Vulnérabilité et intelligence artificielle
Oser la contre-utilité

Béatrice DE MONTERA

Groupe d'Epistémologie et Ethique des Sciences et Technologies (GEEST),
Pôle Biosciences, echnologies, Ethique, UR CONFLUENCE Sciences & Humanités,
Université Catholique de Lyon

Résumé : L'intelligence artificielle va devoir relever des défis pour l'instant inatteignables lorsqu'elle se trouve face aux personnes humaines les plus vulnérables. Comment pallier d'un point de vue éthique les risques qui surviennent en termes de manipulation de ces personnes ? Seule une éthique qui applique la contre-utilité et qui écarterait les plus vulnérables, lors du choix des volontaires pour tester les nouvelles solutions technologiques, pourra respecter la vulnérabilité humaine.

Mots-clés : vulnérabilité, individuation, utilité

L'intelligence artificielle (IA) est de plus en plus sollicitée dans des projets visant un appui aux personnes vulnérables mais aussi dans le cadre d'une médecine de précision. Mais de quel appui et de quelle précision parle-t-on ? On entend ainsi l'appui dans les manipulations domestiques quotidiennes ou pour rester connecté à la réalité environnante, et également la précision dans la mesure des indicateurs physiologiques ou dans l'administration du traitement, ou plus généralement, la précision dans l'approche de soin qui se dit alors « *comprehensive* » c'est-à-dire une approche de soin global ou complet. Mais de quel point de vue dit-on cela ? Les personnes concernées par ces soins, quelles sont-elles et s'expriment-elles sur les soins ou l'aide apportés ? Notre objectif sera d'abord d'identifier qui est vulnérable en priorité face à l'émergence de l'IA dans nos vies et quelle est la définition de la vulnérabilité que l'on peut retenir dans ce cadre. Nous envisagerons ensuite les défis posés à l'IA lorsqu'elle se trouve face à la vulnérabilité humaine et enfin, comment pallier d'un point de vue éthique les risques qui surviennent en termes de manipulation des personnes vulnérables et comment notamment oser la règle éthique de la contre-utilité.

On parle de « *comprehensive care* » ou soin complet lorsque l'on cumule les soins physiques et psychologiques ou émotionnels et cela par différentes approches comprenant aussi les relations des personnes à leur environnement social ou naturel [1]. On se réfère à ce soin global ou complet lors de situations de vulnérabilité complexes et durables pour lesquels une approche de soin classique, ciblée sur un organe, une voie métabolique ou un seul niveau du système biologique (physique *versus* psychique), paraît inopérante. Pour gérer alors la précision de traitement dans une approche de soin global, l'IA serait un atout majeur par exemple pour comparer les bases de données de cas déjà traités, suivre les bio-indicateurs, mesurer avec précision les effets, calculer des risques statistiques, n'oublier aucun traitement possible et coordonner les traitements à donner venant de différentes approches. Mais qui seraient les personnes vulnérables concernées par l'intervention de l'IA et de quelle vulnérabilité parle-t-on ? Pour répondre à cette question, il nous faut d'abord expliquer l'évolution de la notion de vulnérabilité. Dans les philosophies dites modernes, qui sont celles où le sujet humain est dominant, on assiste à l'oubli de la vulnérabilité humaine, puisque grâce à la puissance acquise par la technique, l'humain dépasse sa condition d'être vulnérable [2].

Au XX^{e} siècle et depuis un évènement marquant une puissance sans précédent, c'est-à-dire la capacité à détruire toute vie planétaire par la bombe atomique, on assiste au réveil de la vulnérabilité qui est le corollaire de notre puissance. Comme le dit Roland Schaer : *« De la puissance, nous sommes aussi bien les sujets que les objets, de la vulnérabilité, nous sommes aussi bien les objets que les sujets. Que ce soit les sciences qui nous le rappellent est une invitation faite aux philosophes de penser davantage avec la science »* [3].

Cette vulnérabilité qui nous revient tel un boomerang après chaque tentative de dépassement de notre condition humaine limitée, n'est autre que celle de chaque vivant en tant justement que vivant *perturbable*, *pénétrable*, c'est-à-dire ouvert à l'accident ou à la manipulation. Comme nous l'avons expliqué ailleurs en analysant le cas des biotechnologies, la vulnérabilité est liée à la fragilité du corps, mais pas parce que le corps est occasionnellement fragilisé, par la maladie par exemple, ou encore progressivement fragilisé par la vieillesse [4]. La vulnérabilité n'est pas la fragilité accidentelle, ni simplement la marque de la contingence [5] ; la vulnérabilité est la condition de toute vie qui doit être relationnelle pour être une vie et qui est donc en ce sens pénétrable y compris par l'intervention technique. Et à ce titre, toute entité biotechnique même alors que ses performances sont accrues, est vulnérable en ce qu'elle continue d'être en relation constructive avec l'environnement [6]. Et cette vulnérabilité consti-

