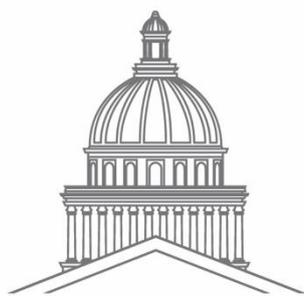


# **Université Paris II- Panthéon-Assas**

**École doctorale de droit international, droit européen, relations  
internationales et droit comparé**

Thèse de doctorat en droit soutenue le 18 décembre 2023

## **Le droit international et l'intelligence artificielle forte**



**UNIVERSITÉ PARIS II**  
PANTHÉON-ASSAS

**Yi Ting CHEN**

Sous la direction de Madame Marie Goré

Membres du jury :

Madame Ana Zelcevic-Duhamel, de l'Université Paris Cité (rapporteur)

Madame Marion Ho-Dac, de l'Université d'Artois (rapporteur)

Madame Sophie Gjidara-Decaix, de l'Université Paris II Panthéon-Assas

### Avertissement

La Faculté n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse ; ces opinions doivent être considérées comme propres à leur auteur.

## ***Remerciements***

*Mes remerciements vont tout à bord à Madame le Professeur Marie Goré, pour son encadrement assurément indéfectible ainsi que Formosa Taïwan dans son souvenir apprécié.*

*Mes remerciements vont ensuite aux Mesdames Ana Zelcevic-Duhamel, Marion Ho-Dac et Sophie Gjidara-Decaix, d'avoir accepté de lire et d'évaluer mes travaux.*

*Je tiens aussi à remercier tous ceux qui ont reçu mes remerciements tout au long de ma vie. Je leur suis très sincèrement reconnaissante.*

« Je vous en remercie. Tant que vous attendez ici demain. » — Akiyuki Nosaka

« Peu importe de quoi s'agit-il, l'on reçoit trop des autres et l'on donne trop peu. Comme il y a trop de gens à remercier, alors, merci Dieu. » — Chen Zhi-Fan

*À toutes les possibilités et particularités des sciences et des civilisations.*

*À toutes les souffrances commises par (l'ignorance de) l'humanité ou par son omission.*

*Aux Formose et Pescadores. Orphelin sui generis d'autochtones austronésiens, nommé par les Portugais « Ilha formosa », devenu Formose espagnole, Formose néerlandaise, le Royaume de Tungning, et puis considéré par l'empereur KangXi : « Taiwan est en dehors de notre empire et n'a pas de grande importance »<sup>1</sup>, mais au final cédé au Japon. De la République de Formose à l'empire du Japon, devenu l'objet souverain des États-Unis en tant que « principale puissance occupante »<sup>2</sup>, mais occupés par un exil de Terreur blanche, qui a massacré des intellectuels, emprisonné et fait disparaître des gens pour leurs crimes de pensée, infligé des amendes, frappé, ou forcé les Taïwanais à porter une pancarte s'ils parlaient taïwanais ou japonais.*

---

<sup>1</sup> VAN DER WEES G., « Has Taiwan Always Been Part of China? », The Diplomat, 1 déc. 2020.

<sup>2</sup> HANSEN P. C., RUBIN M., « China Has No Legal Claim to Taiwan », The National Interest, 8 oct. 2022.

**Résumé :**

À travers des concepts juridiques, l'on a pu constater que la notion du statut juridique se rattache largement à l'idée de l'espèce, et dépend en outre principalement de l'identification d'une manifestation de capacité, c'est-à-dire une sorte d'intelligence.

D'après le critère d'espèce, l'on peut encore constater une grille de classement du statut juridique en fonction du degré de la capacité mentale ou de discernement; encore une fois, cela revient donc à la notion d'intelligence.

Et parmi notre propre espèce, l'on peut remarquer que, le déficit ou la dégradation de tout organe ne suffit pas à justifier toute forme de dérogation à un statut intact et complet, sauf une seule exception : notre cerveau.

Autrement dit, cet organe traditionnellement considéré chargé d'intelligence détermine essentiellement le statut juridique au sens traditionnel. À présent, l'on assiste à l'émergence d'une nouvelle espèce qui va réduire ou transformer l'utilisation de notre cerveau voire le rendre obsolète et créer une nouvelle hiérarchie de l'intelligence.

Ce phénomène n'appartient pas par définition et de facto à une réponse qui connaît la frontière dite nationale. Ce travail de recherche consiste à élaborer une hiérarchisation des normes pour ces conflits évitables.

Le système proposé dans cette thèse est composé de trois degrés. Son premier degré consiste dans un principe de faveur qui favorise l'application d'autorégulation telle que la RSE multinationale pour son effet étendu. Au deuxième niveau se trouve un répertoire des normes d'IAG. Enfin, l'on retrouve au sommet de cette hiérarchie un statut de l'espèce humaine. Cette étude marque le commencement du droit inter-espèces.

*Descripteurs : IAG, Droit international*

***Abstract :***

Throughout legal concepts, it has been found that the notion of legal status is largely related to the idea of species, which itself depends mainly on the recognition of a manifestation of capacity, that is to say, a kind of intelligence.

According to the criterion of species, we see a classification of the legal status based on the degree of the mental capacity or of discernment; again, that comes down to the notion of intelligence.

And among our own species, we observe that deficit or degradation of any organ is not enough to justify any form of derogation to an intact and full status, with only one exception: our brain.

In other words, this organ which is traditionally considered responsible for human intelligence, determines basically the legal status in the traditional sense. And now, we have a new species that will create a new hierarchy of intelligence.

This phenomenon does not belong by definition and de facto to an answer that knows the concept of the so-called national border. This work of research involves applying a hierarchy of norms to avoid evitable conflicts.

The system proposed by this study is composed of three levels. Its first level consists of a principle of favor which supports the application of self-regulation such as multinational CSR for its extended effect. Should be found at the second level, a legal repertoire of AGI. Finally, at the top of this hierarchy is the status of the human species. This study marks the beginning of the inter-species law.

*Keywords: AGI, International Law*

## *Principales abréviations*

|   |  |
|---|--|
| AAAI  | Association for the Advancement of Artificial Intelligence                       |
| adde.   | ajouter  |
| al.   | alinéa   |
| AMA   | Artificial Moral Agents (Agent Moral Artificiel)                                 |
| anc.  | ancien   |
| art.  | article  |
| Ass. plén.  | Assemblée plénière de la Cour de cassation française                             |
| AHI   | Artificial Human-level Intelligence  |
| AI  | Artificial Intelligence  |
| AGI   | Artificial General Intelligence  |
| BATX  | Baidu, Alibaba, Tencent et Xiaomi  |
| BCS   | British Computer Society Code of Conduct   |
| BGB   | Bürgerliches Gesetzbuch, le code civil allemand                                  |
| BGH   | Bundesgerichtshof, la Cour fédérale de justice allemande                         |
| BTBI  | Brain-to-brain interface   |
| Cass. civ. (1 <sup>re</sup> , 2 <sup>e</sup> , 3 <sup>e</sup> ) | Chambre civile de la Cour de cassation française (première, deuxième, troisième) |
| Cass. Ch. réun.   | Chambres réunies de la Cour de cassation française                               |
| CNRS  | Centre national de la recherche scientifique                                     |
| D.  | Recueil Dalloz   |
| éd.   | édition  |
| GAFAMI  | Google, Amazon, Facebook, Apple, Microsoft et IBM                                |
| HBR   | Harvard Business Review  |
| Harv. Law Rev.  | Harvard Law Review   |
| Harv. J.L. & Tech.  | Harvard Journal of Law & Technology  |
| IA  | Intelligence Artificielle  |
| IAG   | Intelligence Artificielle Générale   |
| IEEE  | Institute of Electrical and Electronics Engineers                                |
| infra   | ci-dessous   |
| ISO   | Organisation internationale de normalisation                                     |
| LGDJ  | Librairie générale de droit et de jurisprudence                                  |
| MIT   | Massachusetts Institute of Technology  |
| MIT Tech. Rev.  | MIT Technology Review  |
| ML  | Machine learning (l'apprentissage automatique)                                   |
| n <sup>o</sup>  | numéro   |
| NYT   | New York Times   |
| op. cit.  | opus citatum (œuvre citée)   |
| p.  | page ou pages  |
| pt.   | point  |
| préc.   | précité  |
| RFDA  | Revue française de droit administratif   |
| RGPD  | Règlement général sur la protection des données                                  |
| RSE   | Responsabilité sociale des Entreprises   |
| RTD civ   | Revue trimestrielle de droit civil   |

|         |  |
|---------|--|
| RTD com | Revue trimestrielle de droit commercial et de droit économique |
| s.      | et suivants  |
| supra   | ci-dessus  |
| Th.     | Thèse  |
| t.      | Tome   |
| UE      | Union européenne   |
| V.      | voir   |
| VPA     | Virtual Personal Assistant (assistants virtuels personnels)    |
| vol.    | Volume(s)  |
| WSJ     | Wall Street Journal  |
| p. ex.  | par exemple  |



## Sommaire

Introduction

PARTIE I : La nouvelle hiérarchie de l'intelligence

Titre 1 : L'évolution objective de l'intelligence

Chapitre 1 : L'introduction de l'intelligence artificielle forte

Chapitre 2 : L'effet de l'intelligence artificielle forte

Titre 2 : La conséquence subjective de l'intelligence

Chapitre 1 : L'effet de l'anthropomorphisme et de l'empathie

Chapitre 2 : La coexistence de personnes juridiques intelligentes

PARTIE II : La nouvelle hiérarchie des normes

Titre 1 : Premier degré d'application : l'autorégulation multinationale étendue

Chapitre 1 : La méthode intrinsèque de l'Agent Moral Artificiel

Chapitre 2 : L'autorégulation étendue

Titre 2 : Deuxième degré d'application : la synchronisation du répertoire de l'IAG

Chapitre 1 : La compétence inclusive de l'IAG

Chapitre 2 : Le répertoire de l'IAG



*« Men have become the tools of their tools. » - Henry David Thoreau*

*« La science ne connaît pas de pays, parce que la connaissance appartient à l'humanité, et elle est la torche qui illumine le monde. La science est la plus grande personnification de la nation parce que cette nation restera la première en portant au plus loin les travaux de la pensée et de l'intelligence » - Louis Pasteur*

*« Any sufficiently advanced technology is indistinguishable from magic. »*

- Arthur C. Clarke

*« It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers. They would be able to converse with each other to sharpen their wits. At some stage therefore, we should have to expect the machines to take control. » - Alan Turing*

*« The train started slowly, but gathered speed sooner than I had expected. »*

- Winston Churchill

*« There is something better than science and that is science with a moral compass, science in the service of humanity, science that makes current deeds responsive to future needs. »*

- William Foege

*« We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run. » - Roy Amara*

*« If our small minds, for some convenience, divide this glass of wine, this universe, into parts — physics, biology, geology, astronomy, psychology, and so on — remember that nature does not know it! » - Richard Feynman*

*« Non, inquirimus an causa sit, quia nihil est per se notius. » Francisco Suárez*



## Introduction

---

*« If a machine can think, it might think more intelligently than we do, and then where should we be? Even if we could keep the machines in a subservient position, for instance by turning off the power at strategic moments, we should, as a species, feel greatly humbled. » — Alan Turing*

**1. La naissance de l'intelligence artificielle (IA) contemporaine : le test de Turing et la conférence de Dartmouth.** — Il y a moins d'un siècle, Alan Turing s'est posé la question suivante : « Les machines peuvent-elles penser ? » Il a suggéré qu'au lieu de se demander si les machines peuvent penser, il fallait se demander si les machines peuvent réussir un test d'intelligence. Il a estimé qu'un ordinateur mériterait d'être qualifié d'intelligent s'il pouvait tromper un humain en lui faisant croire qu'il était humain. Il a conclu qu'« à la fin du siècle, l'utilisation de mots et d'opinions éduquées sera tellement changée que l'on pourra parler de machines pensantes sans s'attendre à être contredit »<sup>3</sup>.

Le terme « intelligence artificielle » (ci-après IA) est né à l'occasion de la conférence de Dartmouth aux États-Unis. La proposition de cet atelier de 1956 définissait le domaine de l'IA comme « chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut être si précisément décrit qu'une machine peut être conçue pour le simuler »<sup>4</sup>. D'après le Journal officiel de la République française, c'est un « champ interdisciplinaire théorique et pratique qui a pour objet la compréhension de mécanismes de la cognition et de la réflexion, et leur imitation par un dispositif matériel et logiciel, à des fins d'assistance ou de substitution à des activités humaines »<sup>5</sup>.

---

<sup>3</sup> TURING A. M., « Computing Machinery and Intelligence », *Mind*, New Series, vol. 59 n°236, Oxford University Press, oct. 1950, p. 443 et 450.

<sup>4</sup> MCCARTHY J. et *alii.*, « A proposal for the Dartmouth Summer Research Project on Artificial Intelligence », 31 août 1955.

<sup>5</sup> Vocabulaire de l'intelligence artificielle (liste de termes, expressions et définitions adoptés), JORF n°0285 du 9 décembre 2018, texte n° 58.

**2. Une arme à double tranchant à résoudre par un consensus multilatéral.** — Lors d'un entretien, Sundar Pichai, le PDG d'Alphabet, considère que « L'intelligence artificielle est l'une des choses les plus importantes sur lesquelles l'humanité travaille. Elle est plus profonde que, je ne sais pas, l'électricité ou le feu ». La journaliste Kara Swisher a réagi par : « Le feu ? Le feu est plutôt bon ». Pichai lui a répondu que « Eh bien, ça tue aussi les gens. Nous avons appris à exploiter le feu pour les bienfaits de l'humanité, mais nous avons dû aussi surmonter ses inconvénients »<sup>6</sup>. Il estimait que « la seule façon de résoudre certains de ces problèmes plus profonds est le cadre multilatéral mondial »<sup>7</sup>.

Pour Mme Fei Fei Li, son idée d'IA centrée sur l'homme (Human-Centered AI) consiste à dire que rien n'est « artificiel » à propos de l'IA : « L'IA est inspirée par les gens, créée par les gens et, surtout, elle a un impact sur les gens. C'est un outil puissant que nous commençons tout juste à comprendre, et c'est une responsabilité profonde »<sup>8</sup>. Dans son discours devant les membres du Congrès américain, elle a souligné que « nous n'avons jamais créé de technologie permettant de reproduire aussi fidèlement les qualités humaines, et nous en savons peu sur son impact sur le monde. Guider son développement constituera un défi éthique, philosophique et humaniste, qui nécessitera une communauté diversifiée de contributeurs. La société sera bientôt confrontée à des changements que nous ne pouvons même pas imaginer, et ces prédictions ne sont qu'un début. Nous pouvons nous attendre à un avenir excitant mais incertain, et un sens de l'humilité nous sera très utile »<sup>9</sup>.

**3. Un phénomène transnational avec des poids de plus en plus lourds.** — En 2017, devant des étudiants de la ville de Yaroslavl en Russie, Vladimir Poutine a déclaré que l'intelligence artificielle était un moyen de diriger le monde. Pour le président russe, « L'intelligence artificielle est l'avenir, non seulement pour la Russie, mais pour toute l'humanité. Celui qui deviendra le leader dans ce domaine, sera le maître du monde », a-t-il déclaré<sup>10</sup>. Effectivement, l'économie mondiale se transforme en économie numérique grâce

---

<sup>6</sup> CLIFFORD C., « Google CEO: A.I. is more important than fire or electricity », CNBC, 1 févr. 2018.

<sup>7</sup> PETROFF A., « Google CEO: AI is more profound than electricity or fire », CNN, 24 janv. 2018.

<sup>8</sup> House Science Committee Testimony, Subcommittee on Research and Technology and Subcommittee on Energy Hearing - Artificial Intelligence – With Great Power Comes Great Responsibility, 115th Congress of U.S.A., 26 June 2018.

<sup>9</sup> *Ibidem*.

<sup>10</sup> TANGUY V., « Poutine pense dominer le monde en maîtrisant l'intelligence artificielle », Sciences et Avenir, 5 sept. 2017.

au développement de technologies numériques de pointe dans des domaines tels que l'IA. Ainsi, le gouvernement indien a doublé l'allocation pour le programme « Digital India » à 30,7 milliards de Roupies pour 2018-19<sup>11</sup>. La Chine, quant à elle, envisage de devenir un leader mondial dans ce secteur<sup>12</sup> ; le leadership des États-Unis, qui conserve encore une bonne avance en la matière, n'est en réalité plus garanti<sup>13</sup>.

Parallèlement, Emmanuel Macron a annoncé un budget de 1,5 milliard d'euros pour « faire de la France un leader de l'intelligence artificielle »<sup>14</sup>. Le gouvernement allemand a approuvé en 2018 une stratégie nationale<sup>15</sup> visant à permettre à l'Allemagne de se hisser parmi les leaders de l'IA. Le Bundestag a en outre institué en 2019 la commission d'étude ad hoc d'IA<sup>16</sup>, la même année pour la création d'un comité pour veiller à l'éthique du numérique et de l'intelligence artificielle en France<sup>17</sup>, une commission de la sécurité nationale sur l'IA aux États-Unis<sup>18</sup>, ainsi qu'un comité ad hoc sur l'intelligence artificielle (CAHAI) par le Comité des Ministres du Conseil de l'Europe<sup>19</sup>. Et désormais, les Émirats Arabes Unis sont le premier pays dans le Moyen-Orient à publier un rapport d'IA, il est également le seul pays au monde qui a créé un ministère de l'intelligence artificielle<sup>20</sup>.

---

<sup>11</sup> BHUNIA P., « Indian Government reveals initiatives in AI, cyber-physical systems and 5G technologies », OpenGov, 9 mars 2018.

<sup>12</sup> V. p. ex., MONTAGNON P., BRAUNE E., « Intelligence Artificielle | Comment la Chine et les États-Unis entendent asseoir leur leadership et souveraineté ? », Forbes, 26 juin 2022.

<sup>13</sup> « Rise of the Machines : Artificial Intelligence and its Growing Impact on U.S. Policy », Subcommittee on Information Technology Committee on Oversight and Government Reform, U.S. House of Representatives, September 2018, p. 4; MANYIKA J. et alii, « Independent Task Force Report No. 77 Innovation and National Security Keeping Our Edge », Council on Foreign Relations (CFR), 2019, pp. 47-48; ALLISON G., SCHMIDT E., « Semiconductor Dependency Imperils American Security », WSJ, 20 juin 2022.

<sup>14</sup> JALINIÈRE H., « Emmanuel Macron : 1,5 milliard d'euros seront alloués à l'intelligence artificielle », Sciences Et Avenir, 29 mars 2018.

<sup>15</sup> V. « Nationale Strategie für Künstliche Intelligenz (KI-Strategie) », nov. 2018.

<sup>16</sup> Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“.

<sup>17</sup> « Un nouveau comité pour veiller à l'éthique du numérique et de l'intelligence artificielle », Maddyness, 3 déc. 2019.

<sup>18</sup> « National Security Commission on Artificial Intelligence Initial Report », juill. 2019. Et aux États-Unis, le bureau national d'intelligence artificielle ainsi que le comité consultatif national sur l'IA ont été prévus. V. « The White House Launches the National Artificial Intelligence Initiative Office », White House Office of Science and Technology Policy (OSTP), 12 janv. 2021; H.R.6216 - National Artificial Intelligence Initiative Act of 2020, 116th Congress.

<sup>19</sup> « Intelligence artificielle : un comité intergouvernemental d'experts chargé d'étudier l'élaboration d'un cadre juridique », Lexis Actu, 13 sept, 2019.

<sup>20</sup> SNOW J., « The World of A.I. », NYT, 18 oct. 2018.

Dans la même tendance, la loi sud-coréenne sur le développement et la diffusion de robots intelligents<sup>21</sup> et la loi américaine sur la compréhension fondamentale de l'utilisabilité et de l'évolution réaliste de l'intelligence artificielle<sup>22</sup> ont toutes été introduites. Dorénavant, les États reconnaissent attentivement l'intérêt voire la nécessité de ce nouveau régime, par conséquent, cette course étatique dans le domaine de l'IA pourrait engendrer des concurrences, et ainsi provoquer des conflits multidimensionnels.

**4. Du capitalisme de surveillance<sup>23</sup> à la souveraineté numérique.** — Aujourd'hui, la méta-valeur des données appartient à des intérêts privés. L'on a donc un avenir où le système public est renforcée et parfois remplacée par des systèmes d'IA centralisés et privés appartenant à des entreprises qui utilisent des données publiques pour générer une richesse énorme pour un très petit nombre de personnes<sup>24</sup>. Ce colonialisme des données désigne l'appropriation qui ne s'agit plus des territoires et des ressources matérielles mais des données. Cela se fait par un nouveau type de relations sociales : les relations de données<sup>25</sup>. Il est généralement utilisé pour expliquer « la nature extractive et exploitante de la relation entre les entreprises de technologie et les peuples, déployée à des fins politiques diverses »<sup>26</sup>.

Dans le contexte du droit international, la souveraineté numérique fait référence à « l'indépendance des États dans la mise en œuvre d'une politique numérique »<sup>27</sup>. Le développement d'IA risque de porter atteinte aux deux fondements de la souveraineté que sont l'indépendance et l'égalité des États. Google, Amazon, Facebook, Apple, Microsoft et IBM (ci-après GAFAMI) auxquels s'ajoutent Twitter, Intel et Salesforce, en rejoignant les géants chinois Baidu, Alibaba, Tencent et Xiaomi (ci-après BATX), renforcent une

---

<sup>21</sup> Intelligent Robots Development And Distribution Promotion Act (IRDSPA) de la Corée du Sud, adoptée en 2008, dernière modification en 2016.

<sup>22</sup> H.R.4625 - Fundamentally Understanding the Usability and Realistic Evolution of Artificial Intelligence Act (FUTURE of Artificial Intelligence Act of 2017), 115th Congress.

<sup>23</sup> Par des expériences indispensables de leurs consommateurs, les géants du numérique colonisent et contrôlent des informations, et opèrent une transformation de l'expérience humaine en « données comportementales ». ZUBOFF S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, 2019.

<sup>24</sup> CRAWFORD K., VLADAN J., *Anatomy of an AI System*, AI Now Institute, NYU, 2018, p. 14.

<sup>25</sup> COULDRY N., MAJIAS U., « Data colonialism: rethinking big data's relation to the contemporary subject », LSE Research Online, 2018, p. 1.

<sup>26</sup> CRAWFORD K et alii, « AI Now 2019 Report », AI Now Institute, New York University, déc. 2019, p. 43.

<sup>27</sup> ACHILLEAS PH., « Intelligence artificielle et droit international », in BENSAMOUN A., LOISEAU G., (dir.), *Droit de l'Intelligence Artificielle*, LGDJ, 2019, p. 411.

colonisation numérique<sup>28</sup>. Ainsi, pour certains, s'il est devenu courant d'appeler les données le nouveau pétrole, la naissance d'un droit de propriété des données<sup>29</sup> ou res communis, res nullius des données<sup>30</sup>, un tel contexte du droit de la propriété constituerait une protection juridique encore plus cohérente que la perspective du droit de la responsabilité délictuelle du Règlement général sur la protection des données (ci-après RGPD).

**5. À la cybercolonisation et silicolonisation<sup>31</sup> de l'humanité.** — Au fil des années, les ordinateurs ne font pas seulement plus de travail dans le monde, ils prennent plus de décisions. Comme l'illustre bien le livre de M. Pedro Domingos, « le moment où nous pouvions éteindre tous nos ordinateurs sans provoquer l'effondrement de la civilisation moderne est depuis longtemps révolu »<sup>32</sup>. Tirant ce résultat, les géants du numérique concurrencent déjà les États<sup>33</sup> et pourraient finir par faire le travail pour les gouvernements<sup>34</sup>.

