

## L'algorithme et les droits de l'Homme\*

Par Juan G. Corvalán

### 1. Introduction

L'humanité se trouve à l'aube d'une ère où les *robots*, les *algorithmes intelligents*, les *androïdes* et d'autres formes d'intelligence artificielle, de plus en plus sophistiquées, semblent être sur le point de déclencher une nouvelle révolution industrielle qui touchera probablement toutes les couches de la société. Ces affirmations ne sont pas tirées d'un film de science-fiction. Ce sont les termes mêmes du Parlement Européen dans ses recommandations, approuvées début 2017, concernant les règles de droit civil sur la robotique<sup>1</sup>.

Vers la fin de cette deuxième décennie du XXI<sup>e</sup> siècle, nous avons commencé à nous poser les questions suivantes: qui sont les responsables des conséquences du comportement des machines intelligentes? Comment assurer l'autodétermination humaine dans l'ère des algorithmes intelligents? Est-ce possible et, le cas échéant, comment "programmer" l'intelligence artificielle pour qu'elle puisse avoir une approche juridique, éthique et morale? Comment faire pour éviter que l'intelligence artificielle intensifie les inégalités entre les personnes? Comment réussir à ce que les machines intelligentes soient compatibles avec les droits de l'homme? Les réponses à ces questions demandent un effort important visant à repenser et à mettre en place des outils innovants pour aborder les enjeux de ce qu'on appelle la "quatrième révolution industrielle"<sup>2</sup> et à laquelle nous sommes confrontés. Nous devons explorer des approches durables face à des scénarios disruptifs, complexes, incertains qui présupposent essentiellement des approches transversales, interdisciplinaires, multipolaires, dynamiques, flexibles et intégrées.

Nous nous trouvons pour l'essentiel face à un défi sans précédent que nous pouvons résumer par les deux questions suivantes: Comment faire pour que cette intelligence non humaine soit bénéfique et non nuisible aux êtres humains? D'un point de vue strictement juridique, comment faire pour que le développement de l'intelligence artificielle soit compatible avec les droits de l'homme et avec l'État constitutionnel? Nous essaierons dans cet exposé succinct de mettre en Avant une voie possible à long terme et qui reste à explorer. Le parcours séquentiel de cette proposition peut être synthétisé comme suit: L'intelligence humaine → l'intelligence artificielle → les

---

\* Publié dans Conseil D'État, 2017. [Version espagnole](#). [Bibliographie recommandée](#).

<sup>1</sup> 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), Introduction, points A et B.

<sup>2</sup> Autant le Forum Économique Mondial que l'Organisation Internationale du Travail, soulignent que le monde est en train de traverser une quatrième révolution industrielle. Voir L'initiative du centenaire sur l'avenir du travail. Note d'Information, 2015, p. 2, et, Le futur des emplois. Compétences et stratégie de la main d'oeuvre pour la quatrième révolution industrielle. Forum Économique Mondial, 2016, p. 1, Global Challenge Insight Report, disponible à [www.weforum.org/reports/the-future-of-jobs](http://www.weforum.org/reports/the-future-of-jobs); sur les aspects conceptuels de cette dénomination (quatrième révolution industrielle) voir: Schwab, Klaus, *La quatrième révolution industrielle*, Barcelone, Dèbat, 2016.

problèmes/enjeux → une approche centrée sur les droits de l'homme → les principes juridiques de l'IA.

## 2. L'intelligence humaine, l'intelligence artificielle et les algorithmes intelligents

L'intelligence humaine est liée à un nombre de capacités<sup>3</sup> ou qualités<sup>4</sup> cognitives relativement indépendantes qui sont souvent décrites comme des "profils d'intelligence" ou "intelligences multiples"<sup>5</sup>. Elles sont l'intelligence sociale, l'intelligence linguistique (ou musicale), l'intelligence logique-mathématique, l'intelligence interpersonnelle et intra-personnelle ou émotionnelle<sup>6</sup>, l'intelligence fluide<sup>7</sup>, entre autres<sup>8</sup>. Parmi les divers aspects du concept "intelligence", l'élément commun est la *capacité de traiter des informations pour résoudre des problèmes*<sup>9</sup>. Notre cerveau contrôle en substance cette capacité à traiter les informations<sup>10</sup> provenant de l'environnement et de notre propre corps<sup>11</sup>, et qui sont utilisées pour évaluer et choisir de futures lignes de conduite. Et c'est là qu'entrent en scène le processus de prise de décisions<sup>12</sup> et

<sup>3</sup> Voir Gardner, Howard, *La inteligencia reformulada [L'intelligence reformulée]*, Madrid, Paidós, 2010, p. 115.

<sup>4</sup> Voir Kayser, Daniel, *Diccionario de ciencias cognitivas [Dictionnaire des sciences cognitives]*, Bs. As., Amorrortu, 2003, p. XXIII.