1. Comme par exemple dans les articles suivants : HOPMAN P. *& alii*, Effectiveness of comprehensive care programs for patients with multiple chronic conditions or frailty : A systematic literature review, *Health Policy*, 2016 Jul ;120(7), p. 818-32 ; RAFFERTY J. *& alii*, Ensuring Comprehensive Care and Support for Transgender and Gender-Diverse Children and Adolescents, *Pediatrics*, 2018 Oct., 142(4), p. 818-32

2. SCHAER R., « Science, technique et responsabilité. Une lecture de l'œuvre de Hans Jonas », *Revue Découverte*, n°370, septembre-octobre 2010, p. 64-69.

3. *Idem*, p. 68-69

4. de MONTERA B. & MAGNIN T., « Vulnérabilité et responsabilité au temps de l'homme augmenté », in *La Vulnérabilité au prisme de la technologie*, THIEL M.-J. (dir.), Presses de l'Université de Strasbourg, 2019, sous presse

5. SCHAER R., *Répondre du vivant*, Paris, Le Pommier, 2013

6. Comme par exemple dans le cas des bactéries dont on a remplacé une base du code génétique et qui survivent dans des contions d'habitude létales, cf de MONTERA B. & MAGNIN T., *Op. cit.*

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

tutive est bien ce qui, chez l'humain, fait que l'humain peut être son propre horizon et se dépasser [7]. Il s'agit alors de se demander quel type de perturbation occasionne l'intervention technologique d'une IA cette fois, auprès d'une personne âgée ou malade, c'est-à-dire d'une personne plus vulnérable que la moyenne de la population. En d'autres termes, l'intervention d'une IA risque-t-elle d'augmenter la vulnérabilité des personnes concernées ?

Quelle définition du vulnérable est donc valable ici dans la relation à l'IA ? Nous venons de voir que tout vivant y compris pénétré par l'intervention technologique, et y compris dans une perspective d'augmentation de ses performances, reste vulnérable puisqu'il reste ouvert à la relation au monde environnant. Dans cette perspective, ne pas accroître la vulnérabilité d'une entité vivante, serait justement ne pas mettre en péril par une vulnérabilité supplémentaire nouvelle, la relation qu'entretient cette entité avec cet environnement qui est aussi symboliquement *son monde*, comme l'explique Merleau-Ponty [8]. Quelles seraient alors chez l'humain, les personnes chez qui la relation avec l'IA pourrait engendrer une vulnérabilité nouvelle et non constructive pour elles-mêmes ?

On pourrait penser en première analyse que ce serait les personnes les plus aisément manipulables par une IA parce que croyant avoir affaire à un autrui pensant et bienveillant. Cette croyance les rendrait alors plus pénétrables par la technologie et moins capables peut-être de continuer à s'adapter par eux-mêmes à leur environnement, qui ne serait alors plus tout-à-fait *leur monde. Leur monde* tissé par les relations symboliques que ces personnes ont construites avec ce qui est autre mais présent autour d'elles et qui participe de ce qu'elles deviennent. Ne pouvant plus tout-à-fait s'adapter par elles-mêmes, ces personnes ne seraient plus tout-à-fait sujets mais objets de leur vulnérabilité comme Roland Schaer le disait précédemment. Tous les humains pourraient être concernés par cette objectivation, cette instrumentalisation par la technique, mais *a fortiori* en priorité les plus aisément manipulables que seraient les enfants, les adultes fragilisés par une maladie durable et les personnes âgées. Or vers qui se tourne le marché de l'IA ? Lorsqu'il s'agit d'IA incarnées, comme dans le cas du robot Nao, on cible justement les enfants, ou bien avec les robots auxiliaires de vie, on cible les personnes âgées. Par ailleurs, pour des adultes célibataires prisonniers de leur vie professionnelle trop prenante, les hologrammes proposés par Gatebox Labs peuvent permettre un substitut de relation affective comme dans le cas d'Azuma (version 2018), la jeune fille humanoïde holographique à l'apparence d'un manga dirigée par une intelligence artificielle qui a une fonction d'assistante de maison et petite copine. Les applications de l'IA ciblent donc préférentiellement les personnes les plus vulnérables psychologiquement et affectivement parmi les humains. *CQFD.* L'efficacité de la technique d'IA ici se manifeste sur fond d'instrumentalisation d'une vulnérabilité plus grande.