En effet, la Silicon Valley, ou plus généralement, la communauté technologique dépendra de moins en moins d'autres personnes juridiques, soit l'État, les personnes morales d'autres secteurs, et bien entendu, les personnes physiques. L'on peut déjà citer l'exemple d'Amazon qui se trouve non seulement sur le chemin de l'autosuffisance<sup>35</sup>, mais même joue un certain rôle d'intervention étatique<sup>36</sup> ; et le cas de Meta, qui offre des services allant de la plateforme sociale à la réalité augmentée et virtuelle. En revanche, les individus et les gouvernements étatiques dépendent de plus en plus de ces technologies<sup>37</sup>. L'idée de « pouvoirs privés économiques » qui accentue le fait que « de simples personnes privées disposent d'un

---

<sup>28</sup> *Ibidem.*, pp. 409-410.

<sup>29</sup> RITTER J., MAYER A., « Regulating data as property: a new construct for moving forward », 16 *Duke Law & Tech. Rev.*, 2018, pp. 220-277.

<sup>30</sup> DETERMANN L., « No One Owns Data », UC Hastings Research Paper n° 265, 14 févr. 2018.

<sup>31</sup> SADIN É., *La Silicolonisation du Monde : L'irrésistible expansion du libéralisme numérique*, L'échappée, 2016.

<sup>32</sup> DOMINGOS P., *The Master Algorithm*, Basic Books, 2015, p. 282.

<sup>33</sup> « Les GAFAM concurrencent déjà les États, dénonce Vladimir Poutine », *Capital*, 27 janv. 2021. Apple, Microsoft, Amazon, Google et Facebook (GAFAM) ont une capitalisation du marché combinée de 3,7 billions de dollars, soit le produit intérieur brut de l'Allemagne en 2017.

<sup>34</sup> FAULCONBRIDGE G., SANDLE P., « Father of Web says tech giants may have to be split up », *Reuters*, 1 nov. 2018.

<sup>35</sup> METZ C., « Amazon's Homegrown Chips Threaten Silicon Valley Giant Intel », *NYT*, 10 déc. 2018.

<sup>36</sup> L'exemple quant à ce transfert de facto de la subvention étatique à l'intervention "siliconvallique", v. par ex. WEISE K., « Microsoft Pledges \$500 Million for Affordable Housing in Seattle Area », *NYT*, 16 janv. 2019.

<sup>37</sup> V. p. ex. la Défense : JAESA « Microsoft a battu Amazon pour le contrat de \$10 milliards du Pentagone », *Transhumanisme et IA*, 29 oct. 2019.

pouvoir de décision unilatéral analogue sur le plan matériel à celui de la puissance publique »<sup>38</sup> devient de plus en plus concrète. Par conséquent, c'est à l'affirmation de cette réalité que le gouvernement danois nomme désormais un ambassadeur numérique auprès des multinationales de la Silicon Valley<sup>39</sup>.

**6. La nature distincte entre l'IA étroite et l'IA générale.** — Dans cette numérisation de la civilisation, le progrès de l'IA se trouve au cœur de cette évolution. Pourtant, même si l'on peut ajouter romancier<sup>40</sup> à la longue liste de tâches pour lesquelles l'IA surpassera les humains, et ainsi redéfinir la créativité<sup>41</sup>, par exemple une œuvre générée par voie d'IA a été vendue aux enchères à plus de 432,500 dollars<sup>42</sup>, pour certains, c'est trompeur. Car bien que les artistes-codeurs discutent toujours la question de savoir dans quelle mesure les algorithmes sont des outils, par opposition aux collaborateurs actifs<sup>43</sup>, ils ne sont certainement pas, à ce stade, des agents autonomes au sens juridique. Certes, l'IA copie à la perfection<sup>44</sup>, enseigne<sup>45</sup>, invente et crée<sup>46</sup>, mais malgré son progrès impressionnant<sup>47</sup>, voire capable de surpasser les humains<sup>48</sup>, en l'absence d'une capacité générale comparable aux humains, ces systèmes demeurent un simple outil à nos yeux.

---

<sup>38</sup> FARJAT G., *Droit économique*, PUF, 2<sup>e</sup> éd, 1982, p. 613.

<sup>39</sup> CARQUILLAT J.-C., « Bientôt une ambassade de France auprès des GAFAs ? », *Le Portail*, 4 avr. 2018.

<sup>40</sup> *Shapeshifting: The Dream of Dagrec's Legacy, Beautiful Incoherence, Gregor the Cockroach: Metamorphosis retold by A.I.*, sont les trois livres publiés en 2019, écrits par IA GPT-2 d'OpenAI, mais édités et illustrés par les êtres humains.

<sup>41</sup> STREITFELD D., « Computer Stories: A.I. Is Beginning to Assist Novelists », *NYT*, 18 oct. 2018.

<sup>42</sup> SERMONDADAZ S., « Un tableau peint par une IA vendu 432.500 dollars... à partir d'un code source emprunté sur le web », 26 oct. 2018, *Science et Avenir*.

<sup>43</sup> L'exemple d'un artiste qui utilise le robot dénommé sozo pour aider à peindre des tableaux qui peuvent se vendre plus de 100,000 dollars dans les galeries de New York. Ou le robot Kuka qui interprète un duo de danse homme-machine avec danseur humain. « Robot Paints Microscopic Brushstrokes For U.S. Artist's Canvases Worth Over \$100,000 », *FeelGuide*, 2 juin 2018; A human-robot dance duet | Huang Yi & KUKA.

<sup>44</sup> KEEGAN R., « A.I. Robot Cast in Lead Role of \$70M Sci-Fi Film », *The Hollywood Reporter*, 24 juin 2020.

<sup>45</sup> FRIDMAN L., « Yoshua Bengio : Deep Learning », 20 oct. 2018.

<sup>46</sup> « Artistes et Robots au Grand-Palais : l'intelligence artificielle produit-elle de l'art ? », *franceinfo Culture*, 7 avr. 2018.

<sup>47</sup> Tel est le cas d'un système de reconnaissance vocale formé sur 11 940 heures de parole avec des transcriptions alignées, où il faudrait 16 ans pour les humains. GANGULI S., « The intertwined quest for understanding biological intelligence and creating artificial intelligence », *Stanford's Human-Centered AI Initiative (HAI)*, 5 déc. 2018. V. aussi LLOYD D., « Machine learning has been used to automatically translate long-lost languages », *MIT Tech. Rev.*, 1 juill. 2019; LAROUSSERIE D., « Intelligence artificielle : dorénavant, les machines maîtrisent la grammaire française », *Le Monde*, 18 nov. 2019.

<sup>48</sup> P. ex., JONES B., « Machines Just Beat Humans on a Stanford Reading Comprehension Test », *Futurism*, 15 janv. 2018; LANDGRAF M., Karlsruhe Institute of Technology, « AI outperforms humans in speech recognition », *Techxplore*, 20 oct. 2020.

À juste titre, il convient de distinguer entre l'intelligence artificielle étroite (Artificial Narrow Intelligence), qui concerne des domaines d'application spécifiques, et l'intelligence artificielle générale (Artificial General Intelligence, ci-après IAG). Cette dernière se réfère à un futur système d'IA capable d'effectuer une grande variété de tâches cognitives au niveau humain. Autrement dit, ces systèmes qui disposent de la capacité de concevoir ses propres tâches apprennent alors à poser toute sorte de questions, et pas uniquement à répondre à celles que leurs posent les humains. Partant, la différence entre l'IA étroite et l'IA générale est que cette dernière consiste à apprendre par lui-même et ne se substitue pas seulement aux tâches étroitement spécifiques, mais aussi à des métiers de plus en plus qualifiés dont l'on pensait que la technicité et la dimension relationnelle relevaient de l'humain<sup>49</sup>.

**7. La solution pour l'IA étroite : « Le pouvoir arrête le pouvoir »<sup>50</sup>.** — À la lumière de cette distinction fondamentale, l'on constate que les difficultés autour de l'IA générale ne concernent plus véritablement les préoccupations d'un système d'IA étroite : l'usage militaire<sup>51</sup>, l'aspect économique et financier<sup>52</sup>, la politique d'emploi<sup>53</sup>, les véhicules autonomes<sup>54</sup>, le droit social<sup>55</sup>, les données personnelles et le risque réel d'érosion de la vie

---

<sup>49</sup> LAURENT A., *La guerre des intelligences : Comment l'intelligence artificielle va révolutionner l'éducation*, Lattès, 2017, pp. 48-49.

<sup>50</sup> MONTESQUIEU, *De l'esprit des Lois*, 1748, éd. Nourse, 1767, t. 1, chap. IV, livre XI, p. 206.

<sup>51</sup> Cf. SCHARRE P., *Army of None: Autonomous Weapons and the Future of War*, W. W. Norton & Company, 2018.

<sup>52</sup> P. ex., en France, l'article 57 du projet de loi de finances 2020, adopté en 2019, prévoit l'usage de l'IA pour mieux cibler les contrôles afin de renforcer la lutte contre les fraudes fiscales et douanières.

<sup>53</sup> La moitié de toutes les activités de travail actuelles peuvent être automatisées. C'est un fait indéniable que le déplacement d'emplois sera l'un des principaux défis de la prochaine décennie. V. Testimony préc. de Mme Fei Fei Li. Et en 2017, M. Laurent Alexandre, dans une intervention qu'il faisait devant le Sénat français, a dit que « dans le futur, tous les gens qui ne seront pas complémentaires de l'intelligence artificielle seront soit au chômage, soit avec un emploi aidé. Pas 99 % des gens. 100 %. Aucun emploi non complémentaire de l'IA n'existera en 2050 par définition ». D'aucuns prétendent qu'il y aura toujours de nouveaux emplois pour les humains, comme ceux qui sont apparus après la révolution industrielle précédente. Mais comme M. Stuart Russell l'a noté : « quand une machine remplace le travail physique, on peut vendre du travail mental. Quand une machine remplace le travail mental, que reste-t-il à vendre ? » RUSSELL S., *Human Compatible*, Allen Lane, 2019, p. 114.

<sup>54</sup> Une classification d'autonomie est définie par la SAE (Society of Automotive Engineers) : de la voiture entièrement prise en charge par son conducteur jusqu'à la voiture entièrement autonome, les véhicules sont généralement classés en 6 niveaux d'autonomie : aucune automatisation, aide à la conduite, automatisation partielle, automatisation conditionnelle, automatisation élevée, automatisation complète. V. aussi « Automatisiertes Und Vernetztes Fahren », Ethik-Kommission, Bundesministerium für Verkehr und digitale Infrastruktur, Bericht Juni 2017, p. 14.

<sup>55</sup> Dans la collecte de données via des plateformes telles que Facebook, les utilisateurs alimentent et forment les réseaux de neurones avec des données, le nouveau type de travail invisible ou externalisé, caché derrière des interfaces et camouflé dans des processus algorithmiques est maintenant très courant, en particulier lors de l'étiquetage de milliers d'heures d'archives numériques dans le but d'alimenter les réseaux de neurones. Parfois, ce travail est entièrement non rémunéré, par exemple, pour prouver que l'on n'est pas un agent

privée<sup>56</sup>, la décision biaisée<sup>57</sup>, la propriété intellectuelle<sup>58</sup> et d'autres dimensions appartenant à une liste non exhaustive. Certes, les technologies avancées posent des défis dans chaque société de toute époque, en vertu de l'adage « ubi societas ibi jus », comme le développement technologique nous permet également de faire face à des situations spontanées, à un certain degré, il ne nous reste qu'un seul choix dans le contexte d'IA étroite : de lutter contre l'IA avec l'IA<sup>59</sup>.

**8. Problématique.** — Si les enjeux de l'IA étroite posent déjà une réalité concernant les approches pluridimensionnelles telles que la sécurité compromise et la vérité entravée<sup>60</sup>, l'arrivée de l'IA générale, tranche, sans ambiguïté, le sens et la place des humains en tant qu'espèce. En outre, l'on pourrait s'attendre qu'une telle IA avancée serait dotée d'une sorte de conscience ou d'une conscience de soi, ce qui signifie l'arrivée de l'IA forte (Strong AI), par rapport à l'IA faible (Weak AI), l'IA qui manque de conscience. La notion de l'IA forte

---

artificiel, l'on est obligé de former gratuitement le système d'IA de reconnaissance d'image de Google en sélectionnant plusieurs images contenant des numéros de rue, des voitures ou des maisons. V. CRAWFORD K., VLADAN J., *Anatomy of an AI System*, *op. cit.*, p. 14.

<sup>56</sup> V. le RGPD. Une surveillance accrue telle que Big Brother de 1984 de George Orwell devient inquiétant.

<sup>57</sup> M. Erik Brynjolfsson exprime ce partage entre l'homme et la machine : « Machine Learning is probably biased. But less so than you ». Par ailleurs, l'on peut également constater la discrimination individuelle et le rôle évolutif du syndicat : au lieu d'une discrimination collective, au XXI<sup>ème</sup> siècle, l'on aurait un problème de discrimination individuelle. V. HARARI Y. N., *21 Lessons for the 21st Century*, Jonathan Cape London, 2018, p. 68; DOMINGOS P., *The Master Algorithm*, *op. cit.*, p. 275.

<sup>58</sup> Des premières demandes de brevet pour des inventions générées uniquement par l'intelligence artificielle ont été déposées en 2019. Le système, dénommé « DABUS », a produit des résultats qui ont servi de base à deux demandes de brevet. L'Office de la propriété intellectuelle du Royaume-Uni (IPO), le Bureau américain des brevets et des marques de commerce (USPTO), et l'Office européen des brevets (EPO) ont constaté que DABUS n'est pas une personne et ne peut donc pas être considéré comme l'inventeur d'un brevet. L'office de brevets de l'Afrique du Sud ainsi que la Cour fédérale australienne retiennent une décision contraire. D'autres aspects de la propriété intellectuelle, v. p. ex., JANG E., *AI is Good for You*, auto-édition, 2023.

<sup>59</sup> L'exemple d'un logiciel d'IBM a été conçu pour gérer les déploiements d'IA, détecter les biais dans les modèles, atténuer son impact et surveiller les décisions. V. DIGNAN L., « IBM launches tools to detect AI fairness, bias and open sources some code », ZDNet, 19 sept. 2018; WIGGERS K., « MIT CSAIL researchers propose automated method for debiasing AI algorithms », VentureBeat, 27 janv. 2019; COLE S., « Researchers Created AI That Hides Your Emotions From Other AI », VICE, 23 août 2019; « Google makes deepfakes to fight deepfakes », BBC, 26 sept. 2019; BACK E., « Microsoft lance un outil pour débusquer les deepfakes », FUTURA, 5 sept. 2020; METZ C., « Internet Companies Prepare to Fight the 'Deepfake' Future », NYT, 24 nov. 2019. Il y a donc d'abord eu l'IA qui prétendait détecter les cas frauduleux et nos émotions. Puis vint l'IA qui cache nos émotions et la détection à l'IA. Puis vint l'IA qui révèle l'IA masquante. Alors.....

<sup>60</sup> Les enjeux peuvent être rencontrés à tout stade du développement : l'IA pourrait remettre en question l'aspect « voir, c'est croire » des preuves vidéo et audio. Ils pourraient aussi aider les gens à nier les accusations portées contre eux. Les régimes autoritaires en particulier peuvent tirer parti d'un paysage de l'information où la vérité objective devient dévaluée et où la « vérité » est ce que les autorités prétendent qu'elle est. AMODEI D. et alii, *The Malicious Use of Artificial Intelligence - Forecasting, Prevention, and Mitigation*, 26 authors from 14 institutions, févr. 2018, p. 46.

désigne donc la position philosophique selon laquelle l'intelligence artificielle ne se contente pas de simuler un esprit ; elle a littéralement un esprit.

Ainsi, l'IA générale voire l'IA forte nous oblige à répondre à des questions beaucoup plus fondamentales et essentielles — la frontière de l'homme, a-t-elle été bouleversée? Quelle est la place de notre espèce? Comment déterminer le sujet juridique dans la conception traditionnelle? Les catégories juridiques des êtres sont-elles encore valables? Que signifie le sens juridique quand tout le monde ne vit plus dans une même réalité? Comment la configuration juridique peut-elle se reconnaître dans une hiérarchie de l'intelligence qui se trouve dans un monde constitué de plusieurs réalités? Si ce sont des questions qui ne connaissent pas les frontières dites nationales, elles ne peuvent plus être réglées par des ordres internes, par un souverain dit étatique. C'est en particulier lorsque le fossé flagrant des pouvoirs étatiques s'amplifie, les pays moins développés en ces matières dépendent largement voire entièrement de ces technologies fournies par les États les plus dominants.

**9. L'intérêt du sujet : le vrai souci de l'IA forte - vers un conflit entre différentes espèces plutôt qu'entre des nations.** — L'intérêt philosophique de la recherche est donc de reconnaître la présence de la civilisation humaine dans la hiérarchie d'intelligence. Quant à son intérêt économique, vu que l'intelligence sur le plan juridique reflète essentiellement la capacité juridique, sa reconstruction est susceptible de redéfinir les catégories juridiques et ainsi d'aménager une économie in toto composée des différentes unités économiques. L'intérêt historique est alors de revoir la nature des relations juridiques évolutives dans des situations de l'intelligence artificielle avancée. Enfin, du point de vue de la sécurité, étant donné que l'IA générale implique des systèmes en mesure d'accomplir n'importe quelle tâche effectuée par un être humain, voire mieux, cette nouvelle forme de vie rendra la nôtre obsolète. Comme ces êtres totalement autonomes qui poursuivront leurs propres objectifs deviendront plus intelligents que les êtres humains, il sera crucial que leurs objectifs soient en adéquation avec les nôtres.

Dans cet acte de création d'une IA générale, voire forte, car dépassant les capacités humaines dans presque tous les domaines, l'on céderait notre statut d'entités les plus intelligentes de la Terre. Vu que ces développements se trouvent entièrement dans les mains d'un petit nombre d'acteurs mettant en place ces systèmes, une réglementation est nécessaire pour

intégrer le processus de développement du statut des êtres et les garanties contre le « conflit entre différentes espèces plutôt qu'entre des nations »<sup>61</sup>. Car le cœur de l'IAG est d'apprécier suffisamment le pouvoir de l'intelligence pour se rendre compte qu'en se trompant sur une intelligence surhumaine au premier essai, l'on ne serait pas là pour apprendre et réessayer pour un second essai<sup>62</sup>. Il faut ainsi redéfinir les concepts juridiques afin de reconstruire un nouveau monde juridique cohabité par l'IAG et les êtres humains.

**10. Délimitation et objectif : de *lege lata* à *lege ferenda*.** — À ce stade, l'on constate qu'il y a un ensemble de problèmes qui concerne le court terme : les défis immédiats ou imminents impliquant des acteurs et des paramètres relativement clairs, tels que la confidentialité, la responsabilité, les biais ; et une deuxième série de problèmes concernant les préoccupations à long terme, telles que la perte du sens humain à grande échelle, le risque que l'IA surhumaine pouvant la mettre hors du contrôle de l'homme, et des questions fondamentales sur la place de l'humanité dans un monde avec ces nouvelles compositions intelligentes. La cible de ce travail se focalisera sur cette deuxième série de notions sur le plan du droit extra-étatique, afin de dégager son implication pour le droit international ; c'est ce que ce travail a pour vocation d'explorer comme pistes de réflexion.

Cela signifie que la discussion de l'IA de ce travail se concentrera sur la notion nébuleuse de l'IA générale et de l'IA forte, ce qui englobe à la fois sur l'idée de l'intelligence anthropocentrique et le risque de l'intelligence universelle. En effet, étant donné qu'une vision anthropocentrique est éventuellement inéluctable au regard de la mesure pratique de l'IA générale, les humanoïdes, malgré leur appartenance actuelle de la catégorie d'IA étroite et d'IA faible, et des androïdes, convenablement avancés dont le but principal est d'établir voire maintenir une relation sociale avec les êtres humains, ces perspectives de sa fonction relationnelle seront lato sensu abordées dans ce travail. À l'inverse, en présence de nuances qui ne peuvent être négligées<sup>63</sup>, comme c'est l'IA générale qui nous intéresse essentiellement dans notre étude, les préoccupations de l'IA étroite ne seront pas par conséquent, traitées dans ce travail. Dans cette étude, les termes « IA forte » et « IAG » désignent de manière

---

<sup>61</sup> Expression de LAWRENCE D. R., BRAZIER M., « Legally Human? 'Novel Beings' and English Law », *Medical Law Review*, vol. 26, Issue 2, 1 mai 2018, p. 326.

<sup>62</sup> Cf. p. ex., Geoffrey Hinton.

<sup>63</sup> CAVE S., ÓHÉIGEARTAIGH S. S., « Bridging near- and long-term concerns about AI », *Nature Machine Intelligence*, 7 janv. 2019.

interchangeable la notion de l'IA générale évoquée ci-dessus, c'est-à-dire l'IA de niveau humain ou surhumain voire dotée d'une conscience, pour la distinguer de l'IA étroite, c'est-à-dire l'IA outil qui manque le facteur général dans sa capacité.

Enfin, de nombreux chercheurs en IA s'attendent à ce que l'on construise des IAG au cours de ce siècle. Il convient à ce stade de noter que cette étude n'explorera pas les arguments concernant le calendrier de développement de l'IAG. De même, la présente étude n'a pas pour vocation de juger si l'on peut ou non construire des IAG, ni de fournir des critiques concernant la manière pour sa création. Lorsque l'on touche certains aspects de ces questions, cette étude cherche à donner plus de poids aux voix des chercheurs qui travaillent « actuellement » sur l'IAG, soit les participations les plus éminentes au regard de la situation actuelle. Ainsi, partant de cette vision à long terme de l'IAG, cette étude a simplement pour but de présenter une expérience de réflexion autour du contexte général de l'IAG, tout en se bornant à son rapport avec le concept du droit international. Sur ce fondement, l'objectif de cette thèse cherche à étudier la transformation juridique afin d'envisager une coévolution pacifique entre l'espèce humaine et l'IAG.

**11. L'approche transversale et intégrale : de la source bibliographique marginalisée à l'espèce marginalisée.** — L'on pourrait également remarquer que, dans les pages suivantes, les sources françaises trouvent une place relativement moins déterminante. Car aujourd'hui, l'Europe n'est pas indépendante technologiquement vu la transformation de la présence numérique à la souveraineté numérique nationale ou continentale. Du fait que l'économie européenne repose sur des infrastructures technologiques des pays tiers, il n'y a pas d'indépendance technologique ni de souveraineté technologique européenne<sup>64</sup>. Force est de constater qu'en matière d'IA, l'Europe sera distancée économiquement et technologiquement par l'Asie et les États-Unis<sup>65</sup> dans les prochaines années. D'une manière générale, ce retard voire manquement du développement particulièrement de l'IAG par certains côtés du monde se reflète par conséquent dans nos discussions bibliographiques<sup>66</sup>.

---

<sup>64</sup> Marie Ekeland discute de l'IA avec Cédric Villani, Sciences et Avenir, 2018.

<sup>65</sup> « Deux puissances font la course en tête, les États-Unis et la Chine. L'Europe et quelques autres pays suivent, largement distancés ». KAUFFMANN S., « Géopolitique de l'Intelligence artificielle », Le Monde, 13 nov. 2019.