<sup>5</sup> Voir Gardner, Howard, *Las cinco mentes del futuro [Les cinq esprits du futur]*, Bs. As., Paidós, 2013, p. 17; *Estructuras de la mente: la teoría de las inteligencias múltiples [Les structures de l'esprit: la théorie des intelligences multiples]*, Mexique, Fondo de Cultura Económica [Fonds de Culture Économique], 1987.

<sup>6</sup> Voir Bachrach, Estanislao, *En cambio [En échange]*, 5<sup>e</sup> ed., Bs. As., Sudamericana, 2015, p. 395 et 396.

<sup>7</sup> Voir Manes, Facundo - Niro, Mateo, *Usar el cerebro [Utiliser le cerveau]*, Bs. As., Planeta, 2017, p. 115.

<sup>8</sup> Ampliar en Gardner, Howard, *La inteligencia reformulada [L'intelligence reformulée]*, p. 62 et suiv.

<sup>9</sup> Howard Gardner conçoit l'intelligence comme un "potentiel bio-psychologique pour traiter les informations susceptibles de s'activer dans un cadre culturel pour **résoudre des problèmes ou créer des produits qui ont une valeur pour une culture**" (mis en gras par moi). Et ensuite le même auteur affirme que: "Les intelligences sont des potentiels neuraux qui s'activent ou non en fonction des valeurs d'une culture donnée, des opportunités disponibles dans cette culture et des décisions prises par chaque personne et/ou sa famille, ses enseignants et d'autres personnes" (*La inteligencia reformulada [L'intelligence reformulée]*, p. 52 et 53). Le directeur de l'ingénierie chez Google, Ray Kurzweil, définit l'intelligence comme la "capacité à résoudre des problèmes à partir de ressources finies, y compris des limitations de temps", *La singularité est proche*, Berlin, Lola Books, 2012, p. 339). Manes et Niro font allusion à l'intelligence fluide comme l'ensemble des ressources dont dispose un individu pour s'adapter au milieu ou la "capacité à résoudre des problèmes nouveaux en découvrant les relations qui existent entre les choses et indépendamment des connaissances acquises tout au long de la vie" (Manes - Niro, *Usar el cerebro [Utiliser le cerveau]*, p. 114 et 115).

<sup>10</sup> Benítez, Raúl - Escudero, Gerard - Kanaan, Samir - Masip Rodó, David, *Inteligencia artificial avanzada [L'intelligence artificielle avancée]*, Barcelone, UOC, 2013, p. 14.

<sup>11</sup> D'un point de vue biologique, l'ADN est un porteur essentiel de l'information génétique. Gerard, Michael - Gerald, Gloria E., *El libro de la biología [Le livre de la biologie]*, Madrid, Ilus Books, 2015, p. 354.

<sup>12</sup> À l'égard des questions concernant le processus de prise de décisions chez les personnes, voir Kahneman, Daniel, *Pensar rápido, pensar despacio [Penser vite, penser lentement]*, Barcelone, Debate, 2012, p. 569 à 590.

l'évaluation qui consiste à *sélectionner, découper et organiser les informations disponibles*<sup>13</sup>.

L'intelligence humaine a conduit au développement de multiples innovations technologiques de nature diverse. Nous aborderons ici celles qui concernent le traitement des informations pour résoudre des problèmes et prendre des décisions à partir de machines ou de ce qu'on appelle les algorithmes intelligents<sup>14</sup>. L'intelligence artificielle (ci-après l'IA) repose sur des algorithmes intelligents<sup>15</sup> ou des algorithmes d'apprentissage qui sont utilisés, parmi bien d'autres objectifs, pour identifier des tendances économiques ou des recommandations personnalisées<sup>16</sup>. Un algorithme peut être défini comme un ensemble précis d'instructions<sup>17</sup> ou de règles<sup>18</sup>, ou comme une suite méthodique de pas qui peut être utilisée pour faire des calculs, résoudre des problèmes et prendre des décisions<sup>19</sup>.

<sup>13</sup> Pour plus de précisions, voir Manes, Facundo - Niro, Mateo, *El cerebro argentino [Le cerveau argentin]*, Bs. As., Planeta, 2016, p. 269, 270, 274, 275 à 301, et les mêmes auteurs à *Usar el cerebro [Utiliser le cerveau]*, p. 130.