7. MAGNIN T., *Penser l'humain au temps de l'homme augmenté : Face aux défis du transhumanisme*, Paris, Albin Michel, 2017

8. L'être vivant a un monde qui est *son monde*, on peut le dire d'une plante, d'un animal, d'un humain, voir MERLEAU-PONTY M., *La Nature*, Cours du Collège de France, 1964 ; la question est de savoir si cela reste vrai lors d'une vie avec une IA à nos côtés ou si la symbolique du monde est modifiée par l'IA.

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

Que dit l'éthique à partir d'un tel constat ? Des études de cas en éthique sur la manipulation des plus vulnérables, il y en a beaucoup. Prenons l'approche de bioéthique médicale que décrit Habermas. Selon lui, on doit respecter le principe de symétrie communicationnelle avec tout patient et si son état impose une intervention technique, il faut toujours pouvoir supposer l'autonomie de droit dans les relations interpersonnelles et discuter avec le patient pour avoir son avis et récolter son accord ou son refus. Si la communication n'est pas possible, il faut alors refuser toute intervention sans le consentement du patient [9]. Or dans le cas de nos personnes vulnérables psychologiquement ou affectivement, qui sont manipulables aisément, cela signifie que l'on peut les soumettre à une volonté extérieure et que donc, si consentement il y a, il ne serait plus valable, plus moralement légitime. Ainsi on peut prendre ce cas d'une personne âgée en état de souffrance aigüe, qui se verrait administrer son traitement par un robot programmé pour donner un traitement de précision, dosé en fonction de l'état physiologique du patient. Le robot aurait été programmé bien sûr préalablement par le médecin qui suit le patient. Mais contrairement au médecin, l'IA du robot ne pourra recueillir aucune plainte, aucune décision du patient, aucun refus de traitement. Or la capacité à refuser est la marque de l'autonomie de l'humain et de sa dignité. C'est dans la capacité à rejeter la soumission à une volonté extérieure, que l'on peut distinguer, comme le rappelle Gilbert Hottois, le cas de l'éducation de l'enfant et celui de l'instrumentalisation du patient à l'œuvre parfois en médecine : l'un pouvant toujours dire non, l'autre n'y arrivant parfois plus [10]. Quel gage donner à une IA qui représente de manière non vivante et non humaine la décision d'un médecin alors que déjà le patient a du mal à dire non à un humain lorsqu'il est médecin ?

Quel est donc le problème de fond posé à l'intelligence artificielle lorsqu'elle est face à la vulnérabilité humaine ? La réponse pourtant est omniprésente dans les media : les atouts de l'IA sont nombreux. Sa puissance de calcul et de comparaison de données permet d'avoir en conséquence, une puissance analogico-déductive sans précédent. Mais ce que cela apporte en plus, c'est que l'IA permet aussi d'accéder, grâce à la convergence des approches et les analyses à grande échelle, à une étude des situations les plus complexes comme par exemple, les personnes multi-vulnérables comme les enfants ou adolescents LGBTQ [11] stigmatisés pour leur différence [12], ou dans certains pays, les femmes ou les personnes handicapées dont les problèmes de santé sont sous-diagnostiqués [13]. Ainsi l'IA permet donc de nouvelles fonctions, comme la détection de personnes vulnérables dont la vulnérabilité est amplifiée car elles sont marginalisées et en incapacité de communiquer. Cela permet donc en conséquence, de détecter de nouveaux critères de vulnérabilité, de les mettre en lien avec d'autres et d'avoir une approche multifactorielle de la causalité et de pouvoir distinguer ainsi les causes

9. HABERMAS J., *L'avenir de la nature humaine. Vers un eugénisme libéral ?,* BOUCHINDHOMME C. (*trad.*), Paris, Gallimard, 2002

10. HOTTOIS G., « Quel cadre temporel pour penser les générations futures », *in* Vers la Fin de l'Homme ?, CHRISTIAN H. et ROZENBERG J. J. (*dir.*), Bruxelles, de Boeck, 2006, p. 79.

11. Lesbian, Gay, Bisexual, Transgender, or Questioning

12. RAFFERTY J. *& alii*, *Op. cit.*

13. BASU S. & KING AC., "Disability and chronic disease among older adults in India : detecting vulnerable populations through the WHO SAGE Study", *Am J Epidemiol*, 2013, 178 (11), p. 1620-1628

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

des diverses vulnérabilités [14]. Mais l'IA a un désavantage majeur, elle fait peur à une majorité de personnes [15]. Probablement en raison du fait que son développement et les changements de vie associés sont trop rapides pour que l'acculturation ait pu se faire ou bien par peur que la machine soit plus forte que nous en confondant performance pour une tâche précisément décrite et performance tout court [16]. Certains répondront alors : sauf si on grandit avec l'IA ou si on la prend pour un humain ? Mais ne retombe-t-on pas dans le cercle précédent où l'IA est plus performante quand elle est utilisée chez les personnes les plus aisément manipulables : les enfants et les personnes fragilisées psychologiquement et émotionnellement qui, isolées, s'attacheront alors d'autant plus à cette relation avec une IA ? Sur fond de domestication de l'humain par l'IA donc pourrait-on arguer. Les personnes concernées ont-elles donné leur consentement à cela ?