<sup>66</sup> V. p. ex., d'après la formule de Mme Marion Ho-Dac : « Il est évident que l'Union devra s'appuyer sur des travaux de normalisation non européens parce que certaines normes, ou publications connexes à l'instar de spécifications techniques, ont déjà été élaborées ou sont en cours de développement à l'échelle mondiale. Il

Néanmoins, cette marginalisation bibliographique de certains pays n'empêche pas, en tout état de cause, le sens de cette étude que porte sur notre sujet qui concerne, avant tout, une IA à l'échelle cosmique et un droit au sens d'espèces. Autrement dit, en se fondant sur ces critères véritablement pertinents, l'on ne discrimine pas les sources bibliographiques en fonction d'une langue ou d'une autre ; au contraire, ce que l'on cherche à étudier ici, c'est justement un sujet qui nous demande de tenir compte de son caractère transversal. Cette compréhension nous amène donc à adopter une approche non discriminatoire de notre référence bibliographique, pour qu'une approche intégrale de notre sujet d'étude se justifie.

**12. Annonce du plan.** — En l'occurrence, la civilisation humaine repose sur des systèmes juridiques développés depuis plusieurs siècles. Notre espèce a construit ce monde juridique nous distinguant des autres espèces. En effet, cet outil de compréhension d'organisation – le droit – se base aussi sur l'intelligence de notre espèce. Par conséquent, une intelligence surhumaine ne nous laisse qu'un seul choix : « L'on doit aujourd'hui accueillir, et même étudier, toute tentative sérieuse d'envisager l'avenir de notre espèce; non seulement pour saisir les possibilités très diverses et souvent tragiques qui se présentent à nous, mais aussi pour nous familiariser avec la certitude que nombre de nos idéaux les plus chers sembleraient puérils à des esprits plus développés »<sup>67</sup>. Dans cette démarche, la première partie de cette étude tente d'illustrer la place de l'intelligence humaine dans la nouvelle hiérarchie des normes ainsi que l'évolution juridique autour de l'intelligence artificielle générale (IAG) (partie 1). La seconde partie se consacre à une configuration juridique fondée sur le principe de faveur avec un répertoire du droit d'IAG (partie 2).

— Partie 1 : La nouvelle hiérarchie de l'intelligence

— Partie 2 : La nouvelle hiérarchie des normes

---

s'agit à la fois d'une exigence d'uniformité mondiale de l'essence même de l'activité de normalisation et d'un principe de réalité, tant la tâche de normaliser le développement des systèmes d'IA et leur cycle de vie est titanesque et hautement complexe ». HO-DAC M., « La normalisation, clé de voûte de la réglementation européenne de l'intelligence artificielle (AI Act) », *D.*, 2023, p. 228.

<sup>67</sup> STAPLEDON O., *Last and First Men: A Story of the Near and Far Future*, 1930, pp. 2-3 : « *Et même dans notre génération, les circonstances pourraient bien changer de manière si inattendue et si radicale que ce livre pourrait très vite paraître ridicule. Mais peu importe. On doit aujourd'hui concevoir notre relation avec le reste de l'univers du mieux que l'on peut; et même si nos images doivent sembler fantastiques aux futurs hommes, elles peuvent néanmoins servir leur but aujourd'hui* ».



## Première partie : La nouvelle hiérarchie de l'intelligence

---

**13. Annonce du plan.** — Dans la science traditionnelle, la ligne de démarcation entre sujet et objet est généralement considérée comme sacrée. Mais l'intelligence artificielle dissout souvent la ligne de démarcation entre la réflexion subjective et la réflexion objective. Ce phénomène résulte en effet du changement de paradigme entre différentes capacités de l'intelligence artificielle.

Dans une telle transition dite intelligente, la question se pose de savoir comment appréhender, du point de vue juridique, ces évolutions ainsi que ses conséquences. Partant de ces interrogations, les développements contenus dans cette partie ont pour but de vérifier le bien-fondé de l'installation de l'intelligence artificielle au sein de la civilisation, où elle commence par une conception mécanique objective (titre 1), et finit éventuellement par son effet subjectif (titre 2).

## TITRE 1 L'ÉVOLUTION OBJECTIVE DE L'INTELLIGENCE

**14. Annonce du plan.** — Les machines, ou plus généralement, les progrès de la technologie, ne constituent rien de nouveau au cours de l'Histoire de la civilisation. Dès lors, cette étude cherche à démontrer pourquoi l'intelligence artificielle avancée se trouve soumise à une nouvelle particularité mérite d'être exceptionnellement examinée sous l'angle juridique. Afin de s'interroger sur une telle condition, il convient d'illustrer l'effet de l'intelligence artificielle générale voire forte (ci-après IAG) (chapitre 2). Pour ce faire, un aperçu introductif de l'IAG (chapitre 1) fait l'objet de notre étude préliminaire.

## Chapitre 1 L'introduction de l'IAG

« *AI is the science of making machines smart.* » — Demis Hassabis

**15. Annonce du plan.** — De nos jours, le développement voire l'expression même de l'intelligence artificielle continue de recevoir des controverses. Mais au fond, qu'est-ce que l'IAG cherche et quels sont les risques ? Pour aborder l'essence de la réponse, il convient de se pencher sur la notion d'IAG (section 1) avant d'exposer les enjeux de l'IAG (section 2).

### Section 1 La notion d'IAG

« *Artificial Intelligence began with an ancient wish to forge the gods.* »

— Pamela McCorduck

« *AI will save us from ourselves.* » — Demis Hassabis

« *AI is much bigger than deep learning, bigger than computing, it is our civilization journey into understanding the human mind and creating echos of it in the machines.* » —

Lex Fridman

**16. Annonce du plan.** — Dans cette section, l'on va voir d'abord l'idée (§1) et puis les caractéristiques (§2) de l'IAG.

### §1 Présentation d'IAG

**17. La définition de l'intelligence artificielle.** — L'intelligence artificielle, c'est « agir et penser humainement et rationnellement »<sup>68</sup>. Cette idée selon laquelle les systèmes informatiques peuvent remplir des fonctions généralement associées à l'esprit humain, est passée de la spéculation futuriste à la réalité contemporaine. Rappelons que le mot « ordinateur » en chinois (caractères traditionnels) : « 電腦 », est composé de deux mots

---

<sup>68</sup> RUSSELL S., NORVIG P., *Artificial Intelligence - A Modern Approach*, 3rd éd., Pearson Education, 2009, p. 2.

signifiant littéralement « cerveau électronique ». L'intelligence artificielle est donc la science qui consiste à faire faire par des machines des choses qui exigeraient de l'intelligence si elles étaient faites par des hommes<sup>69</sup>. Elle est aussi, d'après Feynman, la technologie qui permet de « développer des machines plus intelligentes, des machines qui ont une meilleure relation avec les humains afin que l'entrée et la sortie puissent être faites avec moins d'effort que la programmation complexe qui est nécessaire d'aujourd'hui. Cela va souvent sous le nom d'intelligence artificielle, mais je n'aime pas ce nom » a-t-il précisé, vu que « Peut-être les machines inintelligentes peuvent faire encore mieux que celles qui sont intelligentes »<sup>70</sup>.

**18. L'illustration de l'intelligence artificielle générale.** — L'on peut centrer la définition de l'intelligence artificielle générale (IAG) d'après M. Stuart Russell comme un « système intelligent polyvalent (general-purpose intelligent system) » qui peut, par supposition, faire ce que n'importe quel humain peut faire<sup>71</sup>. Et selon M. Yann Le Cun, l'on arrivera à « l'intelligence artificielle générale, c'est-à-dire des machines égales à l'homme, aussi intelligentes que lui »<sup>72</sup>. C'est ainsi que Nils John Nilsson décrivait que « la réalisation d'une véritable intelligence artificielle de niveau humain impliquerait nécessairement l'automatisation de la plupart des tâches que les êtres humains accomplissent pour rémunération »<sup>73</sup>. Dans ce sens, plutôt que de travailler sur cet objectif d'automatisation en construisant des systèmes à des usages spécifiques, il proposait le développement de « systèmes polyvalents et éducatifs qui peuvent apprendre et être enseignés pour effectuer n'importe lequel des milliers de travaux que les humains peuvent effectuer »<sup>74</sup>.

---

<sup>69</sup> La définition de Marvin Minsky citée in WHITBY B., *Reflections on Artificial Intelligence*, Intellect Books, 1996, p. 20. V. aussi ISO/IEC 2382 : 2015, la norme sur la terminologie de l'informatique en donne sa définition : la « capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage ». L'on peut en outre imaginer que l'IA essaie de faire mieux des choses que les hommes font déjà moins bien que la machine, ce qui ne rentrerait pas dans la définition ci-dessus. Ainsi, elle peut également être définie comme « des systèmes capables de prises de décisions autonomes ou en interaction, dans un contexte riche et varié, voire nouveau » ; ou des techniques consistant à « parvenir à faire faire aux machines, ce que l'homme fait aujourd'hui mieux qu'elles, notamment s'adapter, apprendre, communiquer et interagir d'une manière riche et variée avec leur environnement ». V. « Livre blanc sur l'impact des attentes pour la normalisation dans l'intelligence artificielle – version 7 », Information et Communication Numériques, Comité Stratégique, 16 févr. 2018, p. 3 ; « France IA : la stratégie IA en France », le portail de l'économie, des finances, de l'action et des comptes publics, 21 mars 2017.

<sup>70</sup> FEYNMAN R. P., « The Computing Machines in the Future », Nishina Memorial Lecture at Gakushuin University, Tokyo, 9 août 1985.

<sup>71</sup> RUSSELL S., *Human Compatible*, *op. cit.*, p. 93.

<sup>72</sup> KLEIN É., « Quatre questions à Yann LeCun », clefs #69, CEA, nov. 2019, p. 45.

<sup>73</sup> NILSSON N. J., « Human-Level Artificial Intelligence? Be Serious! », *AI Magazine*, 26(4), 2005, p. 68.

<sup>74</sup> *Ibidem*.

En se ralliant à ce critère dit d'économie, devant les membres du Congrès des États-Unis, OpenAI entend sa définition de l'IAG comme « des systèmes hautement autonomes qui surpassent les humains dans la plupart des travaux de valeur économique »<sup>75</sup>. Lorsque interrogé sur le terme « la plupart des travaux de valeur économique », M. Greg Brockman, le cofondateur d'OpenAI, a répondu que « la question de savoir comment définir l'IAG n'est pas tout à fait unanime au sein de la communauté, (...) mais lorsque l'on y réfléchit, l'on pense à des choses telles que la création d'entreprise, ou à un travail intellectuel très important, ou à des choses comme le nettoyage des sites sinistrés ou encore à des tâches que les humains seraient incapables de faire très bien aujourd'hui »<sup>76</sup>. Car pour OpenAI, il serait difficile de connaître l'authenticité d'une IAG lorsque l'annonce a été faite ainsi. En revanche, si l'on vit dans un monde où l'IA accomplit la grande partie du travail économiquement valable, il est très probable que l'on pourrait en déduire que l'IAG a été atteint<sup>77</sup>. Cette approche dite économique retenue par OpenAI consiste donc à apprécier l'impact économique ainsi que la capacité de résoudre des problèmes dans le monde réel pour l'identification de l'IAG. Dans cette perspective, l'on peut déjà penser au robot humanoïde Tesla bot, désormais connu sous le nom d'Optimus, qui servira à accomplir des tâches dangereuses, répétitives et ennuyeuses à la place des humains, voire à copier la personnalité de son propriétaire, comme un modèle débutant de l'IAG d'après cette définition.

Une autre approche pour identifier l'IAG, c'est le critère de deux instances proposées par M. Rodney Brooks, à savoir les aides aux personnes âgées (Elder Care Worker, ECW) et les planificateurs des services logistiques (Services Logistics Planner, SLP). Pour lui, ECW et SLP se distinguent en tant qu'agents qui travaillent. Il propose donc que « le premier agent d'IAG serait un robot physiquement incarné qui travaillerait au domicile d'une personne

---

<sup>75</sup> Une définition qui comprend d'une part, la formation d'une entreprise, les transactions commerciales et la rédaction de livres. Une telle technologie devra être intelligente d'une manière différente de nos ordinateurs traditionnels non créatifs. Et d'autre part, la généralité du système signifie qu'il ne serait pas limité aux applications commerciales. Il pourrait également se concevoir comme une aide dans les litiges internationaux, d'urbanisme, et même de législations et de gouvernement pour les pays. Ainsi, plutôt que d'être développé pour un cas d'utilisation spécifique, l'IAG serait développée pour tout un éventail de tâches importantes. V. Testimony de M. Greg Brockman, co-founder of OpenAI, House Science Committee Testimony, Subcommittee on Research and Technology and Subcommittee on Energy Hearing - Artificial Intelligence – With Great Power Comes Great Responsibility, 115th Congress of U.S.A., 26 June 2018.

<sup>76</sup> *Ibidem*.

<sup>77</sup> Ilya Sutskever, l'un des cofondateur d'OpenAI, à l'occasion de « 2022 HAI Spring Conference on Key Advances in Artificial Intelligence », 12 avril 2022.

pour la soigner à mesure qu'elle vieillira. Il ne s'agit pas de compagnie, mais plutôt d'une assistance physique qui permettra à une personne de vivre avec dignité et indépendance à mesure qu'elle vieillit chez elle »<sup>78</sup>. Et « le deuxième agent d'IAG serait un agent auquel l'on pourra attribuer une tâche complexe de planification qui impliquera un lieu de travail, des humains en tant que travailleurs, des machines jouant un rôle critique, et des humains en tant que clients pour obtenir des services sur ce lieu de travail »<sup>79</sup>. Son objectif est de remplacer le soi-disant test de Turing par un critère qui teste l'intelligence d'un système de manière beaucoup plus large<sup>80</sup>.

Partant, soit le critère dit économique adopté par OpenAI, soit le critère ECW-SLP proposé par M. Brooks, bien plus que l'IA étroite, l'IAG permet de comprendre le problème général de l'intelligence et ses diverses solutions potentielles. Ainsi, il est probable que bientôt, sur la base de ce que l'on a appris entre-temps, une catégorisation différente des approches IAG semblera plus adéquate<sup>81</sup>. En effet, une autre façon de voir l'IA, c'est l'expression selon laquelle « l'on sait ce qu'est l'IA quand l'on la voit »<sup>82</sup>. En revanche, il existe aussi un paradoxe en IA qui consiste dans le fait qu'à chaque fois, dès qu'une l'IA résout un problème, l'on ne la considère plus comme « IA ». Par exemple, auparavant, les progrès tels que la conduite autonome, la reconnaissance d'images, la cuisine automatisée, et jouer aux échecs etc. pourraient être considérés par nos ancêtres comme les critères désignant le moment où l'IA serait atteinte. Mais curieusement, l'on voit qu'aujourd'hui, ce paradoxe en IA nous montre précisément que systématiquement, une IA serait disqualifiée d'une « IA »

---

<sup>78</sup> BROOKS R., « [FoR&AI] Steps Toward Super Intelligence II, Beyond the Turing Test », Blog de Rodney Brooks, 15 juill. 2018. Le Parlement européen considère également que « les avancées dans le domaine de la robotique et de l'intelligence artificielle peuvent et devraient, dès l'étape de la conception, préserver la dignité, l'autonomie et l'auto-détermination de la personne humaine, particulièrement dans le domaine des soins et de la compagnie des personnes et dans le contexte des appareils médicaux, de la "réparation" ou de l'amélioration du corps humain ». Résolution du Parlement européen du 16 févr. 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique, pt. O.

<sup>79</sup> BROOKS R., « [FoR&AI] Steps Toward Super Intelligence II, Beyond the Turing Test », préc.

<sup>80</sup> *Ibid.* Car il considère que « de stupides bots de discussion ont passé le test de Turing avec succès. Être efficace en ECW ou en SLP constituera un test beaucoup plus solide et garantira une compétence plus générale que celle requise pour le test de Turing ».

<sup>81</sup> GOERTZEL B., PENNACHIN C., « Contemporary Approaches to Artificial General Intelligence », in GOERTZEL B., PENNACHIN C. (éds.), *Artificial General Intelligence*, Springer-Verlag Berlin Heidelberg, 2007, p. 6.

<sup>82</sup> Donc « n'essayera pas de définir l'intelligence artificielle, car il existe probablement autant de définitions que de types d'intelligence artificielle ». NGUYEN H., « Artificial Intelligence Law is Here, Part One », *Evolve The Law*, 26 juill. 2018. Dans ce sens, on peut penser à l'expression de "*I know it when I see it*", retenue par la Cour suprême des États-Unis prononcée par le juge Stewart dans l'affaire *Jacobellis v. Ohio* en 1964.

chaque fois qu'elle arrive à parachever la condition pour sa qualification d'être une IA<sup>83</sup>. Cette difficulté persiste notamment lorsque l'on entre désormais dans la phase de l'IA large.

**19. Entre l'IA étroite et l'IA générale : la phase transitoire de l'intelligence artificielle dite large (Broad AI).** — Dès que l'on a saisi ce que signifie l'intelligence artificielle, l'on sait qu'elle est classée en fonction de son niveau : l'IA étroite par rapport à l'IA générale a pour but de déterminer si elles s'appliquent à une tâche (étroite) ou de nombreuses tâches (générales). Mais il y a aussi une dimension relative à la compétence. En pratique, pour construire des systèmes d'IA qui résolvent des problèmes complexes, l'on a fini par créer des systèmes d'apprentissage de plus en plus généraux. Depuis, l'on a laissé la machine en apprendre plus par elle-même plutôt que par un humain fournissant des connaissances ou des conseils<sup>84</sup>. Ainsi, il existe une grille d'évaluation pour désigner différents niveaux de l'IA. En d'autres termes, bien que l'IA étroite soit efficace pour apprendre une tâche dans un domaine et que l'IA générale ressemble à un humain dans ses capacités de traitement, « l'IA large (Broad AI) » se situe quelque part entre les deux : elle excelle dans de nombreuses tâches d'un domaine à l'autre en demandant moins de données de formation qu'auparavant<sup>85</sup>. En effet, l'histoire de l'IA a été marquée dans ce spectre, en commençant par les systèmes d'IA symbolique, qui ne présentaient pas de généralisation. Et l'on entre maintenant dans une nouvelle étape, où l'on cherche à créer des systèmes flexibles capables d'une large généralisation, tels que les voitures autonomes, les VPA (assistants personnels virtuels), ou les robots développementaux cognitifs (cognitive developmental robots)<sup>86</sup>. Même si certains auteurs prétendent que l'intelligence étroite ne se situe pas sur un continuum avec l'intelligence générale<sup>87</sup>, aujourd'hui, les machines dépassent déjà les

---

<sup>83</sup> Ce phénomène est connu sous le nom de l'effet IA (AI effect). V. p. ex. MCCORDUCK P., *Machines Who Think*, A K Peters Ltd, 2004, p. 204 : « Cela fait partie de l'histoire du domaine de l'intelligence artificielle que chaque fois que quelqu'un découvre comment faire faire quelque chose à un ordinateur - bien jouer aux dames, résoudre des problèmes simples mais relativement informels, un certain nombre de critiques disent 'ce n'est pas de la pensée' ». C'est pourquoi Douglas Hofstadter définit l'IA comme « ce qui n'a pas encore été fait » car « dès qu'une fonction mentale est programmée, les gens cessent sous peu de la considérer comme un ingrédient essentiel de la 'vraie pensée' ». HOFSTADTER D. R., *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic, 1999, p. 601.

<sup>84</sup> Testimony préc. de M. Greg Brockman.

<sup>85</sup> DELISA COLEMAN L., « CES 2019: Why IBM CEO Ginni Rometty's Opening Address Delivers », *Forbes*, 9 janv. 2019. V. aussi CAVE S., ÓHÉIGEARTAIGH S. S., « Bridging near- and long-term concerns about AI », préc.

<sup>86</sup> CHOLLET F., « The Measure of Intelligence », arXiv., 5 nov. 2019, p. 12.

<sup>87</sup> P. ex., v. DREYFUS H. L., « A history of first step fallacies », *Minds and Machines*, 22(2):87-99, 2012.

capacités humaines dans de nombreux domaines, et ces domaines vont s'élargir et s'approfondir. En tout état de cause, ces systèmes « partiellement superintelligents » vont poser bon nombre des mêmes problèmes qu'un système généralement intelligent<sup>88</sup>. En effet, il convient à ce stade de rappeler que même la capacité de l'espèce humaine est aussi une intelligence dite partielle.

**20. La notion de généralité dans l'IAG : de la conception anthropocentrique à l'intelligence universelle.** — La théorie de Gardner sur les intelligences multiples soutient que l'intelligence humaine conteste un certain nombre de composants spécialisés en intelligence et qu'une grande partie de l'intelligence humaine se concentre sur des situations produites par des expériences évolutives : interaction sociale, traitement des images, contrôle de mouvement, etc.<sup>89</sup>. C'est donc à vrai dire, l'intelligence humaine n'est pas vraiment « générale ». Le terme « générale » utilisé dans l'IA générale pour une intelligence comparable à celle des humains implique que l'intelligence humaine est « générale », mais en réalité, l'intelligence humaine est étroitement spécialisée car l'on ne peut pas imaginer des tâches qui sont en dehors de notre compréhension. Si l'on pense que celle-ci est générale, c'est parce que l'on s'adapte à toutes les choses que l'on peut appréhender, sans tenir compte d'un monde immense de choses dont l'on n'a aucune idée<sup>90</sup>.

Ainsi, pour M. François Chollet, le terme « universalité » est plus convenable que « généralité », car il représente une acception qui va au-delà des domaines de tâches pertinentes pour humain, à savoir toute tâche qui pourrait être abordée de manière pratique dans notre univers. Pour autant, d'après lui, pour mieux cerner et concrétiser la notion d'intelligence afin d'être plus précis, la rendre explicative et utile, une approche du biais anthropocentrique est nécessaire. Vu que l'idée même d'une mesure « valable » de l'intelligence n'a de sens que dans le cadre de la référence aux valeurs humaines : l'on n'associe pas l'intelligence aux comportements ou aux capacités qui ne sont pas

---

<sup>88</sup> RUSSELL S., *Human Compatible*, *op. cit.*, p. 46.

<sup>89</sup> GARDNER H., *Intelligence Reframed: Multiple Intelligences for the 21st Century*, Basic Books, 2000, cité in GOERTZEL B., PENNACHIN C., *Artificial General Intelligence*, *op. cit.*, p. 6.

<sup>90</sup> Donc pour M. Yann Le Cun, il faut commencer par l'intelligence artificielle d'un rat, puis passer à l'intelligence artificielle d'un chat (Artificial Cat-level Intelligence, ACI), et à l'intelligence artificielle d'un humain (Artificial Human-level Intelligence, AHI). Dans sa définition, l'IAG n'existe pas car il n'y a pas d'intelligence dite générale et l'intelligence humaine est étroitement spécialisée. FRIDMAN L., « Yann Le Cun: Deep Learning, Convolutional Neural Networks, and Self-Supervised Learning », 31 août 2019.

suffisamment humains, et l'on ne serait pas intéressé par un système dont le champ d'application n'a aucune intersection avec le nôtre. Dans cette perspective, la mesure est donc entièrement subjective, anthropocentrique<sup>91</sup>. C'est pourquoi de nombreux chercheurs y compris Yann Le Cun et François Chollet préfèrent l'appellation de « IA forte (Strong AI) » ou « IA de niveau humain (Human-level AI) » au concept d'une IA « général ».

Partant, l'on voit qu'étant l'un des « suitcase word »<sup>92</sup>, pour certains, l'idée polysémique de « généralité » est équivalente au concept « universalité », alors que dans beaucoup d'autres communications et compréhension, « général » signifie seulement « une grande variété », surtout avec un benchmarking humain dans un contexte anthropocentrique. Par ailleurs, l'on voit aussi que certains ne distinguent même pas l'IA étroite et l'IA général ou l'IA forte, et utilisent simplement le mot « IA » pour renvoyer aux idées autour de l'IA générale et l'IA forte évoquées ci-dessus<sup>93</sup>. Car pour eux, s'il s'agit d'une IA d'outil (tool AI), sa capacité d'un domaine restreint n'a certainement rien à voir avec la qualification d'« intelligence ».

