation.

Enjeu majeur pour la santé (S) : ces systèmes de robots de téléprésence, à l'instar de GiraffPlus, offrent l'avantage d'améliorer le suivi du traitement médical auprès d'une personne à distance. Puisque « les accessoires [du système] permettent aux médecins de fournir aux patients une profondeur [d'aide] sans précédent » pour faciliter la détection précoce de la détérioration possible de la santé en raison du vieillissement, mais que cela implique le risque d'effets psychologiques sur la personne elle-même : « l'isolement social, un important

facteur du risque de mortalité prématurée⁹⁹ », on ne peut dépendre totalement du remplacement de l'interaction humaine par la mise en place du soin par des robots.

Enjeu majeur pour les relations humaines (RH) : le système d'assistance GiraffPlus offre certes à la personne âgée l'avantage de vivre avec un robot de service sécuritaire (dont le fonctionnement a pour but de réduire son sentiment de solitude et de l'aider à faire face à ses déficiences, soit en maintenant une activité physique ou en entraînant les fonctions cognitives), ce qui favoriserait le bien-être (absence de soucis, bonheur d'être à la maison) de cette personne, comme en témoigne Lea Mina Rally, qui se sent moins inquiète avec son robot qu'elle appelle « Robin » ; mais cela augmente le risque de l'isolement social par la perte de contact avec les humains (professionnels de la santé, proches et aidants naturels) :

L'isolement social réfère à une carence de contacts interpersonnels. Il s'agit d'un paramètre objectif qui peut être mesuré en examinant si une personne vit seule, son statut conjugal, la taille de son réseau social et sa participation à des activités de groupe. L'isolement social est un problème particulièrement fréquent à des âges avancés lorsque la diminution des ressources économiques, l'invalidité ou encore le décès des personnes proches contribue à diminuer les contacts sociaux¹⁰⁰.

Ce risque d'isolement social est d'autant plus important que le débat international sur les technologies d'assistance à la maison, désigné par le terme SLL – Ambient Assisted Living – se concentre essentiellement sur des technologies qui permettent aux personnes âgées de rester aussi longtemps que possible chez elles de façon sécurisée, en ayant un robot de service à la maison, sans se soucier du fait que ces technologies, loin de constituer une solution au problème standard de l'isolement social, risquent de l'aggraver. Les facteurs aggravants dans le contexte de l'étude des usages des systèmes des robots d'assistance en cours de développement sont les suivants :

- Les robots en développement un peu partout dans le monde deviennent de plus en plus utiles ;
- Les robots sont au service de ces patients pour leur apporter de l'aide en cherchant à éviter de leur faire du mal en raison parfois de leur impatience ou de leur frustration (« Generally speaking, the more human-like the technology, the greater the reaction will be. People cooperate with sufficiently human-likes machines, are polite to them,

99. M. Juneau (2017), « L'isolement social, un important facteur de risque de mortalité prématurée », Observatoire de la prévention, <<https://observatoireprevention.org/2017/05/03/lisolement-social-important-facteur-de-risque-de-mortalite-prematuree>>.

100. *Ibid.*

decline to sustain eye contact, decline to mistreat or roughhouse with them, and respond positively to their flattery¹⁰¹. »);

- Les patients s'attachent à ces robots («An extensive literature in communications and psychology demonstrates that humans are hardwired to react to social machines as though a person were really present¹⁰². »), de sorte qu'une tentative de les en distancer pourrait causer une détresse significative.

Enjeu majeur pour l'économie sociale et solidaire (ESS) : ces systèmes de robots d'assistance, en répondant aux besoins de la collectivité, offrent l'avantage de retombées économiques importantes qui peuvent permettre de répondre aux défis du système économique de protection sociale : pour faire face à la pénurie de main-d'œuvre, à l'augmentation de la population vieillissante et au problème de financement de la protection sociale à la suite de l'augmentation des dépenses en santé et services sociaux, mais cela implique aussi un risque de perte d'emploi pour les professionnels de la santé, donc une augmentation des conflits socioéconomiques.