<sup>14</sup> Il est important de préciser deux questions qu'on perçoit souvent intuitivement lorsqu'on analyse jusqu'à quel point les *robots* ou les ordinateurs peuvent être "intelligents". En premier lieu, on discrédite l'intelligence artificielle parce qu'on soutient qu'il est impossible de reproduire le cerveau humain dans des machines, étant donné sa complexité et l'absence de données fiables permettant de déterminer avec certitude son fonctionnement intégral. Bien que ce dernier élément soit exact, il est vrai que cette objection ignore le fait que l'intelligence artificielle ne doit pas ressembler le cerveau pour réaliser avec succès des activités qui ne sauraient être attribuées qu'à l'intelligence humaine. En outre, il importe prendre en compte que le développement de l'IA ne consiste pas -au moins exclusivement- à imiter ou copier le cerveau humain. L'IA utilise différentes méthodes pour traiter les informations et résoudre des problèmes ou prendre des décisions, de la même manière -permettez-moi cette analogie- que les ingénieurs en aviation n'ont ni imité ni copié la méthode et/ou les techniques d'apprentissage des oiseaux pour construire les avions modernes. En deuxième lieu, de nombreux progrès technologiques marqués par la présence de l'IA deviennent naturels. Si une machine intelligente commence à réaliser certaines activités où elle égale ou améliore la capacité de traitement des informations d'un cerveau humain, on affirme maintes fois que c'est une question d'informatique et qu'il ne s'agit pas d'une véritable "intelligence". Si l'application *Siri d'Apple* nous donne aujourd'hui des réponses intelligentes, nous pensons d'habitude qu'elle n'est pas "tellement" intelligente car elle échoue souvent ou car elle n'est pas en mesure de reconnaître ce que nous exprimons, outre le fait que dans d'autres cas elle donne des réponses efficaces en moins d'une seconde. Ce qui importe pour pouvoir mesurer le défi de réglementer cette nouvelle technologie c'est de comprendre que tout comme le cerveau extrait, sélectionne, découpe et organise l'information disponible afin de prendre des décisions, l'intelligence artificielle fait de même par d'autres méthodes et à une vitesse différente. Pour plus de précisions, voir Bostrom, Nick, *Superinteligencia [La super-intelligence]*, p. 29, 2<sup>e</sup> ed., España, Tell, 2016, et Kurzweil, Ray, *La singularité est proche*, p. 302.

<sup>15</sup> Pour plus de précisions, voir Domingos, Pedro, *The master algorithm: how the quest for the ultimate learning machine will remake our world*, New York, Basic Books, 2015, p. 1 et suiv.; Harari, Yuval N., *Homo Deus*, p. 99 à 107, Bs. As., Debate, 2016; Borruso, Renato, *La ley, el juez, la computadora. Un tema fundamental de la informática jurídica [La loi, le juge, l'ordinateur. Un sujet fondamental de l'informatique juridique]*, "Informática y derecho" [*Informatique et droit*], vol. 5, 1996, p. 30 et suiv.

<sup>16</sup> Pour plus de précisions, voir Palma Méndez, José T. - Marín Morales, Roque, *Inteligencia artificial*, Madrid, Mc Graw-Hill, 2011, p. 683, y Abu-Mostafa, Yaser S., *Técnicas de aprendizaje automático [Les techniques de l'apprentissage automatique]*, p. 50 à 53, Investigación y ciencia, abril, 2013.

<sup>17</sup> Deutsch, David - Ekert, Artur, *Más allá del horizonte cuántico [Par-delà l'horizon quantique]*, Investigación y ciencia, novembre 2012, p. 79.

<sup>18</sup> Benítez, Raúl - Escudero, Gerard - Kanaan, Samir - Masip, Rodó David, *Inteligencia artificial avanzada [L'intelligence artificielle avancée]*, p. 13.

<sup>19</sup> Pour plus de précisions, voir Domingos, Pedro, *The master algorithm: how the quest for the ultimate learning machine will remake our world*, p. 1.

La logique de ces nouvelles technologies s'explique essentiellement par le processus consistant à "algorithmiser"<sup>20</sup> le traitement des données ou des informations, afin de se substituer ou d'améliorer les activités ou les tâches qui sont le fruit de l'intelligence humaine. C'est à partir de ces éléments d'argumentation que nous pouvons poser trois affirmations très importantes à prendre en considération: 1) il existe un consensus pour affirmer que les organismes, et donc les êtres humains, sont des algorithmes<sup>21</sup> complexes et sophistiqués qui fonctionnent à travers les sensations, émotions/désirs et pensées<sup>22</sup>; 2) le processus d'innovation en matière de technologie augmente de façon soutenue sa capacité à rendre "algorithmisables" les processus cognitifs<sup>23</sup>; 3) dans plusieurs domaines, les algorithmes deviennent de plus en plus précis et plus efficaces que les êtres humains<sup>24</sup>.

Or ce processus sans cesse croissant destiné à "algorithmiser" l'intelligence est lié à l'explosion des données et de l'information<sup>25</sup>. Afin de mieux comprendre le phénomène, il convient de souligner deux grandes questions concernant le concept d'IA et le rôle qu'elle joue.