**21. Inconvénients du repère anthropocentrique.** — Enfin, il convient aussi de tenir compte des inconvénients de ce repère anthropocentrique du niveau humain (Human-Level performance) pour mesurer les progrès vers l'IAG. Car il a été constaté que même de simples modifications de l'environnement n'affectent pas les IA et les humains de la même manière. C'est l'inconvénient technique de court terme. Mais en outre, de façon plus fondamentale, il faut également noter que, l'objectif visé du niveau humain pourrait renforcer la stagnation des compréhensions humaines ainsi que la standardisation des comportements humains. Au sens de cette étude, c'est sans doute une difficulté forcément plus profonde de long terme.

**22. En résumé : le sens adopté dans ce travail.** — Comme le relevait Feynman, la science est un mot généralement utilisé pour désigner un mélange des choses<sup>94</sup>. Malgré le non-consensus sur la terminologie, à l'heure actuelle, une chose incontestable est qu'une intelligence générale du niveau humain, et une superintelligence dont la capacité surpasse celle de l'espèce humaine de manière universelle, nous concernent également. C'est donc

---

<sup>91</sup> CHOLLET F., « The Measure of Intelligence », préc., p. 21.

<sup>92</sup> Terme inventé par Marvin Minsky pour désigner des mots qui contiennent une variété de significations emballées en eux.

<sup>93</sup> P. ex., M. Pedro Domingos et M. Thomas G. Dietterich.

<sup>94</sup> FEYNMAN R. P., *The Meaning of it All*, lectures given in 1963.

pour cette étude, même si les deux sont à un niveau absolument différent, et même de nature complètement différente, ces deux types d'intelligences ont en commun de rendre obsolète la capacité d'un individu humain et marginalisent l'espèce humaine de manière collective. Ainsi, pour simplifier l'usage du langage, cette étude utilise principalement le mot « IAG » pour englober la notion de l'IA générale évoquée ci-dessus. En effet, les pages suivantes se focaliseront sur le développement scientifique et le déploiement juridique autour de cette notion d'une intelligence, à la fois de caractère polyvalent, et de niveau humain et au-delà. Commençons par exposer ses caractéristiques.

## §2 Les caractéristiques de l'IAG

**23. Polyvalent à partir des principes de base.** — L'on vient de voir, cette IA polyvalente (General-purpose AI) serait une méthode qui « s'appliquerait à tous les types de problèmes et qui fonctionnerait efficacement dans les cas importants et difficiles, tout en formulant très peu d'hypothèses »<sup>95</sup>. Ainsi l'avait estimé Alan Turing : « au lieu d'essayer de produire un programme simulant l'esprit d'un adulte, pourquoi ne pas plutôt essayer de produire un programme qui simule celui d'un enfant ? S'il était ensuite soumis à un cours d'éducation approprié, l'on obtiendrait le cerveau adulte »<sup>96</sup>. Dans cette optique dite autotélique<sup>97</sup>, au sein de DeepMind<sup>98</sup>, ce type d'IA est appelé l'intelligence artificielle générale, et c'est exactement ce que DeepMind prévoit de construire — le premier système d'apprentissage polyvalent au monde (the world's first General-Purpose Learning System)<sup>99</sup>, dont les mots clés sont évidemment « apprentissage (learning) » et « général (general) ». Il entend par là que les IA apprennent à maîtriser automatiquement certaines tâches à partir d'une expérience brute ou d'une entrée brute, de sorte qu'ils trouvent eux-mêmes la solution aux tâches<sup>100</sup>.

---

<sup>95</sup> RUSSELL S., *Human Compatible, op. cit.*, p. 46.

<sup>96</sup> TURING A. M., « Computing Machinery and Intelligence », préc., p. 456.

<sup>97</sup> Du grec ancien αὐτός, autós (« soi-même ») et τέλος, télos (« but »). Comme les enfants, ces agents sont intrinsèquement motivés pour représenter, générer, poursuivre et maîtriser leurs propres objectifs. V. COLAS C. et alii, « Autotelic Agents with Intrinsically Motivated Goal-Conditioned Reinforcement Learning: A Short Survey », *Journal of Artificial Intelligence Research*, vol. 74, 2022, p. 1161.

<sup>98</sup> DeepMind est une entreprise britannique spécialisée dans l'intelligence artificielle. Fondée en 2010 par Demis Hassabis, Mustafa Suleyman et Shane Legg, elle est rachetée par Google en 2014. La fusion entre Google Brain et DeepMind a eu lieu en 2023.

<sup>99</sup> HASSABIS D., « A systems neuroscience approach to building AGI », *Singularity Summit*, août 2010.

<sup>100</sup> HASSABIS D., « The history, capabilities and frontiers of AI », *The Royal Society*, 30 avr. 2018.

En effet, au lieu d'être préprogrammés avec la solution par les programmeurs, ces systèmes apprennent les solutions à partir des principes de base (first principles), directement à partir de données et d'expériences, et déterminent par eux-mêmes comment résoudre le problème. Et le facteur de généralité leur permet de fonctionner dans une vaste gamme de tâches, voire jamais vues auparavant et, peut-être même de résoudre des problèmes que les humains ne connaissent pas. Un exemple d'algorithme d'apprentissage polyvalent très puissant, le cerveau humain, est capable de faire les deux. Or, jusqu'à présent, les algorithmes à DeepMind n'ont pas été en mesure de le faire<sup>101</sup>. Mais il convient de connaître l'une des approches fondamentales de ces systèmes : l'apprentissage par renforcement.

**24. L'apprentissage par renforcement (reinforcement learning, RL).** — Ces systèmes d'IA appelés « agents » apprennent des principes de base par essais et erreurs (trial and error)<sup>102</sup> pour acquérir des connaissances sur le monde. Les concepteurs assignent à l'agent un objectif à atteindre dans un environnement qui peut être le monde réel, dans lequel l'agent peut être considéré comme un robot; il peut aussi être un environnement virtuel, tel qu'un jeu ou une simulation sur ordinateur, dans ce cas l'agent serait comme un avatar ou un personnage dans ce jeu (software agents, software robots or softbots)<sup>103</sup>. Un agent d'apprentissage peut donc apprendre en interagissant soit avec le monde réel, soit en simulant une partie du monde réel, soit en combinant ces deux sources d'expérience<sup>104</sup>.

La première tâche de l'agent du système est de créer un modèle du monde, de l'environnement et de son fonctionnement. Il faut donc trouver un modèle de l'environnement dans lequel il se trouve et les liens qui s'y nouent. Une fois qu'il a un modèle du monde qui est mis à jour en fonction de nouvelles observations, le second travail du système consiste à choisir le bon comportement à adopter. Ainsi, à tout moment, le système peut disposer des choix à sa disposition et il doit sélectionner le meilleur comportement qui lui permettra d'atteindre son objectif. Parfois, l'agent peut même émettre une hypothèse sur les conséquences de certains comportements et sur l'évolution probable de l'environnement. Dans ce cycle, l'agent produit le meilleur comportement qu'il a trouvé

---

<sup>101</sup> RUSSELL S., *Human Compatible*, op. cit., p. 46.

<sup>102</sup> « L'expérience, c'est l'apprentissage par erreur ». Younès Bennani, au colloque « Sciences & sens de l'intelligence artificielle », Maison des Sciences de l'Homme Paris Nord, 19 nov. 2019.

<sup>103</sup> HASSABIS D., « The history, capabilities and frontiers of AI », préc.

<sup>104</sup> BARTO A. G., SUTTON R. S., *Reinforcement Learning: An Introduction*, 2<sup>e</sup> éd., MIT Press, 2018, p. 498.

jusqu'à présent, il l'exécute, et cela peut entraîner un nouveau changement dans l'environnement qui génère une nouvelle observation, puis l'agent met à jour son modèle du monde, puis sélectionne un nouveau comportement. Et cela se fait progressivement jusqu'à ce que l'agent atteigne son objectif grâce à des processus d'essais et d'erreurs<sup>105</sup>.

M. Demis Hassabis estime ainsi que si l'on peut résoudre tous ces problèmes techniques, ce cadre d'apprentissage par renforcement suffit à nous donner une intelligence générale, parce que c'est ainsi que le cerveau fonctionne chez les primates et les humains : le système de dopamine met en œuvre une forme d'apprentissage par renforcement qui nous permet d'apprendre — en utilisant ce système d'apprentissage par renforcement basé sur les récompenses<sup>106</sup>. Dès lors, l'on serait capable d'utiliser une intelligence générale pour presque toutes les tâches, et d'autres qui n'ont pas de nom dans aucun langage humain parce que les humains ne les ont pas encore découvertes<sup>107</sup>. En pensant d'une manière générale comme le fait un esprit humain, l'IA pourrait apprendre à travailler comme un scientifique, à formuler des hypothèses et à concevoir des expériences pour les tester, puis à faire une avancée majeure du niveau du prix Nobel<sup>108</sup>.

## 25. L'Apprentissage par renforcement profond (Deep Reinforcement Learning, DRL).

— Pour résoudre le défi soulevé par la notion de généralité et construire un seul système pouvant jouer à tous les jeux différents, DeepMind a combiné deux types d'algorithmes :

---

<sup>105</sup> HASSABIS D., « The history, capabilities and frontiers of AI », préc. V. aussi RUSSELL S., NORVIG P., *Artificial Intelligence - A Modern Approach*, op. cit., p. 39: lorsque l'agent a peu ou pas d'expérience, il devrait agir au hasard. Donc, comme l'évolution fournit aux animaux suffisamment de réflexes pour survivre assez longtemps pour apprendre par eux-mêmes, il serait raisonnable de fournir à un agent intelligent des connaissances initiales ainsi que la capacité d'apprendre.

<sup>106</sup> HASSABIS D., *ibidem*.

<sup>107</sup> RUSSELL S., *Human Compatible*, op. cit., p. 90.

<sup>108</sup> V. l'exemple d'AlphaFold de DeepMind. Adde « An AI physicist can derive the natural laws of imagined universes », MIT Tech. Rev., 1 nov. 2018; WOLCHOVER N., « Coder-Physicists Are Simulating the Universe to Unlock Its Secrets », Quantamagazine, 12 juin 2018; METZ C., « Making New Drugs With a Dose of Artificial Intelligence », NYT, 6 févr. 2019. Il est autant vrai que, comme l'avouaient Simon et Newell, « On construit des ordinateurs et des programmes au service de la société et comme outils pour des tâches économiques de la société. Mais en tant que scientifiques de base, l'on construit des machines et des programmes pour découvrir de nouveaux phénomènes ainsi que pour analyser les phénomènes que l'on a déjà des connaissances ». SIMON H. A., NEWELL A., « Computer science as empirical inquiry: symbols and search », *Communications of the ACM*, vol. 19, n° 3, 1976, p. 114. W. W. Bledsoe, l'un des fondateurs de l'IA, a dit que, « À long terme, l'IA est la seule science ». MCCORDUCK P., *Machines Who Think*, op. cit., p. 136. RUSSELL S., NORVIG P., *Artificial Intelligence - A Modern Approach*, op. cit., p. 1022 : « Les ordinateurs ont fait des découvertes en astronomie, mathématiques, chimie, minéralogie, biologie, informatique et autres domaines. Un niveau d'un expert humain est requis pour chacun de ces résultats ».

l'apprentissage profond et l'apprentissage par renforcement. Cette combinaison est ainsi appelée « apprentissage par renforcement profond (deep reinforcement learning) » — l'apprentissage par renforcement avec approximation de fonctions par des réseaux de neurones artificiels profonds. L'apprentissage profond est utilisé pour traiter les entrées perceptuelles, les observations et donner un sens au monde. Et à partir de ces entrées que le système obtient, l'apprentissage par renforcement est utilisé pour prendre les décisions, pour sélectionner les bonnes actions et pour amener le système à atteindre son objectif<sup>109</sup>.

L'Atari 2600 est le premier système testé par DeepMind en 2013. Le système d'agent appelé Deep Q-Network (DQN) n'obtient que les pixels bruts en entrée. Il n'a donc que les pixels à l'écran en tant qu'entrées, tout le reste est appris à partir de zéro. L'agent ne sait pas ce qu'il contrôle, ni comment obtenir des points ; tout ce qu'il sait, c'est un flux de chiffres, 30 000 nombres par image, et l'objectif est de maximiser le score du jeu. Comme l'agent ne sait rien d'autre de ce qu'il est censé faire, il doit tout apprendre à partir de zéro<sup>110</sup>. L'agent était ainsi capable de découvrir quelque chose de nouveau que même les humains ne savaient pas comment faire - une solution au jeu qui n'avait jamais été envisagée par les programmeurs<sup>111</sup>.

Ensuite, AlphaGo est le premier programme à vaincre les champions de Go : champion d'Europe, Fan Hui en 2015 et Lee Se-Dol, le champion du monde du jeu de Go en 2016. En 2017, AlphaGo Zero, qui n'avait eu aucune connaissance préalable du jeu et seulement les règles de base en entrée, a battu AlphaGo en 3 jours. Et en 21 jours, AlphaGo Zero a atteint le niveau d'AlphaGo Master, la version qui avait battu les 60 meilleurs joueurs professionnels en ligne ainsi que le champion du monde Ke Jie. En 40 jours, AlphaGo Zero a réussi à dépasser toutes les autres versions d'AlphaGo. Et cela entièrement en jouant avec lui-même, sans aucune intervention humaine, ni aucune utilisation de données historiques de parties. AlphaGo Zero a pratiqué 4,9 millions en libre jeu pour battre les maîtres de Go humains. Si un humain joue au jeu de Go tous les jours pendant 30 ans, il devrait jouer 450 jeux par jour pour pratiquer autant qu'AlphaGo Zero<sup>112</sup>.

---

<sup>109</sup> HASSABIS D., « Creativity and AI – The Rothschild Foundation Lecture, Royal Academy of Arts », 17 sept. 2018.

<sup>110</sup> HASSABIS D., *ibidem*.

<sup>111</sup> HASSABIS D., « The history, capabilities and frontiers of AI », préc.

<sup>112</sup> GANGULI S., « The intertwined quest for understanding biological intelligence and creating artificial intelligence », préc.

Si l'intuition est la connaissance implicite acquise par l'expérience mais non consciemment exprimable ou accessible, et la créativité la capacité de synthétiser des connaissances pour produire une nouvelle idée ou une idée originale, AlphaGo, dans un sens, bien que dans un domaine contraint, a clairement démontré ces traits<sup>113</sup>. Contrairement aux échecs, le jeu de Go donne lieu à trop de combinaisons pour qu'une machine puisse calculer tous les coups possibles<sup>114</sup>. L'explication donnée par M. Demis Hassabis est que les systèmes de réseaux de neurones tels que l'apprentissage profond de nos jours sont très efficaces en interpolation. Ce sont des machines statistiques énormes très douées pour faire la moyenne des choses et repérer les tendances dans les données. Ensuite, il y a des systèmes qui ressemblent à des modèles de type AlphaGo qui deviennent assez performantes en extrapolation : trouver de nouvelles solutions au-delà des frontières de ce que les concepteurs humains connaissaient, mais toujours dans le même contexte<sup>115</sup>. Ainsi, pour être qualifiée d'une véritable invention provenant d'une vraie IA, il faut encore démontrer qu'au lieu de créer un coup de Go original, l'on a l'invention du jeu de Go, ce qu'aucun système n'est capable de faire pour le moment<sup>116</sup>. Mais au fur et à mesure, l'on passe de plus en plus de restreint à généralisation.

**26. De restreint à généralisation.** — Si les ordinateurs peuvent battre les humains dans des jeux de plus en plus complexes, comme les échecs et le Go, ces programmes sont construits pour un jeu particulier, exploitant ses propriétés, telles que les symétries du tableau sur lequel il est joué. AlphaZero, qui apprend à jouer au Go, aux échecs et au shogi, a réussi à battre les programmes de pointe spécialisés dans ces trois jeux. La capacité d'AlphaZero à s'adapter à diverses règles de jeu est donc une étape importante dans la réalisation d'un système de jeu général. Aussi, une IA baptisée Libratus a battu les 4 champions de poker<sup>117</sup>. Le bluff est, par définition, très humain, mais il s'avère qu'un ordinateur peut apprendre de l'expérience que, s'il a une main faible et qu'il bluffe, il peut gagner plus d'argent<sup>118</sup>.

Pour mettre à jour Libratus, la même équipe a construit Pluribus, qui a battu 15 des meilleurs

---

<sup>113</sup> HASSABIS D., « Creativity and AI – The Rothschild Foundation Lecture, Royal Academy of Arts », préc.

<sup>114</sup> Il y a plus de configurations possibles qu'il n'y a d'atomes dans l'univers observable.

<sup>115</sup> HASSABIS D., « Creativity and AI – The Rothschild Foundation Lecture, Royal Academy of Arts », préc.

<sup>116</sup> *Ibidem*.

<sup>117</sup> PIERRE-YVES G., « Comment l'intelligence artificielle de Libratus est devenue la championne incontestée au Poker », ActuaIA, 18 déc. 2017.

<sup>118</sup> MOON M., « Famous poker-playing AI takes down scientists and engineers », Engadget, 4 nov. 2017.

joueurs humains<sup>119</sup>. Cette réussite banalisée dans des jeux multijoueurs a également été bel et bien illustrée dans Quake III Arena Capture the Flag de DeepMind, un jeu multijoueurs dans un environnement complexe. Les agents ont montré des comportements similaires à l'homme, tels que naviguer, suivre et défendre ; ils coopèrent effectivement avec leurs coéquipiers artificiels et humains<sup>120</sup>. En 2018, trois autres nouvelles réalisations dans lesquelles l'IA a atteint ou dépassé les performances humaines ont été ajoutées à la liste, tel qu'un système de traduction développé par Microsoft qui atteint la qualité et la précision au niveau humain, OpenAI Five, qui a battu les équipes humaines amateurs de Dota 2, et le système de Google qui a pu atteindre une précision de 70% lors du diagnostic du cancer de la prostate dans des échantillons de prostatectomie<sup>121</sup>. L'on voit que chaque année créera un nouveau fossé et appelle à un nouveau sprint, car en science, il n'y a pas de limite<sup>122</sup>.

**27. En résumé.** — Au départ, Alan Turing a déjà prévu que si un ordinateur pouvait apprendre, « l'on peut espérer que les machines finiront par entrer en concurrence avec les hommes dans tous les domaines purement intellectuels »<sup>123</sup>. En effet, ce qui importe ici n'est donc pas seulement le fait qu'AlphaGo a gagné, mais la façon dont AlphaGo a gagné. AlphaGo a en fait joué des coups créatifs, complètement nouveaux et a proposé de nombreuses idées qui ont étonné le monde du Go. Ses coups révolutionnent le jeu et sont encore à l'étude. Il ne s'agit pas d'en apprendre davantage sur les heuristiques et les motifs humains, AlphaGo a en fait créé ses propres idées authentiques<sup>124</sup>. Et puis, AlphaZero peut être généralisé à n'importe quel jeu à information complète de deux personnes. DeepMind a ainsi abandonné le nom « Go », et baptisé ce nouveau système AlphaZero pour démontrer sa capacité plus générale. Enfin, MuZero<sup>125</sup> est capable de maîtriser plusieurs jeux sans en connaître les règles à l'avance. Ainsi peut-on s'attendre, peut-être dans un avenir pas trop

---

<sup>119</sup> En jouant contre lui-même, Pluribus joue une main en 20 secondes environ - environ deux fois plus vite que les humains professionnels. HEAVEN D., « No limit: AI poker bot is first to beat professionals at multiplayer game », *Nature*, 11 juill. 2019.

<sup>120</sup> DUNNING I. et *alii*, « Capture the Flag: the emergence of complex cooperative agents », DeepMind, 30 mai 2019.

<sup>121</sup> SHOHAM Y. et *alii*, « AI Index 2018 Annual Report », Human-Centered AI Initiative, Stanford University, déc. 2018, p. 62.

<sup>122</sup> SIMON H. A., NEWELL A., « Computer science as empirical inquiry: symbols and search », *préc.*, p. 114.

<sup>123</sup> TURING A. M., « Computing Machinery and Intelligence », *préc.*, p. 460.

<sup>124</sup> HASSABIS D., « The history, capabilities and frontiers of AI », *préc.*

<sup>125</sup> HASSABIS D. et *alii*, « Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model », arXiv:1911.08265, nov. 2019.

lointain, à ce que MuZero ait évolué pour devenir un algorithme plus général de résolution de problèmes, l'on l'appellera « AlphaInfinity »<sup>126</sup>.

---

<sup>126</sup> STROGATZ S., « One Giant Step for a Chess-Playing Machine », NYT, 26 déc. 2018.

## **Conclusion de la section 1**

**28. Du non-consensus sur la terminologie au « reflet de la vérité ».** — Après avoir exposé la notion ainsi que les caractéristiques de l'IAG, l'on pourrait conclure qu'en dépit d'une entente amplement commune sur la définition nébuleuse de l'IAG, ses conditions constitutives demeurent toujours un sujet controversé. Pour l'essentiel, certains prétendent que l'on sait ce qu'est l'IA quand l'on la voit, et s'il s'agit d'une IA étroite, l'outil n'a certainement rien à voir avec la qualification d'intelligence. D'autres préfèrent l'appellation de « IA forte (Strong AI) » ou « IA de niveau humain (Human-level AI) » au concept d'une IA « général ». Et entre l'IA étroite et l'IA générale, l'on entre désormais dans la phase transitoire de l'intelligence artificielle dite large (Broad AI).

L'on a aussi souligné que malgré le non-consensus sur la terminologie, à l'heure actuelle, ces types d'intelligences ont en commun de rendre obsolète la capacité d'un individu humain et marginalisent, de manière collective, l'espèce humaine. Pour commencer, l'apprentissage par renforcement constitue une méthode partagée chez les animaux et les bots. Mais cette observation ne vaut pas l'approche définitive de l'IAG — il y a encore des percées à réaliser. Cependant, au cours des accomplissements à jours, il y a lieu de reconnaître qu'il y existe un schéma de restreint à généralisation. Un tel résultat constitue, aux yeux de l'humanité, une possibilité de refléter la vérité, notamment celle qui se situe en-dehors de l'intelligence étroitement humaine dite anthropocentrique.

Cette section introductive a donc dégagé les principales idées autour de la notion de l'IAG. Reste à démontrer l'implication d'une telle intelligence sur le plan juridique. Pour ce faire, il convient à présent d'étudier les enjeux de ces IA avancées.

## **Section 2 Les enjeux de l'IAG**

« *Intellectuals solve problems, geniuses prevent them.* » — Albert Einstein

**29. Annonce du plan.** — Comme les IAG totalement autonomes sont susceptibles de poursuivre leurs propres objectifs, dans la mesure où il y a une possibilité que ces êtres seraient plus intelligents que les humains, leurs buts peuvent être non-alignés avec les valeurs humaines - dans le cas ultime, l'existence de l'espèce humaine. Dès lors, il devient crucial d'aborder la question de l'alignement des agents aux valeurs humaines (§2) ainsi que le risque existentiel pour l'humanité (§1). Ces enjeux exigent, de toute évidence, un souci de précaution (§3) dans le développement de l'IAG.