Enjeu majeur pour la vie personnelle et privée (VPP) : tous ces systèmes de robots de téléprésence opérant au sein d'un réseau de capteurs tel un système technique structuré de télésurveillance à distance, comme le démontre l'étude de l'architecture du GiraffPlus par Girolami *et al.*, en 2013, offrent l'avantage d'améliorer la vie personnelle des personnes vieillissantes en les aidant à aborder ce nouvel épisode de leur vie de la meilleure manière possible en ayant recours à un système de surveillance sécuritaire et interactif : « Surveillance des données vitales – Communication électronique – Systèmes de sécurité (alarme et système d'appel d'urgence) – Animation par des objets technologiques de réalité simulée (robots) – Enregistrement, saisie de données et documentation¹⁰³. » Mais cet avantage d'une nouvelle vie personnelle plus sécuritaire favorisant le bien-être (bonheur d'être) à la maison, grâce au système d'un robot d'assistance, est comme une fenêtre qui s'ouvre sur le cœur de la maison (« Robots : A Window into the Home¹⁰⁴ »), et implique le risque de la « tyrannie du Bigbrother » (« Bigbrother syndrome¹⁰⁵ ») qui suppose qu'une surveillance directe et le

101. M.R. Calo (2012), « Robots and Privacy », dans P. Lin, K. Abney et G.A. Bekey (dir.), *Robots Ethics*, *op. cit.*, p. 192.

102. *Ibid.*

103. H. Rüegger, D. Roulet Schab et N. Egger (2016), *Aspects éthiques liés à l'utilisation des technologies d'assistance dans les institutions de soins de longue durée (EMS)*, *op. cit.*, p. 5. *Ibidem.*

104. M.R. Calo (2012), « Robots and Privacy », dans P. Lin, K. Abney et G.A. Bekey (dir.), *Robots Ethics*, *op. cit.*, p. 192.

105. Coradeschi *et al.* (2014), « GiraffPlus : A System for Monitoring Activities and Physiological Parameters and Promoting Social Interaction Elderly », *op. cit.*, p. 270.

contrôle de la vie personnelle par des personnes inconnues est partout, donc qu'il n'y a plus de vie privée.

Enjeu majeur pour la vie autonome (VA) : ces systèmes sécuritaires des robots de téléprésence offrent l'avantage de prolonger la vie autonome et indépendante de la personne âgée, mais cela implique comme désavantage le risque que cette personne vieillissante perde le droit de dire son mot dans la détermination de son avenir, donc d'une perte du droit à l'autodétermination (liberté de choix) de la vie prolongée de la personne qui ne serait pas le résultat d'une contrainte. L'avantage d'être sous la surveillance et le contrôle du système sécuritaire implique aussi le risque de perdre la liberté de décider des conditions et des moyens possibles pour assurer une vie prolongée.

Difficulté. Pourquoi l'éthique choisit-elle de privilégier l'ensemble de ces enjeux comme étant prioritaires au critère d'évaluation ? La difficulté ici est de justifier le choix de l'ensemble de ces enjeux que l'on va prendre en compte en considérant chacun à la valeur correspondante qu'on y trouve.

En droit, comme nous l'avons vu dans le cas précédent, les enjeux retenus relèvent principalement de la santé, de la sécurité et de l'environnement (EHS – Environment, Health, Security) alors que les enjeux éthiques, légaux et sociaux (ELSI – Ethical, Legal and Social Issues) ne sont pas nécessairement retenus, conformément au positivisme juridique fondé sur une démocratie républicaine, ou d'une démocratie libérale¹⁰⁶. Cela ne veut pas dire que le droit – dans son ensemble – exclut les enjeux ELS. Il faut surtout comprendre que le droit réglementaire qui vise à encadrer l'implantation de ces dispositifs se concentre davantage sur la considération des enjeux EHS au détriment des enjeux ELS. Par exemple, si le débat démocratique international sur les technologies d'assistance, désigné par le terme SLL – Ambient Assisted Living¹⁰⁷ – se concentre essentiellement sur les technologies électroniques installées au domicile et à la télécommunication externe avec du personnel médical et de soin, ou avec tout autre prestataire de service pertinent, afin que les personnes âgées puissent recevoir, quand cela s'avère nécessaire et à tout moment, l'aide nécessaire et le soin approprié leur permettant de rester aussi longtemps que possible et de façon sécurisée à la maison, le droit aura tendance à prioriser l'assistance robotique qui favorisera les impacts positifs sur la santé et la sécurité