Filons l'exemple précédent d'une IA venant en appui d'une médecine de précision chez une personne âgée, à domicile, touchée par une maladie, mais qui serait chronique cette fois. Cette personne âgée et malade bénéficie d'une aide pour les tâches quotidiennes et pour l'administration de son traitement par un robot auxiliaire de vie. Le robot programmé par le médecin donne la nourriture et les médicaments au patient. Mais qu'arrive-t-il si le patient redemande un 2^{e} café ou un 2^{e} verre de vin ? Le robot ne le lui donne pas, sans discussion mais en répétant simplement l'ordonnance médicale qui fait loi. Serait-ce ce que ferait le médecin ? Sans doute, pour commencer. Et puis viendrait finalement le moment où le médecin appliquerait le principe de symétrie communicationnelle, où il rentrerait humainement en relation avec le patient, cet autre être humain, cet *alter ego*. Se jouerait alors la dialectique hégélienne de soumission puis de reprise de pouvoir : le patient est soumis au médecin pour ce qui concerne sa santé, mais il peut prendre le risque de dire non et se faisant, ne soumet plus sa vie, son autonomie, sa dignité. Et le médecin lui, peut comprendre. Comprendre cette vulnérabilité néanmoins autonome qui prend le risque de la liberté. Et accueillir un refus de suivre le traitement. Et dans certains cas, il laissera prendre le 2^{e} café, le 2^{e} verre de vin. Ainsi, dans tous les cas, il est des choses que l'IA ne fera pas et dont il est bon d'évaluer l'importance. Car ce dont a besoin la personne particulièrement vulnérable n'est-ce pas qu'on l'aide à continuer son chemin d'individuation grâce à la relation au monde vivant ?

Peut-on réellement concevoir aider cette personne en lui offrant une relation avec un robot aussi efficace et intelligent soit-il ? En d'autres termes, la relation à la technique a-t-elle le pouvoir de nous individuer comme les relations au monde vivant nous permettent de le faire ? Pour répondre, il nous faut expliquer la notion d'individuation qui a été particulièrement développée par Gilbert Simondon [17]. Dans cette perspective, l'individu est toujours déjà en relation avec son milieu avec lequel il a une relation constructive. L'individuation est alors ce processus de synthèse sans cesse réactualisé

14. grâce par exemple à la WHO SAGE Study (World Health Organization Study on Global Ageing and Adult Health) et le calcul du score WHOSADi (World Health Organization Disability Assessment Schedule 2.0, inverted), cf *Idem*

15. HODSON R., « Digital revolution », *Nature*, 2018, 563(7733), p. S131 ; HECHT J., « Meeting people's expectations », *Nature*, 2018, Vol 563(7733), p. S141

16. HECHT J., *Op. cit.*

17. SIMONDON G., *L'individuation à la lumière des notions de forme et d'information,* [thèse 1958], Grenoble, Million, 2013

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

qui passe par la destruction de l'état antérieur et qui puise dans le milieu associé et qui fait que l'individu se construit toute sa vie [18]. En tant que vivant, l'humain est un processus d'individuation physico-biologique mais son esprit connaît également un processus d'individuation. Or l'équilibre du vivant est un équilibre « métastable » [19], et il faut donc toujours laisser ouvertes les possibilités d'individuation par cette prise de risque même de la métastabilité. L'animal ou l'humain s'individualise à partir de cette genèse que représente le processus d'individuation propre au vivant basé sur l'équilibre métastable qui lui permet de s'adapter. Or que va faire la technologie face au processus d'individuation et d'individualisation du vivant humain ? Pourra-t-elle le suivre, l'alimenter ou bien au contraire va-t-elle trop le contrôler, le stabiliser et le standardiser ? La tendance actuelle au naturalisme de la technologie qui s'insère dans le cours de la biosphère, peut laisser oublier les risques de désindividualisation dont parle Ruud ter Meulen :

[...] there is an inherent tendency in technology to dominate and control our lives and to fragmentise and de-humanise society [...]. Technology can indeed take over human attributes and may promote standardization, destroying what is unique in each individual. An enhanced human being then is at risk of being so closely controlled by technological devices that this being becomes stripped of any individuality [20].