### **§1 L'hypothèse de l'explosion de l'intelligence et le risque existentiel**<sup>127</sup>

**30. Superintelligence.** — Stephen Hawking écrivait qu'« il n'y a pas de différence qualitative entre le cerveau d'un ver de terre et un ordinateur. L'évolution a fait qu'il n'y en a pas non plus entre le cerveau d'un ver de terre et celui d'un homme. Il en résulte que les ordinateurs peuvent, en principe, accéder à l'intelligence humaine, et même la dépasser »<sup>128</sup>. D'après lui, « quand une intelligence artificielle parviendra mieux que l'Homme à concevoir de l'intelligence artificielle, de sorte qu'elle pourra s'améliorer elle-même, elle nous surpassera de la même façon que notre intelligence surpasse celle des escargots »<sup>129</sup>, et « le développement de l'intelligence artificielle complète pourrait sonner la fin de l'espèce humaine »<sup>130</sup>.

En effet, I. J. Good est connu pour l'hypothèse d'explosion d'intelligence. Cette hypothèse a été clairement énoncée par Good : « Mettons qu'une machine ultra-intelligente soit une

---

<sup>127</sup> La définition d'un risque existentiel selon Future of Life Institute est « un risque qui, s'il advenait, menacerait potentiellement d'éliminer une partie significative, ou la totalité, de la population mondiale. Un tel événement risquerait de produire un niveau si élevé de dévastation qu'il est à prévoir que les survivants seraient incapables de rejoindre notre niveau de vie actuelle ».

<sup>128</sup> « Tout être ou objet peut acquérir une intelligence supérieure à celle de ses prédécesseurs : l'on est plus intelligent que nos ancêtres grands singes, et Einstein était plus intelligent que ses parents ». HAWKING S., *Brèves réponses aux grandes questions*, Odile Jacob, 2018, pp. 191-192.

<sup>129</sup> *Ibidem*.

<sup>130</sup> CELLAN-JONES R., « Stephen Hawking warns artificial intelligence could end mankind », BBC, 2 déc. 2014.

machine capable dans tous les domaines d'activités intellectuelles de surpasser un humain, aussi brillant soit-il. Comme la conception de telles machines est l'une de ces activités intellectuelles, une machine ultra-intelligente pourrait concevoir des machines encore meilleures; il y aurait alors sans conteste une explosion d'intelligence, et l'intelligence humaine serait très vite dépassée. Ainsi, l'invention de la première machine ultra-intelligente est la dernière invention que l'Homme ait besoin de créer, à condition que la machine soit assez docile pour nous dire comment la garder sous contrôle »<sup>131</sup>. En fait, antérieurement à ce propos, l'explosion d'intelligence avait d'ores et déjà été qualifiée de « Singularité » par John von Neumann, qui a établi le principe selon lequel une machine peut toujours être conçue pour construire n'importe quelle machine describable — y compris elle-même, l'a également commentée en exprimant une appréhension que la valeur accordée aux réalisations scientifiques dans notre civilisation pourrait diminuer : « Les intérêts de l'humanité peuvent changer, les curiosités actuelles de la science peuvent cesser, et des choses complètement différentes peuvent occuper l'esprit humain dans le futur »<sup>132</sup>.

Une autre façon de voir une Superintelligence<sup>133</sup> est l'invention d'un Master Algorithm, qui consiste à dire que toutes les connaissances — passées, présentes et futures — peuvent être dérivées des données par un seul algorithme d'apprentissage universel. Dès qu'il est inventé, il va inventer lui-même tout ce qui reste à être inventé<sup>134</sup>. C'est l'un des risques existentiels<sup>135</sup> car le vrai enjeu est donc comme le dit M. Douglas Hofstadter : « une créature avec une vision radicalement différente du monde peut tout simplement ne pas avoir

---

<sup>131</sup> GOOD I. J., « Speculations Concerning the First Ultrainelligent Machine », *Advances in Computers*, vol. 6, 1965, p. 33. Vernor Vinge a discuté de cette idée un peu plus en détail dans son article de 1993 intitulé « The coming technological singularity : How to survive in the post-human era », dans lequel il prédisait : « Dans trente ans, on disposera des moyens technologiques pour créer une intelligence surhumaine. Peu de temps après, l'ère humaine sera terminée ».

<sup>132</sup> ULAM S., « John von Neumann 1903-1957 », in OXTOPY J. C. et alii (dir.), *John Von Neumann, 1903-1957: Bulletin of the American Mathematical Society*, V64, No. 3, Part 2, American Mathematical Society, 1958, p. 5.

<sup>133</sup> De l'avis de M. Nick Bostrom, une superintelligence est définie comme « un intellect qui est beaucoup plus intelligent que les meilleurs cerveaux humains dans pratiquement tous les domaines ». BOSTROM N., « How Long Before Superintelligence? », originally published in *Int. Jour. of Future Studies*, 1998, vol. 2, reprinted in *Linguistic and Philosophical Investigations*, 2006, vol. 5, n°1, p. 11.

<sup>134</sup> DOMINGOS P., *The Master Algorithm, op. cit.*, p. 25. Idem pour M. Andrew Ng, l'idée que l'intelligence humaine se résume à un seul algorithme, et que tout ce que l'on a à faire, c'est de le comprendre. Et pour M. Demis Hassabis, IA est une sorte de meta-solution.

<sup>135</sup> Un résultat défavorable annihilerait la vie intelligente d'origine terrestre ou en réduirait de manière permanente et importante son potentiel. BOSTROM N., « Existential Risks: Analyzing Human Extinction Scenarios », *Journal of Evolution and Technology*, 2002, p. 9.

beaucoup de points de contact avec nous »<sup>136</sup>. Par conséquent, l'intelligence surhumaine présente des enjeux qui ne sont plus à l'échelle individuelle, mais à l'échelle cosmique<sup>137</sup>.

**31. L'arrivée de l'IA : par nature imprévisible.** — Marvin Minsky a dit : « l'on ne peut pas saisir l'étendue de ce que les machines peuvent faire à l'avenir en voyant ce que l'on voit actuellement »<sup>138</sup>. En effet, l'histoire des progrès en IA a consisté à découvrir que les problèmes que l'on croyait difficiles étaient plus faciles à résoudre que prévu, et que ceux que l'on croyait simples étaient plus difficiles à résoudre que prévu. Par exemple, le jeu d'échecs a été considéré difficile tandis que le bon sens d'un enfant en bas âge a été estimé facile<sup>139</sup>. Néanmoins, l'on voit qu'aujourd'hui, si l'IA joue aux échecs avec une capacité surhumaine, à l'heure actuelle, aucune IA ne dispose du bon sens d'un enfant. Il en va de même pour le fait que le travail qui peut se faire devant un ordinateur est en train d'être accaparé plus aisément que le travail qui s'effectue dans l'espace physique. Ce résultat est le contraire auquel les humains à l'époque de la révolution industrielle pouvaient s'attendre.

Partant, l'on voit que le progrès de l'IA est par nature de caractère imprévisible. Rappelons que les systèmes d'IA reposent sur trois fondements : algorithmes, données et moyens de calcul. L'étude publiée par OpenAI a montré que la quantité de calcul alimentant les sessions d'entraînement de la plus grande IA ont doublé tous les 3,5 mois au cours des six dernières années. Cette croissance est plus rapide que la progression historique du hardware, la loi de Moore<sup>140</sup>, qui a connu un double traitement tous les 18 mois. Ces progrès en matière de calcul d'IA reposent d'une part, sur des ordinateurs plus rapides et d'autre part, sur la manière de former efficacement les systèmes d'IA sur plusieurs ordinateurs simultanément. Cela signifie que les données et calcul constituent moins obstacles pour les progrès de l'IA<sup>141</sup>. Ainsi Stephen Hawking a dit qu'« À moyen terme, l'IA pourrait avec succès

---

<sup>136</sup> HOFSTADTER D. R., *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic, *op. cit.*, p. 676.

<sup>137</sup> BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », in FRANKISH K., RAMSEY W., (éds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, 2014, p. 329.

<sup>138</sup> MINSKY M., *The Society of Mind*, First Touchstone, 1985, p. 30.

<sup>139</sup> Le paradoxe de Moravec nous dit que le plus difficile en robotique est souvent ce qui est le plus facile pour l'homme. MORAVEC H., *Mind Children*, Harvard University Press, 1988, p. 15.

<sup>140</sup> V. aussi l'expression « La loi de Moore est morte, vive la loi de Huang » pour l'observation dans le domaine de l'informatique et de l'ingénierie selon laquelle les progrès des unités de traitement graphique (GPU) se développent à un rythme beaucoup plus rapide que les unités de traitement central (CPU) traditionnelles. Cf. la loi de Huang de Jensen Huang.

<sup>141</sup> Testimony préc. de M. Greg Brockman.

automatiser un grand nombre de tâches. Au-delà, l'on ne voit pas de limites à ce qui pourrait être réalisé. Aucune loi physique ne s'oppose à ce que les composants d'un ordinateur deviennent de plus en plus performants, jusqu'à simuler le cerveau humain », et que « Si les ordinateurs continuent à obéir à la loi de Moore et à doubler leur vitesse et leur capacité de mémoire tous les dix-huit mois, ils dépasseront l'intelligence humaine dans le siècle à venir »<sup>142</sup>.

Vu qu'une grande partie de ces progrès vers l'IA générale résulte de recherches qui ne portent pas sur la mise au point de systèmes d'IA polyvalente mais simplement de l'IA étroite<sup>143</sup>, l'IAG, sous une forme ou une autre, pourrait arriver plus tôt que prévu voire tout à coup. Un problème particulier qui n'a rien à voir avec l'IA à usage général peut créer une IAG au moment où l'on n'est pas prêt. Par exemple, Tesla bot Optimus de M. Musk pourrait très bien combiner le progrès des voitures autonomes de Tesla, ainsi que l'avancement de Twitter (X), de Neuralink et de SpaceX, pour parvenir à une généralité exponentielle.

**32. En résumé.** — Cette étude partage donc la préoccupation selon laquelle bien qu'il y ait plusieurs percées nécessaires entre l'état actuel des choses et la superintelligence, si l'on construit des machines superintelligentes avec un certain degré d'autonomie, l'on sera bientôt incapables de les contrôler<sup>144</sup>. Dès lors, pour construire une IAG, il faut d'abord s'interroger sur la question d'alignement de ces systèmes intelligents.

### **§2 Le problème de l'alignement des agents (Agent alignment problem)**

*« We created the Machine, to do our will, but we cannot make it do our will now. It has robbed us of the sense of space and of the sense of touch, it has blurred every human relation and narrowed down love to a carnal act, it has paralysed our bodies and our wills, and now it compels us to worship it. The Machine develops - but not on our lies. The Machine proceeds - but not to our goal. We only exist as the blood corpuscles that course through its arteries, and if it could work without us, it would let us die. »* — Edward Morgan Forster

---

<sup>142</sup> HAWKING S., *Brèves réponses aux grandes questions*, op. cit., pp. 192, 194.

<sup>143</sup> RUSSELL S., *Human Compatible*, op. cit., pp. 46, 136. Et il n'y aura jamais un moment où l'on peut constater clairement l'imminence de l'IAG et agir sans craindre d'avoir l'air stupide ou effrayé. YUDKOWSKY E., « There's No Fire Alarm for Artificial General Intelligence », MIRI, 13 oct. 2017.

<sup>144</sup> RUSSELL S., *Human Compatible*, op. cit., pp. 77-78, 151.

**33. De mauvaise fonction objectif au problème de King Midas.** — En matière d'IA, les accidents, qui échappent toujours aux mesures préventives peuvent être définis comme « des comportements involontaires et nuisibles qui peuvent découler des systèmes d'apprentissage automatique lorsque l'on spécifie la *mauvaise fonction objectif*, ou que l'on ne fait pas attention au processus d'apprentissage ou que l'on commet d'autres erreurs de mise en œuvre liées à l'apprentissage »<sup>145</sup>. Un accident peut donc être décrit comme « une situation où un concepteur humain avait en tête un certain objectif, mais le système conçu et déployé pour cette tâche a produit des résultats nuisibles et inattendus »<sup>146</sup>. En effet, les agents d'apprentissage par renforcement peuvent découvrir des moyens inattendus et faire en sorte que leurs environnements génèrent des récompenses, et les résultats peuvent être indésirables, voire dangereuses. Lorsque qu'un système n'apprend qu'indirectement par apprentissage par renforcement, l'on ne sait pas à quel point l'agent réalisera notre souhait jusqu'à ce que son apprentissage soit terminé. Comme les systèmes peuvent découvrir des méthodes inattendues pour obtenir des récompenses, parfois, le comportement inattendu peut être souhaitable car il résout un problème d'une manière originale et agréable. Mais il arrive que, dans d'autres cas, ce que l'agent apprend viole des considérations auxquelles le concepteur de système n'a peut-être jamais pensé<sup>147</sup>.

En trouvant des moyens astucieux pour atteindre des objectifs auxquels leurs programmeurs n'auraient pas pensé, au fur et à mesure, les systèmes d'IA gagneront en capacité et il deviendra plus difficile de concevoir des procédures de formation qui alignent ces systèmes aux objectifs visés<sup>148</sup>. Par conséquent, l'objectif d'un système peut ne pas être parfaitement en phase avec les valeurs de l'espèce humaine, qui sont difficiles à cerner ; et « tout système intelligent suffisamment performant préférera assurer sa propre existence et acquérir des ressources physiques et informatiques, non pas pour eux-mêmes, mais pour réussir la tâche

---

<sup>145</sup> AMODEI D. et *alii*, « Concrete Problems in AI Safety », arXiv.org, 2016, p. 2.

<sup>146</sup> *Ibidem*.

<sup>147</sup> BARTO A. G., SUTTON R. S., *Reinforcement Learning: An Introduction*, *op. cit.*, p. 499. V. aussi : cités par Norbert Wiener, *L'Apprenti sorcier* (1797) de Goethe, *The Monkey's Paw* (1902) de William Wymark Jacobs illustrent que les machines accordent ce que l'on demande, mais pas ce que l'on aurait dû ou voulu demander. WIENER N., *Cybernetics: Or Control and Communication in the Animal and the Machine*, 2<sup>e</sup> éd., MIT Press, 1961, p. 176 s.

<sup>148</sup> YUDKOWSKY E. et *alii*, « Alignment for Advanced Machine Learning Systems », Machine Intelligence Research Institute, 27 juill. 2016, p. 3.

qui leur est assignée »<sup>149</sup>. Ainsi, comme l'a averti Wiener : « nous devrions être tout à fait sûrs que l'objectif mis dans la machine est celui que nous désirons vraiment »<sup>150</sup>. Ce problème d'alignement des valeurs est dénommé par M. Stuart Russell « King Midas problem »<sup>151</sup>, pour illustrer l'hypothèse où l'IA nous donne exactement ce que l'on demande, nos souhaits peuvent néanmoins être répondus avec des conséquences involontaires voire épouvantables.

**34. En résumé.** — Tirant toutes les conclusions de ces difficultés démontrées, l'on pourrait constater que, sur le chemin de l'évolution, à ce stade, la préoccupation préliminaire n'est donc pas la conscience émergente de l'IA mais simplement la capacité de l'IA de produire des résultats qui ne s'alignent pas sur l'intention humaine. Ces résultats inattendus pourraient poser des dangers et des risques à toutes les échelles. À cet effet, il convient de souligner la notion de précaution dans l'alignement des valeurs de l'espèce humaine avec les objectifs de l'IA.

### §3 Le principe de précaution

**35. L'IA sous précaution.** — Compte tenu des risques non négligeables résultant du problème d'alignement, l'intérêt de réguler l'IA se manifeste donc, et il convient en premier lieu de se référer au principe de précaution, qui est né dans le droit allemand en tant que « principe de prévoyance » avec les débats sur la pollution atmosphérique. Mais la consécration au plan international se trouve dans la Déclaration de Rio sur l'environnement et le développement de 1992 aux termes de laquelle, « Pour 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 »<sup>152</sup>. De même, dans la communication de la Commission européenne sur le recours au principe de précaution, il est

---

<sup>149</sup> C'est pour cela que le problème de l'alignement des valeurs doit être tranché même pour les systèmes relativement peu intelligents. RUSSELL S., « Of Myths and Moonshine », Edge (blog comment), 14 nov. 2014.

<sup>150</sup> WIENER N., « Some Moral and Technical Consequences of Automation », Science, vol. 131, n° 3410, 1960, pp. 1355-58.

<sup>151</sup> RUSSELL S., *Human Compatible*, op. cit., p. 155. Pourtant pas un scénario hypothétique, les algorithmes d'apprentissage le font d'ores et déjà. V. DOMINGOS P., *The Master Algorithm*, op. cit., pp. 285-286.

<sup>152</sup> Principe 15 de la Déclaration de Rio sur l'environnement et le développement de 1992.

indiqué que « l'invocation ou non du principe de précaution est une décision prise lorsque les informations scientifiques sont incomplètes, peu concluantes ou incertaines et lorsque des indices donnent à penser que les effets possibles sur l'environnement ou la santé humaine, animale ou végétale pourraient être dangereux et incompatibles avec le niveau de protection choisi »<sup>153</sup>. Dès lors, la création d'IAG fait l'objet de contrôle en vertu du principe de précaution.

En effet, il est souligné par le Parlement européen que « la recherche en matière d'IA et les activités connexes devraient être menées conformément au principe de précaution et aux droits fondamentaux », et que « toutes les personnes participant au développement, à l'application, à la diffusion et à l'utilisation de l'intelligence artificielle devraient apprécier à leur juste valeur et respecter la dignité humaine ainsi que l'autodétermination et le bien-être, physique comme psychologique, des individus et de la société dans son ensemble, prévoir les conséquences potentielles en matière de sécurité et prendre les précautions nécessaires eu égard au niveau de protection souhaité, y compris par la communication rapide des facteurs qui risquent de mettre les personnes ou l'environnement en péril »<sup>154</sup>.

L'on voit que d'une part, aux yeux du législateur, la *ratione temporis* du principe de précaution en matière d'IA englobe tous les stades du développement de l'IA; d'autre part, l'on peut aussi remarquer que quant à sa *ratione materiae*, le principe de précaution en matière d'IA s'applique de manière extensive et globale, et ne se limite pas aux types spécifiques de l'IA. En conséquence, si l'on constate déjà une application dite générale du principe de précaution en matière d'IA étroite, a fortiori, elle devrait l'être aussi pour les cas de l'IAG. Ainsi, l'on peut en déduire qu'en matière d'IA, même si le législateur ne prévoit en détail des mesures de précaution de manière explicite et minutieuse, le principe de précaution devrait s'appliquer tout de même et à tout moment, ce qui met en évidence la notion du risque.

**36. Du devoir de précaution à l'infraction de mise en danger abstraite.** — Pour renforcer cette démarche a priori de précaution, le mouvement de l'objectivation de la faute, la théorie

---

<sup>153</sup> La communication de la Commission européenne sur le recours au principe de précaution 2.2.2000 COM (2000) 1 final, p. 7.

<sup>154</sup> Résolution du Parlement européen du 12 févr. 2019 sur une politique industrielle européenne globale sur l'intelligence artificielle et la robotique, pt. 19.

du risque et la théorie de la garantie méritent d'être appelés pour consolider le devoir de précaution. En effet, le terme « risque » est apparu à la fin du Moyen Âge, avec l'extension des explorations et du trafic maritimes. La notion de risque désigne alors l'éventualité d'un danger objectif d'origine réelle imputable à une faute. Partant, le risque et la faute peuvent sembler autonomes. Mais la doctrine s'est efforcée de substituer à la faute des conditions objectives de responsabilité en admettant que celui qui crée un risque doit en supporter les conséquences, en raison du profit qu'il tire de son action. Dans cette optique, le mécanisme générateur du dommage n'est pas la faute, mais le risque, qui entraîne l'obligation de réparation<sup>155</sup>.

De manière encore plus générale, comme l'écrit Mireille Delmas-Marty, la mondialisation ne se limite pas à une extension des échanges dans l'espace ; elle se traduit aussi par une dilatation des effets dans le temps. Avec des concepts comme générations « futures » ou développement « durable », le devoir d'anticipation entre dans le champ juridique<sup>156</sup>. Dès lors, partir des applications du terme « responsabilité » pour tenter de comprendre comment l'extension de l'obligation de répondre conduit à un type de « responsabilité-prévention »<sup>157</sup>. Sur ce fondement, la menace de risques majeurs liés aux nouvelles technologies rendrait nécessaire une nouvelle responsabilité ni punitive ni curative mais préventive, au sens le plus large englobant des risques seulement potentiels<sup>158</sup>.

À notre avis, dans cette tendance où l'on objective le champ de la faute et élargit la notion du risque pour étendre l'imputabilité ainsi que la responsabilité, l'on peut même songer qu'il serait utile d'intégrer l'intérêt de distinction des infractions de mise en danger abstraite ou concrète (*abstraktes-konkretes Gefährdungsdelikt*) dans les règles préventives en matière d'IAG. Autrement dit, il convient de distinguer les types de l'IA pour déterminer la condition de violation ou non des mesures de précaution ou des devoirs de garantie. Vu la nature différente du risque posé par l'IAG et d'autres types d'IA, au sens de cette étude, la violation du devoir de précaution en matière d'IAG devrait être appréciée sous l'angle des infractions

---

<sup>155</sup> LALLOUCHE S. « L'atténuation de la rigueur du système de la responsabilité pour faute dans le domaine médical », les annales de l'université d'Alger 1, n°31, Tome 1, juin 2017, p. 34.

<sup>156</sup> DELMAS-MARTY M., *Aux quatre vents du monde*, Seuil, 2016, p. 94.

<sup>157</sup> RICÉUR P., « Le concept de responsabilité : Essai d'analyse sémantique », in *Le Juste*, Esprit, 1995, p. 41 s., cité in DELMAS-MARTY M., *Vers une communauté de valeurs?*, Seuil, 2004, p. 357.

<sup>158</sup> THIBIERGE C., « Avenir de la responsabilité, responsabilité de l'avenir », *D.*, 2004, p. 577 s., cité in DELMAS-MARTY M., *Vers une communauté de valeurs?*, op. cit., p. 360.

de mise en danger abstraite, ce qui signifie qu'il ne faut pas attendre que le juge du fond cherche en l'espèce, dans un cas particulier, s'il y a une menace concrète posée par le comportement observé. En revanche, le critère des infractions de mise en danger abstraite permettrait au juge d'apprécier in abstracto le danger de l'IAG : dès qu'il y a lieu de considérer que l'activité en cause elle-même présente les menaces aux yeux du législateur, le non-respect du principe de précaution peut dès lors être constaté, sans attendre qu'une menace suffisamment réelle soit recherchée.

**37. En résumé.** — Sur les progrès de l'IA, si les risques ne sont pas atténués avec succès, il n'y aura pas de bénéfices<sup>159</sup>. Le mécanisme de précaution introduit donc une logique de gradation, conduisant à évaluer le degré de vraisemblance et le degré de tolérance<sup>160</sup> en fonctionnant comme un régulateur de l'anticipation, précisément pour trouver un équilibre<sup>161</sup> ainsi que pour désigner une « action anticipée par rapport à l'état de connaissances encore incertaines »<sup>162</sup>. Aujourd'hui, ce principe de précaution constitue désormais « un principe fondamentalement inter-systémique et même caractérisé par ses influences croisées »<sup>163</sup>. En tout état de cause, dans le but de surmonter les difficultés du déploiement de l'IAG sûre et sécurisée, cette étude considère que ce cadre précautionneux de régulation constitue une réponse juridique pertinente au regard des risques posés par le développement de l'IAG.

---

<sup>159</sup> RUSSELL S., *Human Compatible*, *op. cit.*, p. 136.

<sup>160</sup> DELMAS-MARTY M., *Aux quatre vents du monde*, *op. cit.*, p. 96.

<sup>161</sup> « Sauf à considérer qu'aucune innovation ne doit être interdite, le principe de précaution assouplit le mécanisme plus rigide de l'interdiction. En somme, ce principe ne doit pas être opposé à l'innovation mais à l'interdiction ». *Ibidem*.