106. Voir J. Patenaude *et al.* (2014), « Framework for the Analysis of Nanotechnologies' Impacts and Ethical Acceptability: Basis of an Interdisciplinary Approach to Assessing Novel Technologies », *op. cit.*

107. Voir H. Rüegger, D. Roulet Schab et N. Egger (2016), *Aspects éthiques liés à l'utilisation des technologies d'assistance dans les institutions de soins de longue durée (EMS)*, Édition CURAVIVA Suisse, p. 6, <https://www.curaviva.ch/files/XO1YCQP/brosch_a4_fr_ethische_aspekte_web.pdf>.

de la personne âgée, tout en diminuant les risques possibles pour sa vie personnelle et privée.

En morale, comme nous l'avons vu dans les deux premiers cas, soit 1. Les implants de mémoire BCI et 2. Les systèmes d'aide à la décision, le choix se focalise sur un enjeu qui vise les impacts sur les dimensions fondamentales (nature, dignité, intégrité), qui sont elles-mêmes fondées en théories morales sur des principes (principe de la nature-identité humaine, principe de la dignité-autonomie). Par exemple, si le développement des technologies d'assistance risque d'abaisser la dignité-autonomie de la personne âgée en augmentant chez elle une dépendance artificielle (dépendance au robot de service), l'évaluation portera sur le choix de l'enjeu prioritaire de la vie autonome en tant que critère du jugement moral, et ce choix se fera à partir du principe de la dignité-autonomie afin d'interdire tout risque que les technologies servant à alléger la démarche de l'assistance humaine implique une dépendance artificielle pour la personne âgée :

L'assistance technique [...] ne doit pas [...] remplacer l'assistance humaine, mais l'encadrer (Nass, 2014, p. 132). Elle doit « décharger le soin, pas le remplacer. » (CURAVIVA Suisse, 2014a). [...] Ce faisant elle libère le personnel soignant qui peut se tourner vers sa tâche véritable, c'est-à-dire le soin et l'accompagnement personnalisé des personnes âgées. [...] « La technique doit servir l'humain et s'adapter à ses besoins, souhaits et parcours de vie – non pas l'inverse » (Manzeschke *et al.*, 2013, p. 26). Cela signifie que la technologie ne doit pas être utilisée chez une personne âgée en remplacement de capacités encore existantes, afin de ne pas conduire au développement d'une dépendance artificielle¹⁰⁸.

En éthique, l'évaluation porte sur l'ensemble des enjeux prioritaires. Ce sont tous les impacts possibles sur toutes les dimensions de notre vie (enjeux E³LS – environnement, économie, éthique, loi et société) qui peuvent faire l'objet d'une évaluation éthique (évaluation de tous les impacts sur les enjeux).

iii) Valeur correspondante. L'évaluation éthique ne guide pas la conduite par le biais de règles et d'obligations morales, mais par le biais de valeurs partagées qu'il s'agit de pondérer en cherchant à *établir* la valeur prioritaire et à montrer dans quelle mesure les impacts augmentent ou diminuent cette valeur. Voyons cela d'un peu plus près.

Nous supposons que les enjeux priorisés ci-dessus (la santé humaine, les relations personnelles, les retombées économiques sur le système de protection sociale, la vie privée et l'autonomie de la personne qui subissent les impacts

108. H. Rüegger, D. Roulet Schab et N. Egger (2016). *Aspects éthiques liés à l'utilisation des technologies d'assistance dans les institutions de soins de longue durée (EMS)*, Édition CURAVIVA Suisse, p. 5, <https://www.curaviva.ch/files/XO1YCQP/brosch_a4_fr_ethische_aspekte_web.pdf>.

positifs et négatifs du système d'assistance du robot) sont considérées comme suffisamment importants dans une société pour les qualifier en termes de valeurs ; ils ont de la valeur, si par « valeur » nous entendons ce qui fait l'objet d'une préférence, ce qui est estimé, préféré, désiré.