*Premièrement*, au sens large, lorsque nous parlons d'intelligence artificielle, nous nous référons aux "stratégies de programmation et aux machines physiques en vue de reproduire de la manière la plus efficace et complète possible les tâches aussi bien cognitives que scientifiques et techniques appelées "intelligentes"<sup>26</sup>. L'IA développe certains processus de façon indépendante, souvent sans imiter ou simuler le processus qui se déroule dans le cerveau humain, mais obtient des résultats égaux ou meilleurs dans plusieurs domaines ou secteurs de connaissance.

---

<sup>20</sup> Borruso a inventé le mot "algorithmiser" pour faire une analogie de l'ordinateur comme une "machine algorithmique universelle". Borruso, Renato, *La ley, el juez, la computadora. Un tema fundamental de la informática jurídica [La loi, le juge, l'ordinateur. Un sujet fondamental de l'informatique juridique]*, p. 35.

<sup>21</sup> Harari, Yuval N., *Homo Deus*, p. 350; Kraemer, Felicitas - van Overveld, Kees - Peterson, Martin, *Is there an ethics of algorithms?*, p. 251 à 260, *Ethics and Information Technology*, vol. 13, Springer, Netherlands, 2011.

<sup>22</sup> Selon Daniel Kahneman, 99% de nos décisions sont prises par des algorithmes très sophistiqués que nous appelons sensations, émotions et désirs. *Pensar rápido, pensar despacio [Penser vite, penser lentement]*, Barcelona, Debate, 2012. Par contre, les distributeurs automatiques fonctionnent grâce à des engrenages mécaniques et circuits électriques. Harari, Yuval N., *Homo Deus*, p. 101.

<sup>23</sup> Borruso, Renato, *La ley, el juez, la computadora. Un tema fundamental de la informática jurídica [La loi, le juge, l'ordinateur. Un sujet fondamental de l'informatique juridique]*, p. 35.

<sup>24</sup> Kahneman, Daniel, *Pensar rápido, pensar despacio [Penser vite, penser lentement]*, p. 291 et suiv. Dans une étude, un algorithme informatique a identifié 90% des cas de cancer du poumon qui lui ont été posés, tandis que les médecins n'ont réussi le diagnostic que dans 50% des cas. Harari, Yuval N., *Homo Deus*, p. 347 et Tzezana, Roey, *The guide to the future*, Haifa, 2013, p. 62 à 64.

<sup>25</sup> L'exemple suivant permet de comprendre la portée exacte de cette question. Le 29 mai 2017, dans une minute 2 432 440 845 (presque deux milliards et demi) courriels ont été envoyés, on a envoyé 458 090 tweets, 63 980 photos ont été publiées sur Instagram, 3 629 947 recherches ont été effectuées sur Google et 2 702 994 gigabytes ont été traités sur le Web. Source: Internet Live Stats [www.internetlivestats.com/one-second/#instagraband](http://www.internetlivestats.com/one-second/#instagraband).

<sup>26</sup> Pour plus de précisions, voir Mira, José, *Aspectos conceptuales de la inteligencia artificial y la ingeniería del conocimiento [Des aspects conceptuels de l'intelligence artificielle et de l'ingénierie des connaissances]*, in Méndez Palma, José T. - Morales Marín, Roque, *Inteligencia artificial [l'intelligence artificielle]*, p. 3 à 9; dans le même sens, Ruiz, Francisco E. - Quevedo, Miguel A. C. - Galipienso, María I. A. - Pardo, Otto C. - Lozano Ortega, Miguel A., *Inteligencia artificial [l'intelligence artificielle]*, Madrid, Thomson, 2003, p. 4.



*Deuxièmement*, les machines intelligentes jouent un rôle toujours plus déterminant dans l'explosion de l'information et des données que nous connaissons aujourd'hui. L'IA est essentielle en ce qu'elle optimise le processus d'extraction des informations ou des données utiles. En effet, "l'information" est "savoir" dans la mesure où les activités intelligentes trouvent des modèles pertinents<sup>27</sup>. Ici, nous touchons un point fondamental que nous allons illustrer par un exemple. *Google* n'a pas besoin de connaître la structure grammaticale des différentes langues pour effectuer des traductions plutôt raisonnables et qui supposent a minima une compréhension élémentaire des informations écrites dans une autre langue. À la place, il apprend des modèles qu'il extrait des informations et des données. Ainsi, plusieurs chemins –nombreux et compatibles les uns avec les autres– sont empruntés par l'intelligence artificielle afin d'égaliser, améliorer, renforcer et/ou maximiser les résultats de beaucoup d'activités humaines intelligentes par le biais de l'optimisation continue et exponentielle du traitement des informations.