Or face au risque de désindividualisation et à celui de la perte d'autonomie et de dignité, il existe une priorité éthique : prendre en compte la vulnérabilité accrue des plus vulnérables. En d'autres termes, les personnes en situation de vulnérabilité accrue devraient être les dernières sur lesquelles on teste des technologies nouvelles et vulnérabilisantes pour les individus qu'elles sont et qui continuent tant bien que mal leur individuation.

Face aux risques de désindividualisation et de perte d'autonomie et de dignité quelles sont les solutions que l'on peut imaginer ? La problématique est ici de se demander ce que l'on peut enlever dans l'action contrôlante des machines vis-à-vis de l'humain vulnérable qui est en face. Trois prescriptions nous semblent possibles à envisager :

1, l'interdiction de tout ce qui enlève la liberté et la surprise qu'une personne peut être pour elle-même

2, « enlever le mal des machines » [21] en instaurant une règle de priorité éthique possible à rentrer dans un algorithme

3, Faire le pari de l'adaptabilité des « agents autonomes » [22] que deviennent progressivement les IA

18. *Idem*, p. 204

19. *Ibidem*

20. ter MEULEN R., "The Moral Ambiguity of Human Enhancement", in Inquiring into Human Enhancement. Interdisciplinary and international perspectives, BATEMAN S., et *alii* (éds.), Hampshire, England, Palgrave Macmillan, 2015, p. 97.

21. GRINBAUM A., *Les robots et le mal*, Desclée de Brouwer, 2019

22. XUE, J. et *alii*, "Causality Reconstruction by an Autonomous Agent", *BICA 2018*, 2018 ; GEORGEON O. & GUILLERMIN M., "Mastering the Laws of Feedback Contingencies Is Essential to Constructivist Artificial Agents", *275 OPC*, 2018

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

La première prescription est celle qu'a faite Hans Jonas dans son plaidoyer sur les risques d'eugénisme liés au clonage chez l'homme [23] :

La prescription éthique qui entre ici dans la gamme élargie de nos pouvoirs est la suivante : *respecter le droit de chaque être de trouver sa propre voie et que cela reste une découverte pour lui.* [24]

Son texte de 1974 anticipe de manière troublante les enjeux éthiques des technologies qui peuvent s'immiscer justement dans le processus d'individuation puis d'individualisation comme par exemple celle du clonage. Dans le cas du clone humain, on choisit le génotype (celui de l'original) et on choisit l'environnement dans lequel il vivra, et ainsi on oriente son individuation-mêmede telle sorte qu'il n'est plus unique, ni autonome quant à ce qu'il devient. L'objectif serait donc d'interdire à l'IA face à des personnes particulièrement vulnérables d'agir d'une façon qui les empêche d'avoir les comportements propres à ce qu'ils sont en tant qu'individus humains uniques et en cours d'individuation.

La seconde prescription est fondée sur l'idée que le mal, s'il n'est pas dû à la volonté humaine, vient alors de la machine [25]. Il convient alors d'enlever le mal de la machine en lui apprenant par exemple des règles de priorité éthique où l'IA serait moins contrôlante avec les personnes les plus vulnérables qui ont justement le plus de mal à s'affranchir du joug provoqué par l'usage de la machine. Le principe de justice distributive par exemple qui tient compte de l'état de chacun et qui dose l'action en fonction de la situation du patient peut être introduit dans un algorithme comme l'atteste la recherche actuelle [26].

Enfin la troisième prescription serait celle qui ferait le pari qu'un jour une IA pourrait ne pas devenir désindividualisante par le fait qu'elle serait elle-même non seulement un agent autonome déjà comme le défendent certains chercheurs du domaine, mais aussi un jour capable de créer des relations constructives pour l'humain sans pour autant le déshumaniser. Les recherches sur l'« Epigenetic Robotics » qui datent de 2001 [27] rappellent le lien entre épigénétique et individuation que nous avons décrit ailleurs [28]. L'idée d'une adaptation possible des machines en fonction de leur environnement rejoint l'idée de Simondon selon laquelle les objets techniques s'individualisent également en lien avec leur milieu associé, qui est humain [29]. Ayant un processus d'individualisation eux aussi, on pourrait imaginer que ces objets techniques-là ne contreviendraient pas à l'individualisation des humains eux-mêmes ou du moins ne provoqueraient pas de standardisation. Mais une critique survient alors qui est celle de l'anthropocentrisme et de l'auto-référencement. Les IA seraient individualisées et

23. JONAS J., Ch. VII Biotechnologies. Une vision anticipatrice, de MONTERA B. (trad.), *Essais Philosophiques. Du credo ancien à l'homme technologique*, Paris, Vrin, 2013 [1974], p. 191-222