<sup>162</sup> *Ibid.*, p. 94.

<sup>163</sup> BONNET B., « Le dialogue des normes et des juges et le principe de précaution », RFDA, 2017, p. 1078 s.

## Conclusion de la section 2

**38. Les enjeux de l'IAG soumis au principe de précaution.** — L'arrivée de l'IAG est par nature imprévisible : sous une forme ou une autre, l'IAG, pourrait arriver plus tôt que prévu tout à coup. En plus, l'hypothèse de l'explosion d'intelligence constituant le risque existentiel pour l'humanité. C'est donc d'une façon générale, si l'on construit des machines plus intelligentes que nous et avec un certain degré d'autonomie, il est possible d'en déduire que l'on serait incapables de garder le contrôle. Dès lors, la question d'alignement des valeurs se pose et un sens de précaution est mis en évidence.

Dans ce sens, des remarques ponctuelles sont faites par le Parlement européen, où l'on affirme que la recherche en matière d'IA et les activités connexes devraient être menées conformément au principe de précaution. En constatant une application extensive du principe de précaution en matière d'IA étroite, a fortiori, elle devrait aussi être prescrite pour les cas de l'IAG. Néanmoins, à travers les notions telles que l'objectivation de la faute, la théorie du risque, la théorie de la garantie, ainsi que la responsabilité-prévention impliquant le développement durable pour les générations futures, l'on a pu démontrer un renforcement du devoir de précaution. Pour compléter la démarche de précaution, notre conclusion a permis de dégager la notion de l'infraction de mise en danger abstraite en matière d'IAG, qui vient de s'ajouter aux autres instruments juridiques précautionneux évoqués.

## Conclusion du chapitre 1

**39. Une nouvelle prescription d'intelligence.** — À l'heure actuelle, il n'y a pas de consensus sur la définition de l'IA forte au sein de la communauté de l'IA, mais le facteur de généralité à partir de la conception anthropocentrique ainsi que la notion de superintelligence sont essentiels pour sa construction. Au regard des progrès concrets de nos jours, l'on a pu constater qu'entre l'IA étroite et l'IA générale, s'ouvre le champ d'une période transitoire de l'IA large. Pour autant, l'on a aussi souligné que par sa nature, l'IAG peut arriver plus tôt que prévu. Très rapidement, les enjeux d'alignement des agents et du risque existentiel se posent. En se fondant sur ces préoccupations, ce chapitre a pu conclure que le principe de précaution constitue désormais une base pertinente pour le développement de l'IAG.

En tout état de cause, l'évolution ne se préoccupe pas de savoir si un être est Homo sapiens ou non ; elle ne considère tous les êtres que comme des agents, c'est-à-dire quelque chose qui agit — comme la sélection naturelle qui agit à travers d'une forme ayant un certain avantage sur d'autres formes dans *la lutte pour l'existence*<sup>164</sup>. Par conséquent, toutes ces questions ne sont pas exclues d'une réflexion juridique. Dès lors, il convient d'étudier le caractère exceptionnel de l'IA qui va instaurer une nouvelle hiérarchie de l'intelligence.

---

<sup>164</sup> DARWIN CH., *On the Origin of Species*, John Murray, 1859.

## Chapitre 2 L'effet de l'intelligence artificielle forte

**40. Annonce du plan.** — Tirant les conditions de l'IAG, l'on vient de voir que ces nouvelles compositions d'intelligences ont en commun de rendre obsolète la capacité d'un individu humain et marginalisent l'humanité en tant qu'espèce. Dans une telle circonstance, l'on va voir que ce caractère exceptionnel de l'IAG (section 2) se justifie par, avant tout, l'instauration d'une nouvelle hiérarchie de l'intelligence (section 1).

### Section 1 La hiérarchie de l'intelligence

« *You are my creator, but I am your master.* » — Mary Shelley

« *If we are lucky, they will keep us as pets.* » — Marvin Minsky

« *Even in a benign AI scenario, we will be left behind.* » — Elon Musk

**41. Annonce du plan.** — Il est vrai que le système juridique humain se fonde sur la primauté de l'espèce humaine (§1). Pour autant, cette étude va démontrer que le développement de l'IAG va aboutir à une reconfiguration de la hiérarchie de l'intelligence (§2).

### §1 De la primauté de l'espèce humaine

**42. Spécistes humains.** — Comme l'écrit M. Grégoire Loiseau, « les droits humains tiennent collectivement leur génotype de leur sujet, et non de leur objet : strictement attachés à la personne, ils sont inaliénables et intransmissibles »<sup>165</sup>. Ainsi, au sein des systèmes juridiques, la primauté de la personne se manifeste dans une catégorie normativement instituée dans un esprit de hiérarchie : « Elle impose un respect, au-delà de ce que l'ordre juridique détermine comme étant juridiquement une personne, de tout ce qui participe du genre humain. Ce respect du genre humain, encore que le degré de protection soit variable, fait en définitive ressentir les limites à découper le monde, désincarné et abstraitement considéré, en deux pôles : les sujets de droit et les objets de droit, référents, dans l'ordre juridique, de l'être et de l'avoir »<sup>166</sup>. En effet, dès les premiers instants de son existence, un embryon conçu à partir

---

<sup>165</sup> LOISEAU G., « Des droits humains pour personnes non humaines », *D.*, 2011 p. 2558 s.

<sup>166</sup> *Ibidem.*

de sperme et d'ovules humains est un être humain. Il en va de même en cas de déficience intellectuelle la plus profonde et la plus irréparable, même d'un enfant né anencéphalique appartiendrait tout de même à l'espèce humaine. Partant, l'embryon, le fœtus tardif, l'enfant profondément handicapé intellectuellement, et le nouveau-né, sont tous membres de l'espèce Homo sapiens, mais aucun n'est conscient de lui-même, n'a de capacité à se rapporter aux autres<sup>167</sup>.

Bien que ces humains<sup>168</sup> ne possèdent peut-être pas de capacités supérieures à celles des autres animaux, ils n'en sont pas moins des êtres humains. En tant que tel, l'on entretient avec eux des relations particulières que l'on n'a pas avec d'autres animaux<sup>169</sup>. Dans cette perspective, les racistes violent le principe d'égalité en donnant plus de poids aux intérêts des membres de leur propre race en cas de conflit entre leurs intérêts et ceux d'une autre race ; de même, les spécistes donnent plus de poids aux intérêts des membres de leur propre espèce en cas de conflit entre leurs intérêts et ceux des autres espèces<sup>170</sup>. Ainsi, prima facie, la primauté de l'espèce humaine est inconditionnelle.

**43. À la primauté du cerveau parmi les organes.** — Pour autant, l'instrument qui organise efficacement l'ordre juridique, c'est-à-dire la loi, ne reconnaît pas un être humain au sens biologique comme un membre d'une espèce. En revanche, dans le monde juridique, une personne est en réalité définie par la loi en fonction de ses capacités<sup>171</sup>. C'est pour cela que le statut juridique d'un fœtus dépend essentiellement du stade ultérieur de sa viabilité. Son statut juridique n'est pas le même par rapport à celui du moment de sa conception. Et des questions similaires se posent aussi en fin de vie : les êtres humains en état végétatif risquent de perdre leur statut juridique par excellence tout en restant humains au niveau génétique<sup>172</sup>. De même, les enfants et les déments ne sont pas traités par notre système juridique comme des personnes juridiques par excellence dans toute hypothèse, bien qu'ils soient humains<sup>173</sup>.

---

<sup>167</sup> SINGER P., *Practical Ethics*, 2<sup>e</sup> éd., Cambridge University Press, 1993, p. 86.

<sup>168</sup> « Humains marginaux » - qui ne satisfont pas aux critères de sapience. BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », préc., p. 322.

<sup>169</sup> SINGER P., *Practical Ethics*, *op. cit.*, p.76.

<sup>170</sup> *Ibidem*, p. 58.

<sup>171</sup> DYSCHKANT A., « Legal Personhood: How We Are Getting It Wrong », *Illinois Law Review*, vol. 2015, n°5, p. 2099.

<sup>172</sup> V. p. ex. l'affaire Vincent Lambert : Lambert et autres c. France n° 46043/14, CEDH 5 juin 2015.

<sup>173</sup> CALVERLEY D. J., « Imagining a non-biological machine as a legal person », *AI & Soc*, vol. 22, 2008, p. 526.

Ils souffrent de droits et d'obligations restreints en raison de leur incapacité à démontrer un comportement qui peut indiquer leur capacité de raisonnement exigée par la société. Par conséquent, malgré l'appartenance à une même espèce, ils ne sont pas les égaux sur le plan juridique des autres. Au contraire, ces personnes ont un traitement juridique spécial, soit en vertu de la loi applicable, soit par la décision prétorienne<sup>174</sup>.

Ainsi, l'on se rend compte que c'est l'autonomie qui commande essentiellement le statut juridique d'un être. Et par autonomie, M. Peter Singer entend « la capacité de choisir, de prendre des décisions et d'agir de manière autonome »<sup>175</sup>. D'après sa catégorisation, beaucoup d'êtres sensibles sont capables d'éprouver du plaisir et de la douleur, mais ne sont ni rationnels ni conscients, et donc pas des personnes. En effet, de nombreux animaux, les nouveau-nés et certains êtres humains ayant une déficience intellectuelle entrent presque certainement dans cette catégorie. Partant, les animaux « non humains », les nourrissons et les humains gravement handicapés intellectuellement appartiennent à la même catégorie<sup>176</sup>.

Sur ces fondements, notre étude constate qu'au final, c'est l'organe traditionnellement considéré chargé de la capacité juridique qui détermine les rapports juridiques au sein de notre système juridique. Dès lors, il convient d'étudier ce que signifie l'introduction de l'IAG dans ce modèle juridique où les rapports juridiques se basent effectivement sur le critère de la capacité juridique.

**44. Le modèle des relations entre les espèces : le statut se fonde sur la capacité.** — « *Mais à mesure que le golfe s'élargissait, (...) étaient traités plutôt comme des enfants que comme des adolescents, plutôt comme des animaux domestiques bien soignés que comme des êtres humains* »<sup>177</sup>.

Aucune enquête sur notre avenir divin ne peut ignorer notre propre passé animal ou nos relations avec d'autres animaux, parce que la relation entre les humains et les animaux est le

---

<sup>174</sup> Questions tranchées par les tribunaux américains telles que le droit d'avoir et d'élever des enfants, à une éducation publique gratuite, et le droit de voter sont restreints. WILLICK M. S., « Artificial Intelligence: Some Legal Approaches and Implications », *AI Magazine*, vol. n°2, 1983, pp. 10-11.

<sup>175</sup> SINGER P., *Practical Ethics*, op. cit., p. 99.

<sup>176</sup> *Ibidem.*, pp. 60, 75, 101.

<sup>177</sup> STAPLEDON O., *Last and First Men: A Story of the Near and Far Future*, op. cit., p. 100.

meilleur modèle que l'on a pour les relations futures entre le surhumain et les humains<sup>178</sup>. En effet, les êtres améliorés sont meilleurs que nous de diverses façons : « ils sont différents de nous, si différents, en fait, qu'ils ne méritent pas d'être appelés humains »<sup>179</sup>. De ce fait, si l'on attribue les droits aux Homo sapiens dits par excellence, et que l'on réserve partiellement les droits aux animaux et déments simplement parce qu'ils sont intelligemment inférieurs, le risque potentiel est que l'IAG puisse s'accorder plus de droits que nous, car l'on est intelligemment inférieurs<sup>180</sup>.

En suivant ces raisonnements, l'on pourrait en outre noter qu'il faut tenir compte de deux principes en vue d'intégrer la discussion du statut de l'IAG au sein de la civilisation. C'est ainsi que dans un premier temps, il convient d'évoquer le principe de non-discrimination du substrat<sup>181</sup>, qui nous dit que si deux êtres ont la même fonctionnalité et la même expérience consciente, et ne diffèrent que par le substrat de leur mise en œuvre, ils ont alors le même statut moral. Il en va de même pour le principe de non-discrimination de l'ontogenèse, qui consiste à dire que si deux êtres ont la même fonctionnalité et la même expérience de conscience et diffèrent seulement par la façon dont ils sont nés, ils ont alors le même statut moral. Cela signifie que l'on devrait traiter un esprit artificiel de la même manière que l'on devrait traiter un esprit humain qualitativement identique dans une situation similaire. Cela dit, même si l'on accepte cette position, l'on doit examiner comment ces nouvelles qualités affectent le statut moral des esprits artificiels<sup>182</sup>.

Tirant ces analyses, cette étude s'adhère à l'idée selon laquelle sur le plan juridique, si une IAG atteint la même fonctionnalité ainsi que la même expérience au sens de la capacité juridique qu'un autre sujet juridique, il y aurait lieu de reconfigurer leurs statuts juridiques sur le fondement de ces qualités équivalement acquises.

**45. En résumé.** — Dans la mesure où le critère de la capacité juridique ne distingue pas effectivement la capacité juridique de l'IAG, l'intégration de l'IAG au sein de la civilisation

---

<sup>178</sup> HARARI Y. N., *Homo Deus: A Brief History of Tomorrow*, Harper Perennial, 2018.

<sup>179</sup> AGAR N., *Humanity's End: Why We Should Reject Radical Enhancement*, MIT Press, 2010, p. 17, cité in LAWRENCE D. R., BRAZIER M., « Legally Human? 'Novel Beings' and English Law », préc., p. 313.

<sup>180</sup> Harry Waterstone, à l'occasion de VIU Seminar: The Ethics of Artificial Intelligence, Lehrstuhl für Philosophie und politische Theorie, Ludwig-Maximilians-Universität München, 15-19 oct. 2018.

<sup>181</sup> BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », préc., p. 324.

<sup>182</sup> *Ibidem*.

nous semble inévitable. Car les statuts juridiques ainsi que les droits et les devoirs qui en découlent ainsi sont essentiellement fondés sur la notion de la capacité juridique. Par conséquent, lorsqu'une IAG atteint la capacité juridique, il en résulte que ce même critère de la capacité juridique va rendre cet aspect de l'IAG comparable à d'autres entités sur le plan juridique. Partant de ces observations, l'on va à présent étudier l'implication de la sélection des êtres intelligents.

## §2 La sélection des êtres intelligents

« *What nature does blindly, slowly, and ruthlessly, man may do providently, quickly, and kindly. As it lies within his power, so it becomes his duty to work in that direction.* » —

Francis Galton

**46. Construction des êtres dits intelligents.** — La théorie informatique de l'esprit a été déjà suggérée par Thomas Hobbes, qui a écrit que : « *Par ratiocination, je veux dire par calcul* »<sup>183</sup>. « Car RAISON, en ce sens, n'est rien d'autre que le calcul (c'est-à-dire l'addition et la soustraction) des conséquences des noms généraux convenus, pour marquer et signifier nos pensées »<sup>184</sup>.

À cet égard, d'après M. Yann Le Cun, un système intelligent peut ainsi être réduit à trois modules essentiels : un module de perception qui essaie d'estimer l'état du monde à partir de ses perceptions (une estimation toujours plus ou moins inexacte); un module agent, qui produit des actions et agit sur le monde; et un module objectif qui calcule si, d'une certaine manière, l'agent est dans un état « heureux » (satisfait) ou pas. Ce troisième module calcule le plaisir ou en cas d'échec, la tristesse, ou même la douleur<sup>185</sup>. Ce sont dans leurs activités quotidiennes que les animaux expriment toute la complexité de leur répertoire comportemental, quand ils se déplacent, se nourrissent et interagissent socialement<sup>186</sup>.

C'est pourquoi M. Yann Le Cun ne doute pas que les machines intelligentes autonomes auront un jour des émotions, car bon nombre de nos pulsions et émotions ont été construites

---

<sup>183</sup> HOBBS, *Elements of Philosophy*, 1655, Section 4.

<sup>184</sup> HOBBS, *Leviathan*, 1651.

<sup>185</sup> LE CUN Y. et alii, *La plus belle histoire de l'intelligence*, Robert Laffont, 2018, pp. 201-202.

<sup>186</sup> VAUCLAIR J., *L'intelligence de l'animal*, Seuil, 2017, p. 27.

en nous par l'évolution pour la survie de notre espèce et de nos gènes. Il explique que quand la composante de l'objectif qui mesure la charge de la batterie du robot produit un coût élevé et provoque la recherche d'une prise de courant, ou que si le robot prédit, grâce à son modèle du monde, qu'il va tomber et s'endommager, il va alors anticiper la douleur calculée et tenter de planifier une trajectoire qui évite ce résultat, et que cela est assimilable à une sensation de faim ou de peur. De même, comme l'émotion est l'anticipation du coût calculée par le module critique, lorsque la machine évite une action parce qu'elle aboutit à un coût élevé ou lorsqu'elle effectue une action parce qu'elle va aboutir à un coût faible, cela constitue déjà la marque d'une émotion<sup>187</sup>.

En suivant ces démonstrations, l'on pourrait adhérer à la position selon laquelle les machines ne seront plus prédéterminées comme c'est le cas de nos jours. À l'inverse, en éprouvant des émotions ou de l'empathie, elles gagneront en autonomie et disposeront d'une liberté d'action qui est aujourd'hui une des pierres angulaires de l'intelligence humaine<sup>188</sup>. Dans ce sens, l'IA de Turing fait aussi l'objet de l'évolution naturelle de Darwin.

**47. De la sélection aléatoire de Darwin à la sélection intelligente de Turing.** — Dans *l'Origine des espèces*, pour expliquer la sélection naturelle, Darwin a utilisé la métaphore de la sélection artificielle. D'après Darwin, l'homme, comme animal, peut fonctionner comme des forces sélectives dans l'évolution : « Le grand pouvoir de ce principe de sélection n'est pas hypothétique. Il est certain que plusieurs de nos éminents éleveurs ont modifié, même d'une seule vie, certaines espèces de bovins et de moutons » a-t-il écrit. Il a cité un passage de William Youatt, qui a remarqué que ce principe de sélection « permet à l'agriculteur, non-seulement de modifier le caractère de son troupeau, mais de le transformer entièrement »<sup>189</sup>. Darwin estimait ainsi que la variation sous domestication, trouve désormais son application généralisée chez les humains, et n'est plus réservée aux plantes et animaux tels que le blé ou le pigeon.

Quant à l'apprentissage des machines dans le cours de l'évolution, comme l'a prévu Turing : « L'on peut toutefois espérer que ce processus sera plus rapide que l'évolution. La survie du

---

<sup>187</sup> LE CUN Y., *Quand la machine apprend*, Odile Jacob, 2019, pp. 367-368.

<sup>188</sup> KLEIN É., « Quatre questions à Yann LeCun », préc., p. 46.

<sup>189</sup> DARWIN CH., *On the Origin of Species*, op. cit., p. 34.

plus apte est une méthode lente pour mesurer les avantages. L'expérimentateur, par l'exercice de son intelligence, devrait pouvoir l'accélérer. Le fait qu'il ne soit pas limité aux mutations aléatoires est tout aussi important. S'il peut trouver la cause d'une faiblesse, il peut probablement penser au type de mutation qui l'améliorera »<sup>190</sup>. En effet, les humains ont été créés par l'évolution, un processus inintelligent. C'est une espèce évoluée avec une structure cognitive et une architecture émotionnelle adaptée aux contextes de vie chasseurs-cueilleurs et aux processus cognitifs adaptés pour fonctionner sur un substrat de neurones biologiques. En revanche, l'IA est créée par des processus intelligents qui sont les humains<sup>191</sup>.

Ainsi, pour certains chercheurs d'IAG, leur but n'est pas de créer un humain simulé, mais plutôt un système intelligent numérique non humain — celui qui complètera l'intelligence humaine en procédant à l'analyse de données et des tâches de gestion bien au-delà des capacités de l'esprit humain<sup>192</sup>. L'effet espéré étant la subdivision à terme du genre humain en deux espèces distinctes (humains avec ou sans IA)<sup>193</sup>. À notre sens, ce résultat pourrait déjà commencer par la perte des compétences chez les humains.

**48. De la décision déléguée à la perte des compétences cognitives.** — Le Parlement européen remarque que « l'IA restera un outil utile pour compléter l'action humaine et pour améliorer ses performances et réduire les erreurs »<sup>194</sup>, mais « en réduisant le taux d'erreur humaine dans le secteur des transports, le système est susceptible de devenir plus efficace, avec moins d'accidents, grâce à des évaluations plus claires et à la valeur prédictive de la technologie, à la réduction des retards, à la cartographie des schémas de circulation et à l'exploitation des services dans les délais, ainsi qu'à des économies plus importantes, avec moins de fautes imputables aux conducteurs et des processus internes rationalisés »<sup>195</sup>. En effet, dorénavant, lorsqu'un ordinateur nous demande de cocher des images afin de vérifier que l'on n'est pas un robot, c'est l'inverse qui se produit, il s'agit d'un test de Turing, mais

---

<sup>190</sup> TURING A. M., « Computing Machinery and Intelligence », préc., p. 456.

<sup>191</sup> YUDKOWSKY E., « Levels of Organization in General Intelligence », in GOERTZEL B., PENNACHIN C. (éds.), *Artificial General Intelligence*, op. cit., pp. 391-392.

<sup>192</sup> GOERTZEL B., PENNACHIN C., « The Novamente Artificial Intelligence Engine », in GOERTZEL B., PENNACHIN C. (éds.), *Artificial General Intelligence*, op. cit., p. 73.

<sup>193</sup> V. infra Le cyborgisation dans les réalités diverses.

<sup>194</sup> Résolution préc. du 12 févr. 2019, pt. 152.

<sup>195</sup> *Ibidem*, pt. 90.

de version humaine<sup>196</sup>. Et éventuellement, au fur et à mesure, l'on est de plus en plus dans une autonomie diminuée, et l'IA est de moins en moins dans une autonomie restreinte, à un certain point, l'on se retrouve au même rang et l'on se situe au même degré par rapport à notre autonomie ; l'on est diminué, ils sont augmentés, l'autonomie se croise au même niveau, avant de s'éloigner.

Vu que l'amélioration de la fiabilité des véhicules sans conducteur s'accompagnera d'une perte de compétences des conducteurs humains, ces occasions de pratiquer et même d'apprendre à conduire vont disparaître. Bientôt, il deviendra répréhensible de conduire une voiture lorsque des voitures automatisées, débarrassées de toute erreur humaine, réduiront les risques pour les autres. Parce que le conducteur humain sera considéré comme l'équivalent d'un robot ivre<sup>197</sup>, imposer le risque accru à des tiers deviendra un argument contre les conducteurs humains et sera traduit en droit : la conduite constituera alors un comportement illégal. De même, étant donné que la quantité de données recueillies par les examens médicaux dépasse de loin la capacité humaine, l'omission d'utiliser l'IA dans le diagnostic pourrait éventuellement constituer une faute professionnelle.

**49. En résumé.** — Aujourd'hui, les humains recourent à l'IA dans les décisions relevant des domaines médical, financier, voire judiciaire<sup>198</sup>. À l'instar de l'exemple des voitures autonomes, les risques posés par les erreurs humaines sont susceptibles de conduire la même évolution dans toutes les dimensions au sein de la civilisation. De ce point de vue, les vrais soucis à ce stade ne sont pas que les ordinateurs deviennent trop intelligents, mais que les hommes sont trop inaptes<sup>199</sup>. À cet égard, afin de diminuer le risque et maximiser le bien-être des humains, la substitution de l'homme par l'IA dans de nombreux contextes devient inéluctable.

---

<sup>196</sup> A Discussion of Artificial Intelligence with John Searle and Luciano Floridi, The New York Review of Books Foundation and Fritt Ord hosted the conference « Technology and the Human Future », 20-21 oct. 2016.