### 3. L'intelligence artificielle selon une approche fondée sur les droits de l'homme

L'intelligence artificielle suppose une innovation technologique d'une telle envergure que, en début d'année, plus de deux mille cinq cents experts (dont Stephen Hawking<sup>28</sup>) ont énoncé le principe suivant: "*Les risques que présentent les systèmes d'IA, en particulier les risques de catastrophe et existentiels, devraient appeler à déployer des efforts de planification et d'atténuation conformément à l'impact attendu. Et dans une plus grande mesure, ils devraient faire l'objet de mesures strictes de sécurité et de contrôle*"<sup>29</sup>. Cependant, nous affrontons sur le court terme des défis multiples pour assurer la compatibilité du développement de l'IA avec le droit interne des États et avec le droit international en vigueur. L'utilisation des algorithmes intelligents pose actuellement trois problèmes principaux. Tout d'abord, dans de nombreux cas, l'IA ne peut offrir une explication détaillée sur la manière d'atteindre un résultat donné: l'algorithme présente un défaut de motivation. Dans certains cas, on ne connaît pas le processus de raisonnement utilisé lors de l'interaction entre les informations et les données pour prendre des décisions, établir des résultats ou faire des prédictions. Autrement dit, on ne peut pas déterminer comment l'algorithme évalue et pondère les données et les informations qu'il traite, ce qui entraîne le phénomène appelé "boîte noire"<sup>30</sup>. Ensuite, dans le cas de certaines IA prédictives, la forme ou la méthode de traitement des informations et des données peut enfreindre le principe d'égalité et de non-discrimination, puisque les prédictions sont basées sur un code source qui prend en considération des distinctions de race et de sexe, entre autres. Par exemple, dans

---

<sup>27</sup> Kurzweil, Ray, *La singularité est proche*, p. 426.

<sup>28</sup> Elon Musk, Royal Martin Rees, Smith-Zadeh y Stuart Russell, Sam Harris, Nick Bostrom, Raymond Kurzweil, David Chalmers et Demis Hassabis, parmi bien d'autres, ont souscrit aux principes.

<sup>29</sup> Asilomar AI Principles. Future of Life Institute, disponible sur <https://futureoflife.org/ai-principles>.

<sup>30</sup> On appelle système "boîte noire" l'outil informatique dans lequel on comprend les données saisies et les résultats mais on ne comprend pas la procédure sous-jacente. Le code est ici impénétrable parce que le programme "évolue" et les humains ne sont pas en mesure de comprendre le processus de programmation utilisé pour obtenir une solution donnée. Voir Barrat, James, *Nuestra invención final [Notre invention finale]*, Mexique, Paidós, 2014, p. 92.

l'affaire "*State vs. Loomis*"<sup>31</sup> le requérant soutenait que l'algorithme intelligent utilisait de façon incorrecte les évaluations de sexe<sup>32</sup>. Enfin, les développeurs en IA se prévalent généralement du secret des affaires et des droits de brevet.

Nous pourrions fournir une longue liste de problèmes, mais ces courts exemples mettent en évidence la complexité et les difficultés relatives au phénomène de l'IA. Tous ces aspects ne pouvant être développés dans cet article, notre proposition se concentrera sur les droits de l'Homme. Selon la théorie du droit, les droits de la personne humaine sont définis par cinq caractéristiques particulières: 1) universels, 2) fondamentaux, 3) abstraits ou indéterminés<sup>33</sup>, 4) moraux -existence et validité morale- et 5) prioritaires<sup>34</sup>. Le "modèle des droits de l'homme" se cristallise à partir d'un "paradigme protectif" résultant des conventions internationales et reposant essentiellement sur l'attribution des deux traits suivants aux droits essentiels de la personne humaine: l'égalité<sup>35</sup> et leur caractère "inaliénable"<sup>36</sup> ou "indérogeable"<sup>37</sup>. Il existe un lien direct entre ces droits, la dignité humaine, la paix, la protection des minorités, des plus vulnérables ou des plus faibles<sup>38</sup>. C'est un mécanisme en vertu duquel les États et la communauté internationale<sup>39</sup> sont obligés d'assurer l'effectivité des droits, principes et règles consacrés dans les constitutions, les conventions internationales et la

<sup>31</sup> Cour Suprême de Wisconsin, "*L'État de Wisconsin c/Eric L. Loomis*", le 13 juillet 2016, disponible sur [www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690](http://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690).

<sup>32</sup> Pour plus de précisions, voir Cour Suprême de Wisconsin, "*L'État de Wisconsin c/Eric L. Loomis*", 13 juillet 2016, voir en particulier les considérants 17, 28, 34, 51, 93 et 94.

<sup>33</sup> Beitz, Charles R., *La idea de los derechos humanos [L'idée des droits de l'homme]*, p. 244, Madrid, Marcial Pons, 2012. Dans le même sens, en Amérique Latine, voir Rincón Córdoba, Jorge I., *Las generaciones de los derechos fundamentales y la acción de la Administración pública [Les générations des droits fondamentaux et l'action de l'Administration publique]*, 2<sup>e</sup> ed., Université Externado de Colombie, juillet 2004, p. 335 et 336.