24. *Idem*, p. 217

25. GRINBAUM A., *op. cit.*

26. RAJKOMAR A. & *alii*, *op. cit.*

27. BALKENIUS C., *et alii.* (éds.), *Proceedings of the First International Workshop on Epigenetic Robotics : Modeling Cognitive Development in Robotic Systems*, Lund, Lund University, 2001 ; PRINCE C. G. & DEMIRIS Y., "Introduction to the Special Issue on Epigenetic Robotics", *International Society for Adaptive Behavior*, 2003, Vol 11(2), p. 75–77

28. Voir de MONTERA B., « L'hérédité épigénétique : un changement de paradigme ? », *Implications philosophiques*, ISSN 2105-0864, 2014, p. 27-49

29. SIMONDON G., Du mode d'existence des objets techniques [1958], Paris, Aubier, 2012

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

agiraient de façon autonome *pour nous* en tant que produits de notre science et de notre volonté [30]. Ce faisant nous ne ferions comme le disait Nietzsche que nous impressionner nous-mêmes et les IA seraient en fait incapables d'êtres individualisées et autonomes *pour elles*. Comment seraient-elles alors capables d'avoir le souci de l'autre ou de parler la langue du corps comme nous le faisons, nous humains,en réponse à la vulnérabilité humaine ?

Se soucier de l'autre et « parler le corps [31] » constituent probablement les défis inatteignables pour l'IA puisqu'ils ont trait à l'émotion vécue et constituent le moment affectif de la formation de la volonté morale. Selon Hans Jonas en effet, l'origine de la sollicitude, n'est pas l'amour, la pitié, mais la peur. Pas n'importe quelle peur en l'occurrence, mais celle qui accompagne cette question : « *[. . .] que lui arrivera-t-il, si moi je ne m'occupe pas de lui ?* [32] » Cette peur devant la vulnérabilité humaine qui est une culpabilité par anticipation sur le futur est l'émergence de notre responsabilité et elle a besoin d'être vécue pour exister. Or le vécu c'est aussi les intentions du corps qui ne sont pas sans effet et sans sens :

Les intentions corporelles, pour inconscientes qu'elles soient, n'en sont pas moins actives, car la perception met en œuvre dans le cerveau des scénarios de décision, comme le précise Alain Berthoz, avant que l'action corporelle n'en réalise un seul. Le corps incarne du sens avant de pouvoir l'exprimer dans une structure linguistique susceptible d'être reconnue par la communauté lors de l'interaction proxémique [33].

Face à un objet technique comme l'IA, non suffisamment intégré dans la vie humaine en ce qu'elle est corps et sollicitude envers autrui, il va falloir donc protéger les personnes les plus vulnérables de deux risques majeurs : que l'on ne communique plus avec elles et qu'on ne les fasse plus communiquer.

Après cette analyse, quel périmètre protecteur pouvons-nous proposer autour des personnes les plus vulnérables à l'IA ? L'objectif éthique ici est de protéger les personnes les plus vulnérables du risque de l'intuition de l'utilité de l'IA et de son corollaire quiest la manipulation accrue des vulnérables par la règle de la contre-utilité. Nous sommes dans une époque pragmatiste où l'on n'hésite pas à instrumentaliser les corps au seuil de la mort dans ce qui est une manière d'étendre le règne de la matière et de l'utilité [34]. Dans cette lignée, en voulant proposer l'appui de l'IA aux personnes les plus vulnérables que sont les enfants, les adultes malades et fragiles psychologiquement et émotionnellement et les personnes âgées, on suit en réalité des objectifs utilitaristes qui sont la diminution de la souffrance et un meilleur contrôle effectif de la contingence (accidents) dans une logique de réduire les coûts globaux et de maximiser ce qui sera utile au plus grand nombre. La question de la réelle amélioration que l'IA apportera aux vies, aux relations personnelles, à la société n'est pas incluse dans le paquet de

30. NIETZSCHE F., *Vérité et mensonge au sens extra-moral,* Arles, Actes Sud, Collection Babel, 2002

31. ANDRIEU B., ch. II « Parler le corps », *La langue du corps vivant. Emersiologie 2*, Paris, Vrin, 2018, p. 167-207

32. JONAS J., *Le principe Responsabilité. Une éthique pour la civilisation technologique*, GRIEISH J. (trad.), Paris, Cerf, 1990 [1979], p. 421-422

33. *Idem*, p. 178-179

34. Voir BOSSI L., « Hans Jonas et les critères de la mort à l'ère des greffes d'organe », in *L'éthique de la vie chez Hans Jonas*, LARRERE C. et POMMIER E. (dir.), Paris, Publications de la Sorbonne, 2013, p. 211