On peut associer à chacun de ces enjeux la valeur correspondante qu'on y trouve. Généralement, il s'agit de la qualité. Par exemple, pourquoi l'enjeu de la santé est-il pour nous si important ? Parce que nous cherchons à maximiser la qualité de la santé pour chacun de nous. Si le système du robot Giraff contient des capteurs servant à la surveillance des données vitales et à la communication par le robot, c'est pour maximiser la qualité de la santé des personnes âgées, il cherche donc à maximiser la qualité de la santé pour la personne âgée. Par contre, si la personne âgée est exposée au risque d'effet psychologique sur elle-même (au risque de remplacer l'interaction humaine thérapeutique par le robot auquel elle s'attache), cela minimise la qualité de la santé humaine.

Donc, lorsque nous faisons l'évaluation éthique des impacts sur les enjeux, nous essayons de faire un consensus entre les participants sur les valeurs qui sont associées aux enjeux et qui renvoient plus précisément à la qualité de tout ce qui était visé par le développement et la commercialisation du système du robot :

- i) qualité de la santé humaine (physique et psychologique),
- ii) qualité des relations humaines,
- iii) qualité des retombées économiques sur le système de protection sociale,
- iv) qualité de la vie personnelle et privée dans la maison,
- v) qualité de l'autonomie (liberté de choix) de la vie prolongée de la personne vieillissante.

Difficulté. La première difficulté d'opérationnalisation du jugement de valeur consiste à nous entendre sur le degré de maximisation ou de minimisation de la qualité engendré par l'impact positif (avantage) ou négatif (risque) potentiel du système d'assistance du robot (comme Giraff). Si nous nous fions à notre désir, nous voulons tous la qualité totale : i) qualité totale de notre santé, qualité totale des relations personnelles, qualité totale des retombées économiques, qualité totale de notre vie personnelle et privée, qualité totale de l'autonomie (liberté de choix) de notre vie prolongée. Autrement dit, nous désirons tous le paradis terrestre, le jardin d'Éden tant convoité.

L'évaluation va ainsi varier d'une personne à l'autre selon le degré d'atteinte ou de diminution (peu, moyennement, beaucoup, presque totalement) de la qualité (valeur) totale envisagée :

- i) Jusqu'à quel point le système du robot augmente-t-il ou diminue-t-il la qualité de la santé de la personne âgée? L'augmentation est-elle majeure ou mineure?
- ii) Jusqu'à quel point le système du robot diminue-t-il la qualité des relations entre les personnes (personne âgée, proches, aidants naturels et professionnels)? La diminution est-elle majeure ou mineure?
- iii) Jusqu'à quel point le système du robot augmente-t-il la qualité des retombées économiques pour répondre aux défis du système de protection sociale? L'augmentation est-elle majeure ou mineure?
- iv) Jusqu'à quel point le système du robot diminue-t-il la qualité de la vie personnelle et privée? La diminution est-elle majeure ou mineure?
- v) Jusqu'à quel point le système du robot diminue-t-il la qualité de l'autonomie (liberté de choix) de la vie prolongée? La diminution est-elle majeure ou mineure?

Quel sera le jugement final d'évaluation permettant de répondre à chacune de ces questions? Le problème est qu'il n'y a pas de données probantes comme une approche permettant d'intégrer les meilleures preuves scientifiques disponibles, l'expérience clinique et en consultation avec les personnes âgées, afin de choisir quelle pondération des jugements finaux de valeurs s'applique le mieux pour résoudre les deux conflits de valeurs suivants :

- i) Jusqu'à quel point les gains en qualité de santé sont-ils supérieurs aux pertes en qualité des relations humaines, et vice versa?
- ii) Jusqu'à quel point les gains en retombées économiques de ce développement des systèmes d'assistance pour répondre aux défis du système de protection sociale sont-ils supérieurs en perte de qualité de l'autonomie (liberté de choix) de la vie prolongée et privée des personnes âgées, et vice versa?

Le dialogue entre les personnes impliquées en éthique devient particulièrement utile, en contexte de conflits de valeurs et d'incertitude sur les faits, pour établir : dans un premier temps, l'accord sur le principal conflit de valeurs agissantes, et ensuite, le résoudre en établissant la valeur prioritaire permettant d'aider ces personnes à prendre la meilleure décision consensuelle finale (jugement de valeur final), et ce, en maintenant un équilibre entre la valeur prioritaire et les autres valeurs conflictuelles secondaires selon : 1) l'efficacité du moyen (modalité et cadre) de l'action pour atteindre la fin (la valeur prioritaire), et 2) la diminution des inconvénients (c'est-à-dire réduction au minimum des impacts négatifs sur les autres valeurs) par le moyen. Mais

qu'est-ce qui nous fera dire que nous prendrons la meilleure décision consensuelle dans la pondération ?

iv) Principe. Le principe justifiant une telle décision dans le processus de l'évaluation en éthique est le bien vivre-ensemble. Ce vivre-ensemble devient dès lors le fondement d'une « vie bonne, avec et pour autrui, dans des institutions justes » (Ricœur). C'est dire qu'en éthique, des institutions justes peuvent contribuer à un meilleur vivre-ensemble. C'est le terme d'une visée éthique des institutions justes¹⁰⁹.