<sup>34</sup> Alexy, Robert, *La doble naturaleza del derecho [La double nature du droit]*, Madrid, Trotta, 2016, p. 60 et 61.

<sup>35</sup> Voir, Déclaration Américaine des Droits et Devoirs de l'Homme, approuvée à la IX<sup>e</sup> Conférence Internationale Américaine à Bogotà, Colombie, 1948, premier paragraphe du Préambule; voir aussi le Préambule, premier considérant, la Convention Internationale des Droits Économiques, Sociaux et Culturels, adoptée et ouverte à la signature, la ratification et l'adhésion par l'Assemblée Générale dans sa résolution 2200 A (XXI), du 16 décembre 1966; Préambule, premier considérant de la Convention Internationale des Droits Civils et Politiques, approuvé à New York, États-Unis d'Amérique, le 19 décembre 1966; Préambule, premier considérant, premier paragraphe de la Convention relative aux Droits des Enfants, adoptée par l'Assemblée Générale des Nations-Unies le 20 novembre 1989; premier considérant de la Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes (désignée aussi par son acronyme anglais CEDAW), approuvée par l'Assemblée Générale des Nations-Unies le 18 décembre 1979 et premier considérant de la Convention contre la torture et autres peines ou traitements cruels, inhumains ou dégradants, approuvée par l'Assemblée Générale des Nations-Unies le 9 décembre 1975.

<sup>36</sup> Voir premier considérant, premier paragraphe de la Convention contre la torture et autres peines ou traitements cruels, inhumains ou dégradants et le Préambule, premier considérant, premier paragraphe de la Convention relative aux Droits des Enfants.

<sup>37</sup> Voir deuxième considérant, premier paragraphe de la Convention Interaméricaine sur la disparition forcée des personnes.

<sup>38</sup> Ferrajoli, Luigi, *Sobre los derechos fundamentales [Sur les droits fondamentaux]*, in *Teoría del neoconstitucionalismo [La théorie du néo-constitutionnalisme]*, Madrid, Trotta, 2007, p. 73 à 75.

<sup>39</sup> La Déclaration du Millénaire de l'ONU affirme que des efforts doivent être réalisés pour assurer "*dans tous les pays, la promotion et la protection intégrale des droits civils et des droits politiques, économiques, sociaux et culturels de chacun... De renforcer, dans tous les pays, les capacités nécessaires pour appliquer les principes et pratiques de la démocratie et du respect des droits de l'homme, y compris les droits des minorités*" (point 25). Disponible sur [www.un.org/es/comun/docs/?symbol=A/RES/55/2](http://www.un.org/es/comun/docs/?symbol=A/RES/55/2).

législation interne<sup>40</sup>. Sur la base de cette plateforme, pour que le développement de l'IA soit compatible avec un “modèle des droits de l'homme”, il faut encourager l'adoption d'une réglementation contenant un nombre de principes que nous énoncerons cidessous<sup>41</sup>.

#### 4. Les principes juridiques visant à une intelligence artificielle respectant et préconisant le maintien de l'État constitutionnel et l'effectivité des droits de l'homme

Placer l'IA dans le cadre de l'État constitutionnel invite à aborder sa conception, son développement et son utilisation au regard du respect de la dignité humaine et des droits de l'homme. Il s'agit, pour l'essentiel, de construire un système de protection plus fort par l'incorporation au bloc de juridicité d'une série de principes généraux visant à réguler l'IA. Ces principes pourraient être les suivants.

*Prévention/précaution.* Ces principes ont deux fonctions différentes, avec un dénominateur commun: le besoin d'agir avant la production d'un dommage. De façon schématique, nous pourrions dire qu'ils répondent à différents types de risques. Face au risque potentiel correspond, la précaution. Face au risque vérifié correspond, la prévention<sup>42</sup>. Le principe de précaution dans l'IA (de façon analogue à ce que l'on rencontre dans le droit de l'environnement) est lié à un *manque total ou absolu de certitude scientifique sur l'absence de risques*, laquelle dans le cas d'algorithmes intelligents doit, au moins, être associée aux circonstances suivantes: 1) un code source fermé ou l'existence d'un système dans lequel on comprend les données chargées et les résultats, mais dont on ne peut pas déduire le traitement sous-jacent (“boîte noire”); 2) l'absence de traçabilité; 3) l'impossibilité de fournir un «bouton d'arrêt» ou un *mécanisme sûr* de contenance de l'IA; 4) l'IA porte atteinte de manière avérée ou potentielle aux droits fondamentaux de l'homme, aux droits d'incidence collective, aux services publics, aux services essentiels ou services universels; 5) des biais considérables sont constatés à une étape quelconque –conception, développement ou application–, ou la transparence et/ou le principe de non-discrimination ne peuvent être garantis.