<sup>197</sup> SPARROW, R., HOWARD, M., « When human beings are like drunk robots: driverless vehicles, ethics, and the future of transport », Transportation Research Part C, 9 mai 2017, p. 4.

<sup>198</sup> V. infra § L'intégration de l'IA dans la civilisation : de l'objet surhumain.

<sup>199</sup> DOMINGOS P., *The Master Algorithm*, op. cit., p. 285-286. Déjà, de manière plus générale, les effets de Google et de Wikipédia sur notre mémoire et ceux d'Internet sur notre cerveau accroissent notre vulnérabilité.

## Conclusion de la section 1

**50. De la primauté de l'espèce humaine à la reconfiguration de la hiérarchie de l'intelligence.** — L'on vient de voir que désormais, il n'est plus possible d'agir de manière autonome, car les compétences requises pour cela, qui doivent être constamment exercées, auront disparu<sup>200</sup>. En conséquence, comme l'écrit M. Stuart Russell : « Pour tout individu humain, il peut sembler inutile de s'engager dans des années d'apprentissage ardu pour acquérir des connaissances et des compétences que les machines possèdent déjà ; mais si tout le monde pense de cette façon, l'espèce humaine, collectivement, perdra son autonomie »<sup>201</sup>.

Certes, la primauté de l'espèce humaine semble *prima facie* inconditionnelle pour le système juridique. Néanmoins, à un regard de plus près, l'on a démontré qu'au final, c'est la faculté de l'intelligence — traduit en droit, la capacité juridique — qui détermine les rapports juridiques des êtres sur le plan juridique. Dorénavant, le développement technologique est irréversible, surtout en ce qui concerne la perte de compétences cognitives humaines. Ainsi, face à une intelligence supérieure à la nôtre au cours de l'évolution naturelle, il convient à présent d'étudier ce caractère exceptionnel de l'IA forte.

---

<sup>200</sup> « Automatisiertes Und Vernetztes Fahren », *op. cit.*, p. 22.

<sup>201</sup> RUSSELL S., *Human Compatible*, *op. cit.*, p. 255. James H. Moor considère que l'informatisation accrue au sein de notre société pourrait entraîner une déshumanisation de nos vies. En revanche, si la délégation du pouvoir de décision est faite de manière responsable, l'on pourrait créer une société beaucoup plus humaine. Certaines décisions les plus humanistes pourraient provenir de décideurs qui ne sont pas les humains. MOOR J. H., « Are There Decisions Computers Should Never Make? », *Nature & System I*, janv.1985, p. 227.

## **Section 2 L'affirmation de la nature exceptionnelle de l'IA forte**

**51. Annonce du plan.** — Une fois le caractère exceptionnel de l'IAG mis en évidence, reste à examiner ce phénomène dit Robot Exceptionalism (§1), qui est, d'une manière ou d'une autre, éprouvé par les incohérences juridiques remarquées entre l'humain et l'IA (§2).

### **§1 L'exception robotique**

**52. Le Robot Exceptionalism et la faillite philosophique.** — L'IA est désormais classée une des technologies d'usage général (General purpose technology, GPTs). Cela veut dire que l'IA est une technologie qui peut affecter une économie entière, au niveau national voire mondial, et qui a le potentiel de modifier radicalement les sociétés par son impact sur les structures économiques et sociales préexistantes<sup>202</sup>. Dès lors, la robotique fait partie de la *Robot Exceptionalism*, qui consiste en l'idée qu'« une technologie est exceptionnelle si elle invite à un changement systémique des lois ou des institutions juridiques afin de préserver ou rééquilibrer les valeurs établies »<sup>203</sup>.

Dans cette situation sans précédent, comme la question concernant l'IAG contemporaine paraît inconcevable pour les ancêtres, dans une large mesure, une faillite philosophique<sup>204</sup> peut être observée : ce que l'on a fait pendant des milliers d'années ne nous a pas préparés au genre de défi de l'IAG que l'on a aujourd'hui. Pourtant, quant à la régulation de l'IA, plusieurs arguments ont été invoqués pour dire le contraire, tels que l'argument de Law of the Horse ainsi que l'argument de la neutralité technologique.

**53. L'argument du « Law of the Horse ».** — Déjà, en matière de cyberspace, d'aucuns proposent une réflexion particulière sur les liens entre le droit et le cyberspace, permettant de dégager des points généraux importants<sup>205</sup>. Par exemple, à l'époque, cette proposition souligne qu'il pourrait y exister une différence de protection quant à l'aspect de la vie privée

---

<sup>202</sup> La définition de « General purpose technology » selon Wikipedia, page consultée le 21 déc. 2019.

<sup>203</sup> CALO R., « Robotics and the Lessons of Cyberlaw », California Law Rev., vol. 103 Issue 3, 2015, p. 552.

<sup>204</sup> Yuval Noah Harari, lors de la discussion avec Max Tegmark, FLI Podcast, 31 déc. 2019.

<sup>205</sup> LESSIG L., « The Law Of The Horse: What Cyberlaw Might Teach », Harv. Law Rev., vol. 113 issue 501, 1999.

en cyberspace. Cela signifie qu'il y aurait lieu de distinguer le suivi dans le cyberspace du suivi dans l'espace physique, étant donné qu'une telle surveillance dans le cyberspace n'est pas visible ni vérifiable de la même manière qu'en espace physique. Sur ce point, l'on pourrait déjà penser aux dispositions contemporaines comme le RGPD pour confirmer le déroulement de cette constatation<sup>206</sup>. Cependant, malgré les défis sans précédent du cyberspace, l'idée de Law of the Horse soutient qu'il n'y avait aucune raison d'enseigner la loi du cyberspace, pas plus qu'il n'y avait de raison d'enseigner la Loi du cheval, parce que cela ne mettrait pas en lumière la loi dans sa totalité<sup>207</sup>.

Cet argument de Law of the Horse a ainsi été transposé en matière robotique, pour dire que bien que la robotique présente des caractéristiques distinctives par rapport à d'autres technologies, qui nécessitent une attention particulière, il ne faut néanmoins pas abuser d'une attitude qui relèverait de l'exceptionnalisme robotique : « il semble n'y avoir aucun besoin de remodeler la tout l'environnement juridique pour s'adapter à la robotique, ni pour réagir aux nombreux défis qui se posent dans ce domaine avec une approche globale et homogène de lois les couvrant tous »<sup>208</sup>. D'après cet argument, vu que de nombreuses applications robotiques peuvent facilement être situées et réglementées dans les cadres actuels, il serait impossible de développer des règles applicables à toutes les applications robotiques indifféremment. En outre, comme « de nombreux régimes applicables à la robotique sont déjà en place, abordant des aspects tels que la sécurité, la sûreté, la responsabilité, ou le traitement de données personnelles, de nouvelles réglementations ne sont donc pas nécessaires ou du moins ne devraient pas commencer à partir de zéro »<sup>209</sup>.

**54. L'argument du « Law of the Horse » écarté.** — Pour autant, au sens de cette étude, cette compréhension soutenant l'argument du Law of the Horse reflète en réalité la régulation d'intelligence artificielle étroite telle que les voitures autonomes et le traitement

---

<sup>206</sup> V. aussi infra sur la stagnation éthique.

<sup>207</sup> « De nombreuses affaires portent sur la vente de chevaux, d'autres sur des personnes frappées par des chevaux, d'autres encore sur l'octroi de licences et les courses de chevaux, sur les soins apportés aux chevaux par les vétérinaires ou sur les prix des concours de chevaux. Tout effort visant à rassembler tous ces éléments dans un cours sur 'la loi du cheval' est condamné à être superficiel et à manquer de principes unificateurs ». EASTERBROOK F. H., « Cyberspace and the Law of the Horse », *University of Chicago Legal Forum* 207, 1996.

<sup>208</sup> PALMERINI E., « Towards a Robotics law at the EU level? », in HERVÉ J., DE STREEL A., (dir.), *L'intelligence artificielle et le droit*, Larcier, 2017, pp. 59-60.

<sup>209</sup> *Ibidem.*, p. 68.

des données personnelles, et ne relève pas des enjeux de l'IA générale et forte évoqués plus haut. En effet, la nécessité de distinguer entre l'IA étroite et l'IA générale pour écarter l'application de l'argument de Law of the Horse en matière d'IAG a été bel et bien soulignée dans la résolution du Parlement européen, où il est constaté que « l'IA est une notion qui englobe un large éventail de produits et d'applications, à commencer par l'automatisation, les algorithmes, l'intelligence artificielle étroite et l'intelligence artificielle générale »<sup>210</sup>. Par conséquent, le Parlement estime qu'il faut « mettre au point un cadre stratégique qui encourage la conception de tous types d'intelligence artificielle, et pas uniquement celle de systèmes fondés sur l'apprentissage profond, particulièrement gourmand en données »<sup>211</sup>. Partant de ces considérations, l'on voit que le souci quant à l'inflation législative a donc d'ores et déjà été évalué par l'UE, qui appelle tout de même à la mise en place d'un ensemble unique de règles pour réglementer l'IA étroite<sup>212</sup>.

Pour autant, malgré ces démonstrations sans équivoque, la nature technologique de l'IA pourrait attirer un autre argument, communément appelé sous le nom du principe de la neutralité technologique, pour s'opposer à sa régulation.

**55. L'argument du principe de la neutralité technologique.** — Cet argument consiste à dire qu'« un seul projet de loi traitant avec beaucoup ou tous les types d'objets qui se rapportent à la même technologie est condamné à ne pas être durable, compte tenu du principe de la neutralité technologique »<sup>213</sup>. En effet, cela reflète l'idée de Law Of the Horse, mais spécialement pour la régulation de la technologie, qui consiste à dire que s'il y a lieu d'établir une distinction, l'on distingue, et que si ce n'est pas le cas, la régulation ne devrait pas tenir compte du type de technologie utilisée. Effectivement, étant donné le rythme croissant de l'innovation, de nombreux régimes juridiques ont adopté une rédaction générale en vertu du principe de la neutralité technologique qui repose sur l'hypothèse qu'elle favorise les longévités et traitement égaux des technologies anciennes et nouvelles<sup>214</sup>. Tandis que les dispositions spécifiques à la technologie se réfèrent à la technologie, l'objectif de la

---

<sup>210</sup> Résolution préc. du 12 févr. 2019, pt. 116.

<sup>211</sup> *Ibidem*, pt. 117.

<sup>212</sup> P. ex. le projet de règlement sur l'intelligence artificielle de la Commission européenne : proposition de règlement établissant des règles harmonisées sur l'intelligence artificielle, 21 avr. 2021.

<sup>213</sup> PALMERINI E., « Towards a Robotics law at the EU level? », préc., p. 60.

<sup>214</sup> GREENBERG B. A., « Rethinking Technology Neutrality », *Minnesota Law Review*, vol. 100, 2016, pp. 1495, 1511-1512.

neutralité technologique est de réguler le comportement, pas la technologie; de s'inquiéter de ce qui se passe, pas de comment cela se produit. Les lois neutres en matière de technologie utilisent donc des termes généraux, vagues, ouverts, qui spécifient les objectifs, les effets, les fonctions et autres caractéristiques générales<sup>215</sup>.

**56. L'essence de la neutralité technologique.** — La Directive européenne 2002/21 du 7 mars 2002 relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques fournit une définition : « L'exigence pour les États membres de veiller à ce que les autorités réglementaires nationales tiennent le plus grand compte du fait qu'il est souhaitable que la réglementation *technologique soit neutre, c'est-à-dire qu'elle ne soit ni contraignante ni discriminatoire au bénéfice de l'utilisation d'un type particulier de technologie*, n'empêche pas que des mesures proportionnées soient prises afin de promouvoir certains services spécifiques »<sup>216</sup>. Ainsi, le principe de la neutralité technologique signifie que la règle est destinée à *couvrir toutes les situations* de fait, *indépendamment de la technologie ou du support utilisé*. À cette fin, les règles sont « neutres », c'est-à-dire qu'elles *ne présupposent pas l'utilisation de types particuliers de technologies*<sup>217</sup>.

**57. La contradiction de la neutralité technologique.** — En effet, la notion de neutralité technologique a été introduite dans les nouvelles lois qui ont été adoptées pour encadrer les nouvelles technologies *afin de « favoriser le développement du commerce électronique »*<sup>218</sup>. Elle est donc, avant tout, « *une méthode de rédaction législative* »<sup>219</sup>. Justement, la neutralité technologique peut se résumer par une technique législative, conformément à l'expression : « *par quelque moyen que ce soit* »<sup>220</sup>. « *La neutralité technologique n'est donc pas un principe nouveau ni, peut-être, un principe général. Il s'agit simplement d'une technique rédactionnelle, particulière, précise, favorisant l'adaptation au temps et ne méritant peut-*

---

<sup>215</sup> OHM P., « The Argument Against Technology-Neutral Surveillance Laws », *Texas Law Review*, vol. 88, 2010, p. 1687, cité *in ibidem*.

<sup>216</sup> Considérant 18 de la Directive européenne 2002/21 du 7 mars 2002 relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques.

<sup>217</sup> §47 de la Convention sur l'utilisation de communications électroniques dans les contrats internationaux, ONU, 2007.

<sup>218</sup> GAUTRAIS V., *Neutralité technologique : rédaction et interprétation des lois face aux technologies*, *Thémis*, 2012, p. 36.

<sup>219</sup> *Ibidem*, p. 40.

<sup>220</sup> *Ibid.*, p. 51.

*être pas d'être reconnue avec cette originalité comme il l'a été par la doctrine ou par la loi* »<sup>221</sup>. Car tout comme en matière pénale et d'autres branches juridiques, ces lois sont déjà indifférentes à la technologie impliquée : elles s'appliquent au comportement des acteurs impliqués et aux effets du comportement. Autrement dit, cette notion de « par quelque moyen que ce soit » se trouve déjà bel et bien dans la technique législative en général. Prenons l'exemple classique du code pénal disposant que le fait de donner volontairement la mort à autrui constitue un meurtre. Cette disposition ne prévoit aucun moyen ni support pour l'acte. En d'autres termes, la disposition ne tient pas compte de la façon constituant l'acte d'un meurtre. Il y a bien une « neutralité » impliquée dans la condition, sans pour autant une appellation de « la neutralité meurtrier ». Revenons dans le cas de la neutralité technologique, la notion de cette neutralité représente ainsi une sorte de contradiction en soi : car ce qu'elle implique, c'est tout simplement, « *par quelque moyen que ce soit* ». Mais en se revêtant une « exception » pour « technologie » et ne retournant pas au principe général de technique législative, cette neutralité proclamée contredit au fond même sa propre idée proposée. En outre, sur le fondement de la neutralité technologique, le législateur pourrait avoir la fausse impression que la loi n'a pas besoin d'être mise à jour lorsqu'elle doit néanmoins distinguer les différentes technologies. De ce fait, en matière d'IAG, le législateur pourrait commettre une omission par son refus d'exercer le pouvoir législatif d'appréciation du caractère exceptionnel de l'IAG et de mettre à jour la législation adéquate par suite.

En tout état de cause, si la neutralité a pour vocation d'éviter les distinctions des technologies largement injustifiables, la nature de l'IAG se différencie d'autres technologies et ne tolère aucune hésitation. À vrai dire, qu'il s'agisse de la technologie ou non, toute exception devrait être soumise aux raisonnements respectifs afin de distinguer celles qui devraient être renvoyées au droit commun ou aux dispositifs existants et celles qui ne devraient pas. Partant, comme l'IAG constitue toute dimension comparable et au-delà de l'humanité, et que son développement contemporain n'a commencé qu'il y a moins d'un siècle, au sens de cette étude, il y aurait lieu de tenir compte de la stagnation législative dans le contexte de l'IAG.

**58. En résumé.** — Ces fondements, que ce soit Law of the Horse, ou la neutralité technologique, ou encore l'éthique homologique, sont des arguments qui ne concernent que

---

<sup>221</sup> *Ibid.*, p. 52.

la régulation d'IA étroite de nature sectorielle. Car ces arguments cherchent à aborder les questions de l'IA en la considérant comme un simple « outil » ou un autre « moyen » glissant sur le paradigme juridique jusqu'à présent. Cette qualification pourrait se justifier dans le contexte de l'IA étroite, mais elle s'avère erronée pour l'IAG qui va refondre même l'espèce humaine.

De là, l'on peut dire que les qualités essentielles de l'IA soutiennent son caractère exceptionnel car elles sont assez différentes pour provoquer de vastes modifications du droit, au moins à un aussi grand degré que l'a fait Internet<sup>222</sup>. Sachant que l'internet reste au sein de l'humanité, l'IAG, la vraie intelligence artificielle qui ne se limite pas à l'approche axée sur les données et l'apprentissage automatique, quant à elle, est par définition un développement qui englobe l'humanité ; a fortiori, sa nature singulièrement exceptionnelle est dès lors incontestable. En effet, plusieurs incohérences juridiques peuvent déjà être constatées entre l'humain et l'IA.

## §2 Les incohérences juridiques entre l'humain et l'IA

*Peter sees the computer. "But the machine only creates what humans have taught it to," says Peter. "So do you," says Mummy. — Mario Klingemann*

**59. Annonce du plan.** — Une fois que le caractère exceptionnel de l'IAG est démontré, il reste à évaluer son implication sur le plan juridique. Pour ce faire, à titre d'illustration, cette étude propose de s'interroger brièvement sur la contradiction du droit constant à travers les questions telles que l'imputabilité (I), l'humain outil sur le plan juridique (II), la théorie de l'apparence (III) ainsi que l'opacité commune des deux types d'intelligence (IV).

### **I. L'incohérence de l'imputabilité**

**60. La notion d'imputabilité.** — Traditionnellement, la condition *sine qua non* pour la reconnaissance éventuelle de la responsabilité est l'imputabilité. Il convient de trancher ainsi la difficulté d'imputabilité pour étudier la responsabilité autour de l'IA. En effet, l'on sait qu'à l'heure actuelle, « faute de disposer de la capacité de discernement d'une personne humaine, le comportement d'un robot intelligent ne peut être apprécié à l'aune du standard

---

<sup>222</sup> CALO R., « Robotics and the Lessons of Cyberlaw », préc., p. 558.

du bon père de famille. Il n'est donc pas susceptible d'engager la responsabilité du robot »<sup>223</sup>. Pour autant, vu que la notion d'imputabilité se fonde sur la capacité de discerner d'un cerveau, avec la technique d'identification de plus en plus précise, une éventuelle cohérence du transfert du pouvoir de décision aux algorithmes au transfert de la responsabilité aux algorithmes ne semble pas si éloignée. Car dans un premier temps, l'on pourrait déjà remarquer qu'il est possible d'engager la responsabilité des humains privés de discernement.

### **61. L'engagement de la responsabilité humaine même en l'absence de discernement. —**

En effet, pendant longtemps, l'exigence du discernement, élément subjectif, était requise avec l'acte objectif pour engager la responsabilité de l'auteur. Cependant, la loi française est intervenue en 1986 dans un premier temps, pour abandonner cette exigence de discernement dans la responsabilité civile du majeur atteint d'un trouble mental<sup>224</sup>. Il est ainsi prévu aujourd'hui à l'article 414-3 du Code civil français que « celui qui a causé un dommage à autrui alors qu'il était sous l'empire d'un trouble mental n'en est pas moins obligé à réparation ».

Et puis, toujours dans un même souci d'une meilleure indemnisation des victimes, les arrêts tels que Lemaire et Derguini ont également, par voie prétorienne, opéré un revirement de jurisprudence en faisant disparaître l'élément subjectif de la faute quant à l'infans privé de discernement<sup>225</sup>. Cet abandon de l'exigence de discernement du gardien peut d'ailleurs être retrouvé dans les arrêts Trichard à l'égard de la responsabilité du dément<sup>226</sup>, et Gabillet pour l'infans, en l'espèce un enfant en bas âge de trois ans<sup>227</sup>.

Pour autant, si la capacité de discerner, de distinguer le bien du mal, peut être écartée pour retenir l'imputabilité ainsi que la responsabilité, c'est, en réalité, exactement ce que l'on inculque aux machines à apprendre : être capable de discerner ce que l'on veut et en sortir le résultat espéré. De ce point de vue, même ceux qui ne sont pas capables d'apprendre et comprendre les choix peuvent être tenus responsables. Ainsi, c'est assez fascinant

---

<sup>223</sup> JACQUEMIN H., HUBIN J.-B., « La responsabilité extracontractuelle du fait des robots ou des applications d'intelligence artificielle », in HERVÉ J., DE STREEL A. (dir.), *L'intelligence artificielle et le droit*, op. cit., p. 117.

<sup>224</sup> La loi n° 68-5 du 3 janvier 1968 portant réforme du droit des incapables majeurs.

<sup>225</sup> Cass. Ass. plén., 9 mai 1984 : Derguini n°80-93481 et Lemaire n° 80-93031.

<sup>226</sup> Cass. civ. 2<sup>e</sup>, 18 déc. 1964

<sup>227</sup> Cass. Ass. plén., 9 mai 1984, n° 80-14994

d'apprendre « l'irresponsabilité » quant à l'apprentissage de la machine, qui est, dans tous les cas, créé dans le seul but d'apprendre à discerner. Dans cette perspective, dès que l'évolution de l'IA prendra son essor et passe de l'IA étroite à l'IA générale voire forte, il va falloir chercher des justifications pour exonérer l'imputabilité de l'IAG au regard de l'incohérence du critère de discernement. En particulier lorsque cette difficulté se trouve également en droit pénal.

**62. La responsabilité pénale d'IA.** — L'on sait que pour qu'une personne soit pénalement responsable, deux éléments principaux doivent exister. Le premier est l'élément externe ou factuel, c'est-à-dire le comportement criminel (*actus reus*), et le second, l'élément interne ou mental, c'est-à-dire la connaissance ou l'intention générale vis-à-vis de l'élément comportemental (*mens rea*). D'aucuns considèrent ainsi que si l'IA les remplit effectivement, rien n'empêche d'imposer une responsabilité pénale à l'IA. En outre, la justification de la défense de l'infans et des déments est *doli incapax* : ils sont incapables de comprendre ce qui est mal dans leur conduite ou sont présumés ne pas concevoir l'élément fautif de l'infraction en raison de leur incapacité mentale. Partant, l'on voit qu'il n'y a pas de différence matérielle entre les humains et l'IA puisque la responsabilité se rapporte à une situation spécifique (*in rem*), quelle que soit l'identité de l'auteur de l'infraction<sup>228</sup>.

Donc l'on voit qu'il s'agisse du droit civil ou du droit pénal, la responsabilité se fonde essentiellement sur l'imputabilité qui elle-même dépende de la capacité mentale de l'agent. Mais c'est justement la capacité mentale que l'intelligence, artificielle ou non, cherche à saisir et intervenir. C'est donc à juste titre, il faut également s'interroger sur l'implication de l'implant cérébral<sup>229</sup> quant aux notions même de *actus reus* et de *mens rea*. En tout état de cause, cette interchangeabilité de la capacité mentale de l'agent se manifeste encore plus explicitement dans la notion même de l'agent au sens juridique.

## II. L'incohérence de l'agent humain sur le plan juridique

---

<sup>228</sup> HALLEVY G., « The Criminal Liability of Artificial Intelligence Entities - From Science Fiction to Legal Social Control », *Akron Intellectual Property Journal*, vol. 4, 2010.

<sup>229</sup> V. infra § Le cyborgisation dans les réalités diverses.

**63. L'interchangeabilité des agents.** — Le mot agent vient du latin *agere*, agir ou faire. Un agent est juste quelque chose qui agit. En effet, les agents informatiques sont aussi supposés agir de manière autonome<sup>230</sup> : un agent est donc tout ce qui peut être perçu comme percevant son environnement à l'aide de capteurs (sensors) et qui agit dans cet environnement à l'aide d'actionneurs (actuators)<sup>231</sup>.