Difficulté. « Les institutions justes sont-elles possibles¹¹⁰ ? » La justesse de la pondération (balance) des décisions pour assurer la qualité du bien-vivre en commun des personnes impliquées dans des institutions justes, comme le sont par exemple appelés à déterminer la Commission gouvernementale de l'éthique de la science et de la technologie au Québec, le Comité consultatif national d'éthique et la Haute Autorité de Santé en France, qui ont pour mandat d'évaluer l'acceptabilité sociale et éthique de développements technologiques dans la société et les systèmes de la santé, demeure un travail difficile. Il n'y a pas que la difficulté de mettre en œuvre une théorie éthique ou une approche commune qui est uniquement procédurale en maximisant les avantages potentiels et en réduisant au minimum les risques potentiels pour favoriser la justice des décisions pour toutes les personnes impliquées. Toute décision des institutions décisionnelles s'élabore sur fond de tension sociale et de visées (valeurs) contradictoires.

La difficulté n'est pas que de résoudre, cependant, le débat international entre l'ultralibéralisme économique sans restriction que défendent des transhumanistes et la critique des humanistes antilibéraux qui en reviennent à la question des limites et au renouvellement des régulations (« Comment les envisager au niveau mondial ? Existe-t-il des alternatives au pouvoir techno-économique tel qu'il est vu dans le cadre de l'ultralibéralisme afin [...] de préserver la liberté et la justice¹¹¹ ? »), mais de voir comment, en décidant d'autoriser ou non, avec ou sans conditions, les décideurs vont implicitement prendre en compte le contexte des tensions sociales du point de vue économique.

109. J.-M. Gaté (2001), « Entretien avec Paul Ricœur », *Le philosophe*, 3, n° 15, <<https://www.cairn.info/revue-le-philosophe-2001-3-page-9.htm>>.

110. J.-L. Schlegel (2017), « Paul Ricœur : "Une vie bonne, avec et pour autrui dans des institutions justes" », RadioFrance, Conférences, <<https://www.franceculture.fr/conferences/revue-esprit/paul-ricoeur-penseur-des-institutions-justes>>.

111. T. Magnin (2017), *Penser l'humain au temps de l'homme augmenté. Face aux défis du transhumanisme*, Paris : Albin Michel, p. 157.

Autrement dit, la question économique est une pression exercée sur la prise de décision. Cela ne fait pas de l'institution quelque chose de moins juste. Pour Ricœur, l'institution juste est celle qui a été créée par des règles démocratiques.

CONCLUSION

Quelle conclusion pouvons-nous tirer de l'analyse de ces difficultés d'opérationnaliser l'évaluation morale, juridique ou éthique à la lumière des débats qu'elle soulève pour chacun de ces quatre produits ?

Le premier niveau de difficulté est de mesurer les impacts positifs et négatifs et, à défaut de mesures certaines, prendre des décisions sociales incontournables : interdire, réglementer ou favoriser.

Le second niveau de difficulté désigne le sens pratique de notre démarche intellectuelle dans le processus simplifié d'évaluation (*i. Impact, ii. Critère, iii. Valeur correspondante, iv. Principe*) qui montre : 1) qu'il importe de suivre une analyse ascendante (dite *bottom-up*) où nous partons de l'échelon le plus « bas » (au sens hiérarchique) ou le plus « terrain » (au sens opérationnel) pour être pris en compte par les échelons supérieurs dans ce processus de l'évaluation, et 2) que nous serons toujours coincés dans la difficulté de prendre des décisions en manque de certitudes sur le plan des faits et des valeurs¹¹².