*Autodétermination algorithmique.* L'autodétermination est un droit fondamental issu de la dignité de la personne humaine. Il s'agit d'assurer le “libre développement de la personnalité», à partir de la reconnaissance de l'autodétermination informative visant à garantir la liberté de choix –liée à la liberté de l'information–, le “droit à savoir”,

---

<sup>40</sup> Toutefois, tel que l'affirme Charles Beitz, il n'est pas plausible de trouver un fondement unique ou de formuler une liste de droits. Beitz, Charles R., *La idea de los derechos humanos [L'idée des droits de l'homme]*, p. 141, 142 à 244.

<sup>41</sup> Il est important de prendre en considération l'approche de l'Organisation des Nations- Unies lorsqu'elle décide de “faire en sorte que les avantages des nouvelles technologies, en particulier des technologies de l'information et de la communication, soient accordés à tous, conformément aux recommandations contenues dans la Déclaration ministérielle du Conseil économique et social de 2000”, résolution adoptée par l'Assemblée Générale 55/2, Déclaration du Millénaire, titre III, point 20, cinquième paragraphe.

<sup>42</sup> Allende Rubino, Horacio L., *La acción de prevención en el Código Civil y Comercial. Su relación con el principio de precaución en el derecho ambiental [L'action de prévention au Code civil et commercial. Son rapport avec le principe de précaution au droit de l'environnement]*, Microjuris online, 2016, cite: MJ-DOC-9989-AR | MJD9989.

à la “connaissance” et à “l’auto-régulation de l’information”<sup>43</sup>. Sur ce fondement, les États et la communauté internationale, sont obligés d’investir de manière responsable et de déployer de grands et divers efforts, afin de garantir l’autodétermination humaine face à l’utilisation des algorithmes intelligents. Puisque l’IA intervient de plus en plus entre les données / l’information et les décisions des personnes, il devient indispensable de protéger leurs droits par la promotion du respect des principes de nécessité, finalité, proportionnalité et propriété des données personnelles<sup>44</sup>.

*Transparence algorithmique.* La conception, le développement et l’utilisation de l’intelligence artificielle doivent être transparents et ouverts. La transparence algorithmique est liée à l’interdiction de “boîtes noires” dans les algorithmes ou la présence de failles, face aux dommages ou lésions pouvant survenir<sup>45</sup>. L’intelligence artificielle doit donc être transparente par rapport à ses décisions, ce qui signifie que l’on puisse inférer ou déduire une “explication compréhensible”, conformément aux critères utilisés pour arriver à une conclusion déterminée, une proposition ou un résultat<sup>46</sup>.

*Traçabilité de l’intelligence artificielle.* La traçabilité ou le traçage est “l’aptitude à suivre l’histoire, l’application ou la localisation d’une entité par le biais des indicateurs enregistrés”<sup>47</sup>. Une IA, établie sur une approche des droits de l’homme, doit exprimer, pas à pas, les opérations techniques réalisées depuis le début jusqu’à la fin d’un processus déterminé. Il faut, comme règle, *garantir l’intelligibilité et la traçabilité du processus de prise de décisions des algorithmes intelligents*. En contrepartie, il convient d’empêcher la configuration des “boîtes noires”. La traçabilité contribue à garantir les principes de transparence et de motivation algorithmiques.

*Accès maximal. Droit d’accès à l’information algorithmique.* L’État est tenu d’assurer l’accès au code source et aux systèmes sur lesquels reposent les algorithmes intelligents, lorsqu’il peut s’avérer une violation à la Constitution nationale, aux traités internationaux et aux normes en vigueur. Lorsque l’État et les personnes publiques non étatiques, pour soi ou pour un tiers, conçoivent, développent ou utilisent des technologies de l’information ou de la communication fondées sur l’IA ou sur des algorithmes intelligents (ce qui inclut tout type de machine ou *robot* intelligent), ils sont tenus

---

<sup>43</sup> Pitschas, Rainer, *Derecho administrativo de la información [Droit administratif de l’information]*, pp. 226, 227 et 236; *Innovación y reforma en el derecho administrativo [Innovation et réforme au droit administratif]*, Global Law Press, 2<sup>e</sup> ed., Seville, 2012. Cet auteur fait référence à un changement de paradigme du *droit administratif de l’information*, dans lequel chaque individu doit être capable de décider, sous sa seule responsabilité et autonomie, parmi les possibilités et les risques issus de la liberté de communication (p. 236).

<sup>44</sup> Rodotà, Stefano, *El derecho a tener derechos [Le droit à avoir des droits]*, Madrid, Trotta, 2014, p. 182.