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

la technologie qui nous est livré [35]. Il nous faut donc apprendre à discerner les fausses bonnes idées et les partager afin de guider les personnes concernées, c'est-à-dire nous tous à terme. Le risque de l'intuition de l'utilité rappelons-nous est le suivant : la technologie est plus utile si on l'applique à des personnes chez qui elle a le plus de chance de marcher, c'est-à-dire chez des personnes qui en ont un besoin assez crucial et qui y croient. Comme nous l'avons vu, le corollaire de ce risque est en fait, sous couvert du recrutement des meilleures cibles pour la technologie, une manipulation des personnes les plus vulnérables à cette technologie. Le besoin et la croyance en l'utilité de la technologie suffiront à engendrer des comportements de censure, d'auto-censure chez les personnes utilisatrices de la technologie comme cela est décrit dans les cas d'emprise par l'autorité d'un mentor. Dans ce cadre, une bonne intention comme la médecine de précision appliquée à l'approche de soin complet, peut s'avérer devenir un contrôle complet et déshumanisant. Nous ferions bien alors de choisir le sens de la « contre-utilité » que décrit Hans Jonas à propos du recrutement de volontaires ou de malades pour des essais cliniques [36] :

[...] plus le savoir, la motivation et la liberté de décision sont réduits – et malheureusement cela signifie une plus grande disponibilité en termes de nombre et de manipulations possibles-, plus le réservoir doit être utilisé avec modération et réticence, et partant, la justification qui fait le contrepoids doit être encore plus forte. Notons que nous avons là le contraire de la norme sociale, l'inverse de l'ordre par « disponibilité et utilisation [37] *» : [...].*

L'inverse de l'intuition de l'utilité est donc la règle de la contre-utilité : ne pas utiliser les plus vulnérables pour tester l'IA mais plutôt ceux qui en ont le moins besoin, qui ne la prendraient pas pour un humain et qui donc en général sont les plus éduqués et en bonne santé, donc ceux qui ont le plus de valeur pour la société. Faire cela c'est viser en faitune utilité « plus haute » comme le dit Jonas, une utilité qui « *garantit une âme à la communauté* » [38]. Il s'agit en effet d'une forme de justice qui dit que l'on ne doit pas mettre en péril les plus vulnérables et laisser aux plus forts la prise de risque. Forts de cela, nous pouvons à présent décrire les rapports entre l'ordre décroissant de contre-utilité et d'acceptabilité et l'ordre croissant de la vulnérabilité (et la manipulation des plus vulnérables qui va de pair) et de la responsabilité (figure 1).

En conclusion, une éthique de la vraie utilité c'est une éthique qui applique la contre-utilité sociale et qui est à l'inverse des intérêts pragmatiques à court terme et de l'intuition sociale dominante où l'on manipulera qui on peut aisément manipuler. Dans le cas de l'IA appliquée aux personnes les plus vulnérables, on est dans le cas d'une expérimentation à grande échelle, où sans le dire on a recruté des volontaires sains ou des malades parmi les plus vulnérables car la technologie est encore dans une phase où elle doit passer le test de la réalité, et donc de l'acceptabilité sociale après avoir passé le test du laboratoire. Ne pas accepter la règle de la contre-utilité nous rendrait collectivement tous hautement responsables car nous saurions que nous

35. ter Meulen, *op. cit.*, p. 98

36. JONAS J., Ch. V Réflexions philosophiques sur l'expérimentation humaine, RAVEZ L. (trad.) *Essais Philosophiques. Du credo ancien à l'homme technologique*, Paris, Vrin, 2013, [1965-1973], p. 149-179

37. *Idem*, p. 169

38. *Ibidem*

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

aurions répondu à la question : « [...] que lui arrivera-t-il, si moi je ne m'occupe pas de lui ? » par : « une machine s'occupera de lui », or nous avons peur de ce que cela veut dire. Le nouvel impératif catégorique de notre époque pourrait être alors : arrête toute expérimentation technologique dès lors que l'individualisation ne peut plus se faire, dès lors que la personne ne peut plus tracer son chemin d'individuation et qu'elle n'a pas encore pu intégrer l'objet technique dans ce processus unique qu'elle préside.