Par contre, l'enjeu éthique (le plus haut niveau de difficulté) est d'assurer la qualité du vivre-ensemble et le « comment » les diverses institutions peuvent jouer leur rôle dans le domaine.

Il nous faudra insister sur le rôle des institutions en place pour relever le défi de cette évaluation complexe qui suit une démarche intellectuelle *bottom-up* en vue du meilleur vivre-ensemble.

Il sera alors utile de distinguer différentes façons de réguler le défi posé :

- le rôle du droit : interdire ou limiter ce qui est jugé intolérable ;
- les institutions conseils (CEST, INESSS, CCNE, HAS, etc.), dont la tâche est de prendre en compte ces difficultés pour recommander aux instances des choix ;
- les entreprises qui développent les technologies avec l'intégration de l'éthique selon l'approche des innovations responsables.

Il nous faudra, en même temps, affermir la nécessité du dialogue interdisciplinaire qui désigne le pilotage participatif (ascendant) où l'animation de

112. L'éthique n'est-elle pas un juste rapport à l'incertitude ?

l'évaluation *bottom-up* commence en éthique appliquée. Il n'en demeure pas moins que le dialogue est le seul moyen susceptible d'ouvrir les portes du vivre-ensemble. C'est la pierre angulaire de la maison commune.

Présentation des auteurs

Jean-Pierre Béland

Jean-Pierre Béland, détenteur d'un Ph. D. et d'un postdoctorat en philosophie de la Sorbonne, est professeur en éthique au Département des sciences humaines et sociales à l'Université du Québec à Chicoutimi. Membre du groupe de recherche InternE³LS au 3IT de l'Université de Sherbrooke, ses recherches portent plus particulièrement sur la gouvernance du développement responsable de l'innovation technologique en robotique et en intelligence artificielle. Il a publié plusieurs ouvrages sous sa direction, dont *L'homme biotech : humain ou posthumain?* (2005) et *Asimov et l'acceptabilité des robots* (2012). Il est co-directeur de la collection « Enjeux éthiques contemporains » aux Presses de l'Université Laval.

Pascale Boucaud

Docteur en droit et spécialiste de droit international des droits de l'homme, Pascale Boucaud a fondé l'Institut des droits de l'Homme de Lyon, a été titulaire de la Chaire UNESCO « Mémoire, Cultures et Interculturalité » et est actuellement doyenne du pôle facultaire Droit, Sciences économique et sociales de l'Université catholique de Lyon. Membre du Laboratoire de recherche sur la personne, ses recherches portent actuellement sur les droits de la personne en son corps et plus particulièrement sur les droits des personnes vulnérables confrontées aux nouvelles technologies.

Christian Byk

Christian BYK est juge à la Cour d'appel de Paris et préside le comité d'éthique des sciences de la Commission française pour l'UNESCO. Spécialiste de droit international et du droit des sciences de la vie et des techniques, il a rédigé le premier projet de Convention européenne sur la biomédecine et les droits de l'homme, et il participe aux activités normatives de l'UNESCO dans

le domaine de la bioéthique et de l'éthique des sciences. Il dirige depuis 1989 l'Association internationale droit, éthique et science, est membre fondateur de l'International Association of Bioethics et préside le Comité intergouvernemental de bioéthique de l'UNESCO. Il a été professeur de droit à l'Université de Poitiers et est professeur invité dans différentes universités ; il a créé le Journal international de bioéthique et d'éthique des sciences et est rédacteur en chef de la revue *Droit, santé et société*. Il a publié une dizaine d'ouvrages et environ 300 articles.

Charles-Étienne Daniel

Charles-Étienne Daniel est avocat et professeur de droit et technologies aux Facultés de droit et de génie de l'Université de Sherbrooke. En tant que membre du groupe interdisciplinaire de recherche InternE³LS, ses intérêts de recherche portent sur le droit, l'éthique et la gouvernance du développement technologique, dans les domaines des biotechnologies, des nanotechnologies, de la robotique et de l'intelligence artificielle. Il s'intéresse plus particulièrement aux forces, aux limites et aux dynamiques à l'œuvre dans l'encadrement normatif du développement des technologies émergentes.