<sup>45</sup> Tel qu’il est affirmé à l’un des 23 principes d’Asilomar, si un système d’IA provoque un dommage, il serait possible de déterminer la cause (principe 7 dénommé transparence dans la faille).

<sup>46</sup> La transparence comprend ce qu’on appelle “motivation algorithmique”. La motivation algorithmique est entendue comme la description, dans un langage formel et intelligible, de l’architecture de la conception computationnelle; l’algorithme doit être capable d’exprimer et faire intelligible comment, quoi et quelles sont les questions tactiques, juridiques et /ou éthiques remarquables à prendre en compte pour adopter telle ou telle décision. La motivation algorithmique permet l’*auditabilité humaine* qui est liée aussi à la règle de la traçabilité des systèmes d’IA et au principe de non-discrimination algorithmique, ce qui suppose l’*interdiction des biais de machine*.

<sup>47</sup> Définition conformément à la norme ISO 8402, complément de la série des normes ISO 9000.



de garantir un maximum d'accès au système de traitement de l'information réalisé par ces technologies<sup>48</sup>.

*Principe de non discrimination algorithmique.* La conception et/ou la mise en oeuvre des algorithmes intelligents doit respecter le principe de non-discrimination, lequel empêche les machines intelligentes –fondées sur des systèmes d'IA– de traiter l'information ou les données sous des biais ou des distinctions entre êtres humains, pour des motifs de race, de couleur de peau, de sexe, de langue, de religion, d'opinions politiques ou d'autres types, d'origine nationale ou sociale, de position économique, de naissance ou de toute autre condition sociale (art. 2 alinéa 2, Pacte des Droits Economiques Sociaux et Culturels).

## 5. Conclusion

Le fait d'appeler à un cadre normatif relatif au développement de l'IA peut être regardé de façon similaire au phénomène survenu avec la naissance des Technologies de l'information et de la communication (TIC) au niveau de l'Organisation des Nations Unies (ONU)<sup>49</sup>. En fin de compte, cette innovation technologique est survenue par l'optimisation artificielle du traitement de l'information.

Il semble que la tâche visant à rendre compatible l'IA avec les droits de l'Homme est le défi le plus complexe et indispensable de l'avenir, et il convient de prendre en compte la configuration émergente de deux scénarios: d'un côté, comment garantir l'intervention humaine face aux décisions ou prédictions d'algorithmes intelligents, conformément aux principes signalés au point précédent? Et de l'autre côté, si les spécialistes en protection de données étudient l'intervention humaine par rapport aux algorithmes<sup>50</sup>, il faut donc travailler sur la question suivante: *combien* d'interventions de l'homme sont nécessaires pour que le résultat du traitement de l'information et des données devienne légitime, respectueux et promoteur de l'efficacité des droits des personnes?

De manière prospective, nous estimons que les défis posés par l'IA concernent notre identité en tant qu'espèce. Si les personnes humaines sont caractérisées par la *diversité, le caractère aléatoire et l'imperfection*<sup>51</sup>, nous sommes en train d'entrer dans une ère d'automatisation qui pourrait mettre en péril ces caractéristiques. Même si cela semble peu probable, dans un avenir proche, il conviendra de penser sérieusement à

---

<sup>48</sup> Il faut garantir cet accès à toute personne humaine ou morale publique ou privée, liée aux fins publiques ou fonds publics reçus, et qui utilise ces technologies, pourvu que la conception ou l'utilisation de l'intelligence artificielle soit liée aux fins publiques ou fonds publics reçus.

<sup>49</sup> Par exemple, depuis l'UNESCO la création d'un milieu législatif propice est motivé dans le domaine des TIC. Dans le même sens, l'ONU croit qu'il faut appuyer le développement des technologies, la recherche et l'innovation nationale, tout en garantissant un milieu normatif propice pour la diversification industrielle et la majoration de valeur aux produits de base. En plus, les États doivent s'abstenir d'utiliser la technologie de l'information et les communications en violation du droit international. V. Assemblée générale, résolution n° A/71/307, 05/08/2016, considérant 8, p. 4, Assemblée générale, résolution n° A/RES/70/1, 21/10/15, considérant 9b, p. 23 et Assemblée générale, résolution n° A/RES/71/101 A-B, 23/12/16, considérant 4, p. 4.

<sup>50</sup> Rodotà, Stefano, *El derecho a tener derechos [Le droit à avoir des droits]*, p. 302.

<sup>51</sup> Kurzweil, Ray, *La singularidad está cerca [La singularité est proche]*, p. 167 et 168.



garantir un nouveau droit fondamental, lequel deviendrait le fondement de l'ère de l'intelligence artificielle: le droit à la diversité aléatoire et imparfaite, inhérente à l'homme.

© Editorial Astrea, 2018. Todos los derechos reservados.