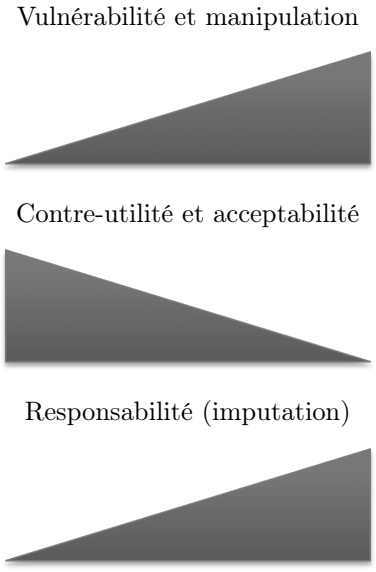


Figure 1 : Le sens de la contre-utilité

Références

ANDRIEU B., *La langue du corps vivant. Emersiologie 2*, ch. II « Parler le corps », Paris, Vrin, 2018, p. 167-207 ;

BALKENIUS C., *et alii.* (éds.), *Proceedings of the First International Workshop on Epigenetic Robotics : Modeling Cognitive Development in Robotic Systems*, Lund, Lund University, 2001. (*Cognitive Studies series*, 2001, Vol 85) ; http://www.lucs.lu.se/Abstracts/LUCS_Studies/LUCS85.html ;

BASU S. & KING AC., "Disability and chronic disease among older adults in India : detecting vulnerable populations through the WHO SAGE Study", *Am J Epidemiol*, 2013, 178(11), p. 1620-8 ;

BOSSI L., « Hans Jonas et les critères de la mort à l'ère des greffes d'organe », in *L'éthique de la vie chez Hans Jonas*, LARRERE C. et POMMIER E. (dir.), Paris, Publications de la Sorbonne, 2013, p. 199-216 ;

de MONTERA B. & MAGNIN T., « Vulnérabilité et responsabilité au temps de l'homme augmenté », in *La Vulnérabilité au prisme de la technologie*, THIEL M.-J. (dir.), Presses de l'Université de Strasbourg, 2019, sous presse ;

GEORGEON O. & GUILLERMIN M., "Mastering the Laws of Feedback Contingencies Is Essential to Constructivist Artificial Agents", *275 OPC*, 2018 ; XUE, J. *et alii*, "Causality Reconstruction by an Autonomous Agent", *BICA 2018*, 2018 ;

GRINBAUM A., *Les robots et le mal*, Desclée de Brouwer, 2019

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874

HOPMAN P. *et alii*, Effectiveness of comprehensive care programs for patients with multiple chronic conditions or frailty : A systematic literature review, *Health Policy*, 2016, 120(7), p. 818-32 ;

HOTTOIS G., « Quel cadre temporel pour penser les générations futures », in *Vers la Fin de l'Homme ?*, CHRISTIAN H. et ROZENBERG J. J. (dir.), Bruxelles, de Boeck, 2006, p. 67-87 ; JONAS J., *Essais Philosophiques. Du credo ancien à l'homme technologique*, Ch. V RAVEZ L. (trad.) et Ch.VII de MONTERA B. (trad.), Paris, Vrin, 2013 [1965-1974] ;

JONAS J., *Le principe Responsabilité. Une éthique pour la civilisation technologique*, GRIEISH J. (trad.), Paris, Cerf, 1990 [1979] ;

MAGNIN T., *Penser l'humain au temps de l'homme augmenté : Face aux défis du transhumanisme*, Paris, Albin Michel, 2017 ;

MERLEAU-PONTY M., *La Nature*, Cours du Collège de France, 1964 ;

PRINCE C. G. & DEMIRIS Y., "Introduction to the Special Issue on Epigenetic Robotics", *International Society for Adaptive Behavior*, 2003, Vol 11(2), p. 75–77 ;

RAFFERTY J. *et alii*, Ensuring Comprehensive Care and Support for Transgender and Gender-Diverse Children and Adolescents, *Pediatrics*, 2018,142(4), p. 818-32

RAJKOMAR A. *et al.*, "Ensuring Fairness in Machine Learning to Advance Health Equity", *Ann Intern Med.*, 2018, 169(12), p. 866–872 ;

SCHAER R., « *Science, technique et responsabilité.* Une lecture de l'œuvre de Hans Jonas », *Revue Découverte*, n° 370, septembre-octobre 2010, p. 64-69 ;

SCHAER R., *Répondre du vivant*, Paris, Le Pommier, 2013 ;

SIMONDON G., *L'individuation à la lumière des notions de forme et d'information,* [thèse 1958], Grenoble, Million, 2013 ;

SIMONDON G., *Du mode d'existence des objets techniques* [1958], Paris, Aubier, 2012 ; ter MEULEN R., "The Moral Ambiguity of Human Enhancement", in *Inquiring into Human Enhancement. Interdisciplinary and international perspectives*, BATEMAN S., *et alii* (éds.), Hampshire, England, Palgrave Macmillan, 2015, p. 86-99.

Published in : PIATTI, M.-C. ; GUILLERMIN, M. (dir.) (2020), *Intelligence(s) artificielle(s) et Vulnérabilité(s) : kaléidoscope. Les contreforts de l'éthique et du droit*, Editions des archives contemporaines, ISBN : 9782813003874, doi : https ://doi.org/10.17184/eac.9782813003874