Mathieu Guillermin

Docteur en physique et en philosophie, Mathieu Guillermin est maître de conférences à l'Université catholique de Lyon et membre du Groupe d'épistémologie et d'éthique des sciences et des technologies du Laboratoire de biologie générale. Au sein de cette équipe, il explore notamment les articulations entre démarche éthique, recherche scientifique et développement technologique. Le cœur de son travail de recherche et d'enseignement porte sur les questions philosophiques et éthiques soulevées par les nouvelles technologies numériques (*big data*, intelligence artificielle, robots).

Mathieu Lefort

Mathieu Lefort est maître de conférences en informatique à l'Université Lyon 1 au sein du laboratoire LIRIS. Durant son parcours, il a travaillé en lien avec différents domaines de recherche : neuroscience computationnelle, robotique développementale et apprentissage automatique. Ses travaux en intelligence artificielle portent sur l'étude d'agents apprenants autonomes bio-inspirés, plus particulièrement sur l'apprentissage développemental et la fusion de données.

Georges A. Legault

Ph. D. LL. L. *Doctorat honoris causa* de l'Université de Sudbury, Georges A. Legault est, depuis sa retraite, professeur associé à la Faculté de droit de l'Université de Sherbrooke. Membre du groupe de recherche InternE³LS au 3IT de l'Université de Sherbrooke, il travaille plus particulièrement sur la gouvernance du développement responsable de l'innovation technologique. L'équipe de recherche a développé, entre autres, un cadre d'analyse d'impact et d'acceptabilité éthique pour les nouvelles technologies, ainsi qu'une approche d'accompagnement du développement technologique par le biais des Usages responsables. Maints travaux cernent les difficultés à intégrer l'éthique, comme discipline philosophique, à l'évaluation des technologies dans nos sociétés démocratiques.

Margherita Merucci

Docteur en psychologie et enseignante-chercheuse à l'Université catholique de Lyon-France, les activités didactiques de Margherita Merucci concernent la psychologie du développement. Professionnellement, elle exerce une activité clinique auprès d'une institution qui accueille des enfants handicapés moteurs, polyhandicapés et présentant des handicaps rares. Dans ce contexte elle accompagne les enfants, les familles et les professionnels. Elle participe aussi à des activités de formation auprès d'institutions de formation en thérapie familiale systémique. Ses activités de recherche sont orientées vers les dynamiques familiales et la réorganisation du système après l'annonce du handicap, en particulier le devenir de la famille dans le temps, les problématiques d'attachement et le rôle du père et de la fratrie; les problématiques d'individuation/séparation de la personne handicapée (en particulier à travers la mise en place et l'animation de groupes de préadolescents); et les politiques inclusives (inclusion des enfants en situation de handicap dans l'école ordinaire, accompagnement et suivi des expériences).

Marjolaine Monot-Fouletier

Docteur en droit, titulaire de l'Habilitation à diriger des recherches (Université Paris-Descartes) depuis 2019, Marjolaine Monot-Fouletier est professeur à la Faculté de droit de l'Université catholique de Lyon. Elle est directrice pédagogique des cliniques juridiques de la Faculté de droit. À ce titre, elle organise en partenariat avec les juridictions lyonnaises, le barreau, une école d'ingénieurs et des industriels des procès fictifs menés par des étudiants sur des cas prospectifs en droit des nouvelles technologies. Elle est par

ailleurs membre de la commission de recrutement de l'Université catholique de Lyon et membre de l'European Law Institut (ELI). À partir de 2019, elle prend la direction du déploiement de cette clinique auprès de cinq universités partenaires. Elle concentre ses activités de recherche et d'enseignement dans les domaines du droit administratif, droit des libertés fondamentales et droit administratif des biens. Sa thèse relative à l'équité en droit public français est publiée à la LGDJ ; elle a publié une vingtaine d'articles scientifiques, ainsi qu'un manuel de droit administratif des biens. Elle est auteur au sein de l'Encyclopédie du Jurisclasseur.

Célia Vaz-Cerniglia

Célia Vaz-Cerniglia est directrice du Département de psychologie (Licence et Master « Psychologie clinique de l'enfant et de l'adolescent ») à l'Université catholique de Lyon. Enseignante et chercheuse en psychologie, sa réflexion se fait à partir de la pratique avec une approche développementale et psychodynamique. Elle est responsable du groupe de recherche GIRP (Groupe interdisciplinaire de recherche en psychologie) et présidente de l'association R-PSYCLY.