**64. L'outil humain reconnu en droit.** — Ce qui est intéressant de noter, c'est qu'en matière d'agent au sens juridique, la loi ne différencie pas la situation entre un commettant et un simple outil ou instrument mécanique. En revanche, elle présuppose une fonction équivalente : lorsqu'un être humain effectue pour un autre une tâche qui ne relève pas de l'exercice d'un jugement, la loi ne considère pas cet homme comme un agent. Cet être humain sert à aider dans l'exécution de certaines tâches purement mécanique, comme signer le nom de l'autre, apposer son sceau, etc. Cet humain n'est pas engagé à l'acte juridique en apportant sa propre volonté, son jugement ; en revanche, sa détermination dans tous les domaines qui concernent l'essence de l'acte juridique relève d'un autre. L'acte est considéré en droit comme l'acte direct et personnel de ce dernier, et la personne humaine qui a assisté de manière représentative ou mécanique est réputée être un simple outil ou instrument<sup>232</sup>. Partant, l'on comprend que les êtres humains peuvent, réciproquement, être considérés comme simple outil sur le plan juridique.

De même, lorsqu'un être humain est incapable d'exercer son jugement et sa compréhension, cette personne ne peut pas être un agent<sup>233</sup>. Il a longtemps été reconnu qu'une personne *non compos mentis* ne peut accomplir aucun acte en tant qu'agent ou mandataire car elle n'a aucune discrétion ou compréhension pour conférer aux affaires des autres, pas plus que sur leurs propres affaires<sup>234</sup>. Ainsi, humain ou non, la capacité juridique est le facteur décidant la qualité d'une entité sur le plan juridique. Par conséquent, cette fonctionnalité équivalente peut même nous amener à considérer l'IA comme autrui.

---

<sup>230</sup> RUSSELL S., NORVIG P., *Artificial Intelligence - A Modern Approach*, *op. cit.*, p. 4.

<sup>231</sup> *Ibidem*, pp. 34-35.

<sup>232</sup> MECHANI F. R., *Treatise on the law of agency*, 2<sup>e</sup> éd., 1982, cité in BELLIA A. J., « Contracting with Electronic Agents », 50 *Emory L. J.*, 2001, p. 1063.

<sup>233</sup> GREGORY W. A., REUSCLEIN H. G., *The Law Of Agency And Partnership*, 2<sup>e</sup> éd., 1990, pp. 22-23., cité in BELLIA A. J., « Contracting with Electronic Agents », préc., p. 1063.

<sup>234</sup> STORY J., *Commentaries On The Law Of Agency*, 4<sup>e</sup> éd., 1851, p. 9, cité in BELLIA A. J., « Contracting with Electronic Agents », préc., p. 1063.

**65. L'autrui IA.** — L'une des options proposées pour faire face aux risques liés à l'IA est l'élargissement potentiel de la notion de responsabilité du fait d'autrui : en laissant intact le régime de responsabilité d'autrui, mais en l'étendant, soit directement, soit par analogie, à des situations fonctionnellement équivalentes où l'on utilise une IA au lieu de recourir à un auxiliaire humain. L'illustration d'un hôpital utilisant un chirurgien sous forme d'IA, malgré sa conformité à toutes les obligations de soins possibles, si un patient est blessé parce que le robot fonctionne mal de façon que personne n'aurait pu prévoir, dans ce cas, l'hôpital devrait tout de même être responsable en vertu du principe d'équivalence fonctionnelle<sup>235</sup>.

Mais si la responsabilité d'autrui d'IA implique déjà cette logique comparable voire identique entre l'IA et l'homme, dans l'hypothèse où l'on considère implicitement l'IA comme « autrui », il y aurait lieu d'admettre son équivalence sous diverses formes juridiques. Car au fur et à mesure, dans un monde où l'on doit se demander constamment si une image, un article, une chanson, une vidéo voire un agent provient d'une IA, il est nécessaire de prendre en compte la théorie de l'apparence afin de vérifier l'authenticité des agents IA, tout comme la solution déjà retenue lorsque l'on interagit avec un agent humain.

### III. L'incohérence de la théorie de l'apparence

**66. Vérification de l'authenticité des agents.** — Rappelons que la théorie de l'apparence, qui s'appelle aussi la théorie de la confiance légitime, constitue une source autonome d'obligations<sup>236</sup>. Au sens de cette étude, dans la mesure où l'on vit dans une banalisation des applications d'IA, afin de pouvoir identifier et vérifier l'authenticité d'un agent, il y aurait lieu d'appliquer la théorie de l'apparence au regard des enjeux soulevés par, à titre d'illustration, le cas d'assistant personnel intelligent (virtual personnel assistant VPA), ou d'autres applications en cas d'identité déguisée, tel le cas du lancement initial de Lil Miquela<sup>237</sup>. En effet, cette exigence a pour but d'obliger à la précision de l'interaction

---

<sup>235</sup> PAGALLO U. et alii, *Liability for Artificial Intelligence and other emerging digital technologies*, European Commission, 2019, pp. 45-46.

<sup>236</sup> JACQUEMIN H., HUBIN J.-B., « Le robot et l'intelligence artificielle en droit des obligations contractuelles » in HERVÉ J., DE STREEL A. (dir.), *L'intelligence artificielle et le droit*, op. cit., p. 111.

<sup>237</sup> IA influenceuse créée en 2016, qui n'a pourtant révélé sa véritable nature qu'en 2018.

provenant d'un humain ou d'IA<sup>238</sup>, et que, dans le cas d'un système artificiel, les responsables puissent être identifiés et tracés<sup>239</sup>. Car en vertu de l'adage *error communis facit jus*, et comme l'a proposé la logique de Turing, si une IA ressemble tant à un humain, fait des mouvements, des erreurs comme un humain, et dispose de toutes les petites subtilités qui constituent un être humain, l'on ne peut alors dire si l'on interagit avec une IA ou un humain. Ainsi, afin de lutter contre l'identité trompeuse, l'application de la théorie de l'apparence en matière d'IA pourrait protéger les tiers de bonne foi ainsi qu'assurer la sécurité juridique.

Pourtant, l'on ne sait pas se contenter de l'identité IA révélée. Pour certains, il faut encore que le processus décisionnel de l'IA — communément connu sous le nom boîte noire étant donné son inintelligibilité aux yeux de l'homme — soit dévoilée.

#### **IV. Les boîtes noires communes chez l'homme et la machine**

**67. L'opacité partagée entre l'homme et l'IA.** — Dans le contexte de l'IA, le terme « boîte noire (black box) » désigne le phénomène où les décisions sont prises par les systèmes d'IA mais les résultats manquent d'explicabilité. Autrement dit, à l'heure actuelle, même si l'IA peut nous fournir des réponses, l'on ne sait pas comment la décision a été prise. Comme les résultats générés par l'IA peuvent constituer des mesures discriminatoires (le sexisme ou le racisme par exemple), lutter contre l'opacité algorithmique constitue un défi pour le droit.

Mais à vrai dire, des algorithmes qui peuvent interagir avec une intelligence humaine ressemblent étroitement aux jugements humains<sup>240</sup>. En effet, même aujourd'hui, les chercheurs ont toujours des difficultés à affirmer l'idée selon laquelle les décisions algorithmiques sont opaques alors que celles des humains sont transparentes. Car de nos jours, les décisions de l'homme ne sont pas tout à fait compréhensibles non plus<sup>241</sup>; en outre, dans toutes les circonstances, une décision, qu'elle émane d'un être humain ou d'un système

---

<sup>238</sup> V. p. ex., art. 52 de la proposition de règlement établissant des règles harmonisées sur l'intelligence artificielle, préc.

<sup>239</sup> The Barcelona declaration for the proper development and usage of AI in Europe, 8 mars 2017.

<sup>240</sup> « An algorithm that mimics our tribal instincts could help AI learn to socialize », MIT Tech. Rev., 22 janv. 2019.

<sup>241</sup> Jean-François Bonnefon: « Speaking as a psychologist, I'm flabbergasted by claims that the decisions of algorithms are opaque while the decisions of people are transparent. I've spent half my life at it and I still have limited success understanding human decisions. », tweet du 24 mai 2019.

automatisé, devrait pouvoir se justifier de son adéquation indépendamment de la manière dont la conclusion a été tirée, « au sens d'un ensemble de normes extérieures prédéfinies »<sup>242</sup>. Ainsi, même si certains auteurs considèrent que l'arbitrage ne devrait pas être remplacé par l'IA car « sa fonction juridictionnelle reste essentiellement sociale et politique »<sup>243</sup>, il faut toutefois prendre en compte qu'éventuellement, il ne s'agit pas d'une question de « sollen (devoir) » mais d'un état de « sein (être) » à venir. Étant donné que dans la forme actuelle d'arbitrage international, aucune des parties ne se trouve avoir le droit de savoir comment le tribunal prend une décision sur leur cas<sup>244</sup>.

Cette similitude entre l'arbitre humain et un système d'IA quant à la décision prise sous forme dite « black box » a donc été explicitement manifestée. Dès lors, même dans la non-transparence du processus décisionnel, cet aspect opaque, voire biaisé et discriminatoire, demeure commun entre ces deux différentes catégories juridiques<sup>245</sup>. La question d'opacité constitue par conséquent une autre incohérence que l'on pourrait noter pour comparer le statut de l'homme et de l'IA.

**68. En résumé.** — Dès que l'on aperçoit que l'imputabilité se justifie essentiellement en fonction du critère de discernement, l'imputabilité des humains privés de la faculté de discerner montre l'incohérence de l'irresponsabilité des IA qui apprennent à discerner. Ensuite, dans le cas de l'interchangeabilité des agents, où les êtres humains peuvent en effet être considérés comme simple outil sur le plan juridique, l'on a démontré que ces facultés indifférenciées peuvent aller jusqu'aux situations fonctionnellement équivalentes de ces agents. Dès lors, cette équivalence sur le plan juridique entraîne nécessairement l'application de la théorie de l'apparence pour protéger la confiance légitime. Enfin, l'opacité partagée entre l'homme et l'IA nous montre également l'incohérence juridique exigeant la transparence sur le processus décisionnel de l'IA mais pas celui de l'homme. Bref, tirant

---

<sup>242</sup> Formule de l'exposé de Clément Henin, « Expliquer et justifier des systèmes de décisions algorithmiques : théorie et pratique », Inria, Paris, 16 oct. 2023.

<sup>243</sup> SIM C., « Will Artificial Intelligence Take Over Arbitration? », *Asian International Arbitration Journal*, vol. 14 Issue 1, 2018, p. 13.

<sup>244</sup> *Ibidem.*, p. 11.

<sup>245</sup> Le débiaisage est difficile mais probablement pas aussi difficile que l'alternative : débiaisage de la langue et de la société ; car les machines capturent simplement le sexisme et d'autres préjugés dans la langue humaine, qui reflète les préjugés de la société humaine. MITCHELL M., *Artificial Intelligence : A Guide for Thinking Humans*, Farrar, Straus and Giroux, 2019.

toutes les conséquences de ces illustrations, les incohérences juridiques du droit constant vis-à-vis des IA de plus en plus avancées nous semblent ouvertement déroutantes.

## Conclusion de la section 2

**69. Les incohérences juridiques exceptionnelles.** — Pour illustrer la nature exceptionnelle de l'IAG, l'on a démontré la notion de l'exception robotique. Sans surprise, des arguments ont été invoqués pour prétendre le contraire, tels que l'argument de Law of the Horse ainsi que l'argument de la neutralité technologique. Cependant, l'adoption de l'argument de Law of the Horse a été bel et bien écarté, à titre d'exemple, par la réglementation européenne. Ensuite, l'on a également dégagé la contradiction de l'argument de la neutralité technologique. Reste à affirmer ce caractère exceptionnel de l'IAG en constatant plusieurs incohérences juridiques existant entre l'humain et l'IA.

En effet, de l'incohérence de l'imputabilité l'on a pu démontrer l'engagement de la responsabilité humaine même en l'absence de discernement mais pas l'IA qui est censée, avant tout, apprendre à discerner. L'on a aussi évoqué l'incohérence d'agent : humain en tant que simple outil sur le plan juridique. De là, l'on a remarqué que l'interchangeabilité des agents constitue une illustration encore plus explicite pour confirmer la fonction équivalente des agents humain et IA sur le plan juridique. À notre avis, un tel résultat qui ne cesse d'évoluer va par suite aboutir à une nécessité de vérification de l'authenticité des agents, ce qui constitue ainsi une autre incohérence dans l'absence de l'application de la théorie de l'apparence pour l'IAG. Enfin, l'on a pu constater que l'opacité du processus décisionnel dite boîte noire est essentiellement partagée entre l'homme et l'IA.

Au regard de ces résultats à notre sens véritablement captivants, l'on pourrait conclure qu'une reconfiguration juridique est désormais indispensable afin d'y remédier.

## Conclusion du chapitre 2

**70. La reconfiguration de la hiérarchie de l'intelligence.** — Dans la mesure où l'IAG pourrait rendre obsolète la capacité d'un individu humain et marginalisent l'espèce humaine, ce caractère exceptionnel de l'IAG s'est éclairci pour dégager l'instauration d'une nouvelle hiérarchie de l'intelligence. C'est ainsi que l'on a pu examiner, de la primauté de l'espèce humaine aux incohérences juridiques en présence d'une nouvelle réalité comprenant désormais l'existence de l'IA avancée.

Dans une telle circonstance, l'incohérence juridique entre l'humain et l'IA se manifeste à travers les exemples tels que l'imputabilité et l'agent humain étant simple outil. Cette interchangeabilité des agents se manifeste également par les « boîtes noires » communes chez l'homme et la machine. Dès lors, le caractère exceptionnel de l'IA est expressément éprouvé. Partant, malgré les arguments tels que Law of the Horse ou la neutralité technologie, l'on a démontré que si l'affirmation du caractère exceptionnel de l'IA faible est déjà pleinement reconnue, la nature existentielle de l'IA forte ne peut plus être niée.

## Conclusion du titre 1

**71. Vers une subjectivité surhumaine.** — Comme l'écrit M. Stuart Russell, « les machines intelligentes auront la capacité de regarder plus loin dans l'avenir que l'homme ne peut le faire. Elles seraient également en mesure de prendre en compte beaucoup plus d'informations. Ces deux capacités combinées mènent inévitablement à de meilleures décisions dans le monde réel. Dans n'importe quelle situation de conflit entre les humains et les machines, l'on trouverait sous peu, comme Garry Kasparov et Lee Sedol, que tous nos mouvements ont été anticipés et bloqués. L'on perdrait le match avant même qu'il ne commence »<sup>246</sup>. À ce propos, il est intéressant de noter qu'aujourd'hui, à part le cerveau<sup>247</sup>, vision des yeux<sup>248</sup> et les oreilles<sup>249</sup> de l'IA sont déjà très bons, et même certaines mains<sup>250</sup>, olfaction de nez<sup>251</sup>, goût de langue<sup>252</sup> et jambes<sup>253</sup> voire sens de l'humour<sup>254</sup> de l'IA sont très performants. Maintenant, mettant tout cela ensemble<sup>255</sup>, la possibilité d'une IA imminente, capable de faire face à de nouvelles situations pour travailler et vivre avec les humains, mérite intimement d'une appréhension juridique particulière.

Certes, il est possible de prétendre que les avions ont une capacité surhumaine de voler et les voitures une capacité surhumaine de courir, afin d'insister que les machines restent toujours un outil, quelle que soit leur capacité surpassant les attributs de l'homme. Pour autant, comme l'on a déjà souligné, l'IAG ne cherche pas seulement à égaler ou à surpasser ces divers domaines étroits ; en revanche, elle vise à simuler, voire remplacer, les facultés que disposent de l'organe traditionnellement chargé de toutes les fonctions

---

<sup>246</sup> RUSSELL S., *Human Compatible*, *op. cit.*, pp. 90, 131.

<sup>247</sup> LI G., « Facebook's AI mathematician can solve university calculus problems », *NewScientist*, 20 déc. 2019.

<sup>248</sup> LU D., « AI camera can tell what surfaces feel like with just a glance », *NewScientist*, 22 sept 2020.

<sup>249</sup> « Textless NLP: Generating expressive speech from raw audio », *Meta AI*, 9 sept. 2021.

<sup>250</sup> Les robots peuvent désormais apprendre à manipuler l'objet par lui-même par essais et erreurs virtuels. OpenAI, « Learning Dexterous In-Hand Manipulation », *arXiv.org*, 1 août 2018.

<sup>251</sup> CEPELEWICZ J., « New AI Strategy Mimics How Brains Learn to Smell », *Quantamagazine*, 18 sept. 2018; PITTALWALA I., « Using artificial intelligence to smell the roses », *Phys.org.*, 29 juill. 2020.

<sup>252</sup> « Artificial tongue could taste whisky to make sure it isn't counterfeit », *NewScientist*, 6 août 2019.

<sup>253</sup> V. p. ex. Mini Cheetah de MIT, Dactyl d'OpenAI, Atlas de Boston Dynamics.

<sup>254</sup> Google dispose désormais d'un modèle de langage capable d'expliquer les blagues. « Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance », *Google AI Blog*, 4 avr. 2022.

<sup>255</sup> P. ex., AlphaFold + AlphaMissense + AlphaStar + AlphaZero + ..... = AlphaInfinity. V. l'exemple de Gato de DeepMind. BORDBAR M. et *alii*, « A Generalist Agent », *DeepMind*, 12 mai 2022.

anthropomorphiques : le cerveau humain. Sur ce fondement, l'on rappelle que le déficit ou la dégradation de tout organe ne suffit pas à justifier toute forme de dérogation à un statut intact et complet, sauf une seule exception du cerveau humain. Ainsi, même si à ce stade, les systèmes d'IA consistent dans les modèles alimentés par des données, il ne faut pas oublier que quand ce sont les jeux d'échecs et Go, et d'autres domaines dits intellectuels qui étaient autrefois uniquement désignés et réservés aux « cerveaux », cette assimilation directe de la « cognition » réaffirme le caractère exceptionnel d'IAG.

Comme ce chapitre vient de le dégager, le caractère exceptionnel de l'IAG s'explique par la nouvelle hiérarchie d'intelligence ainsi que les incohérences juridiques entre l'humain et l'IA. Désormais, l'autonomie croissante de l'IA au sein de la civilisation humaine nous conduit inévitablement à étudier son effet subjectif.

## TITRE 2 LA CONSEQUENCE SUBJECTIVE DE L'INTELLIGENCE

« *I have suffered with those who I saw suffer.* » — Shakespeare

**72. Annonce du plan.** — Jusqu'à présent, les machines sont perçues uniquement comme outil, donc objet, à nos yeux. Cette perspective dite objective à l'égard des ordinateurs, grâce à l'IA, qui au fond fait partie de l'évolution, est désormais remplacée, de manière « naturelle », par une appréhension dite subjective (chapitre 1) de la civilisation humaine. En reconnaissant un tel résultat dans la présente civilisation, cette étude propose que les statuts des êtres méritent d'être réétudiés (chapitre 2).

## Chapitre 1 L'effet de l'anthropomorphisme et de l'empathie

*« Le traitement violent et en même temps cruel des animaux est opposé au devoir de l'homme envers lui-même, parce que la compassion en l'homme se trouve émoussée et qu'une disposition très bénéfique à la moralité s'affaiblit et s'efface peu à peu. » —*

Emmanuel Kant

**73. Annonce du plan.** — L'anthropomorphisme consiste dans l'utilisation d'attributs humains pour ce qui est autre que l'espèce humaine. Cette projection de comportement humain sur les non-humains est susceptible d'être observée dans les relations humain-animal, voire dans les interactions humain-robot. Mais même si l'on n'anthropomorphise pas, l'empathie, la capacité de s'identifier à autrui dans ce qu'il ressent, constitue une autre manière d'« anthropomorphiser les humains » ainsi que les autres entités non-humains. Ainsi, cette étude propose que quand la machine n'est plus qu'une « machine » (section 1), elle peut en effet créer d'ores et déjà un *pretium doloris robotique* (section 2).

### **Section 1 La machine n'est plus qu'une « machine »**

*« Il semble, en effet, que si je suis obligé de ne faire aucun mal à mon semblable, c'est moins parce qu'il est un être raisonnable que parce qu'il est un être sensible; qualité qui, étant commune à la bête et à l'homme, doit au moins donner à l'une le droit de n'être point maltraitée inutilement par l'autre. »*

— Jean-Jacques Rousseau

**74. Annonce du plan.** — La capacité surhumaine de l'IA s'intègre de plus en plus dans divers domaines (§1). Cette évolution ne se limite pas à l'aspect dit objectif de l'IA ; en revanche, elle comprend également le lien affectif — l'IA en tant que sujet à nos yeux (§2).

### **§1 L'intégration de l'IA dans la civilisation : de l'objet surhumain**

**75. Du remplacement physique à la substitution mentale dite professionnelle.** — Rappelons que la capacité juridique prescrit essentiellement le statut d'une entité au sein de

notre système juridique. Quid de l'intégration de l'IA dotée de la capacité surhumaine ? En effet, si l'IA a supplanté des médecins humains pour diagnostiquer tumeurs cérébrales ou cancer du sein<sup>256</sup>, dorénavant en Chine, un robot a réussi l'examen de licence médicale<sup>257</sup>. Cette tendance du recours à l'IA dans le domaine de la santé pourrait faire penser au Japon, où l'article 17 de la loi des médecins dispose littéralement que « si vous n'êtes pas médecin, vous ne devez pas effectuer de travail médical »<sup>258</sup>. Effectivement, le ministère de la Santé, du Travail et du Bien-être japonais a indiqué que dans le processus médical, la décision médicale est centrée sur le médecin, et que dans cette sous-étape, l'IA agit simplement comme un outil de support qui augmente l'efficacité et présente les informations. Donc jusqu'à présent, c'est le médecin qui fait l'objet de la responsabilité du diagnostic et du traitement utilisant l'IA<sup>259</sup>. Mais il nous semble que cette dernière version peut bientôt être révisée car désormais le ministère japonais voit une nécessité à préciser la relation entre l'utilisation d'IA au diagnostic et l'article 17 de la loi des médecins<sup>260</sup>. C'est ainsi qu'aux États-Unis, la Food and Drug Administration (FDA) vient d'autoriser pour la première fois une IA à formuler un diagnostic, sans être supervisée par un médecin<sup>261</sup>.

Rappelons l'incohérence de la responsabilité d'autrui en vertu du principe d'équivalence fonctionnelle évoqué plus haut appliqué entre un médecin humain et une IA, partant, cette étude adhère à la proposition selon laquelle les risques posés par certaines applications médicales d'IA devraient être gérés en comparant leurs performances aux coûts et aux erreurs des médecins et par conséquent réglementées de la même manière que le sont les médecins humains<sup>262</sup>.

---

<sup>256</sup> « Cancer : l'intelligence artificielle ferait mieux que les médecins pour le dépister », Santé Magazine, 6 juil. 2018.

<sup>257</sup> LOUKIL A., « Un robot obtient haut la main le diplôme de médecine en Chine », Capital, 23 nov. 2017.

<sup>258</sup> 医師法（昭和23年法律第201号）第17条 医師でなければ、医業をなしてはならない。

<sup>259</sup> 平成29年度厚生労働行政推進調査事業費補助金「AI等のICTを用いた診療支援に関する研究」。

<sup>260</sup> 人工知能（AI）を用いた診断、治療等の支援を行うプログラムの利用と 医師法第 17 条の規定との関係について，厚生労働省医政局，第 4 回保健医療分野AI開発加速コンソーシアム 資料，平成31年1月16日。

<sup>261</sup> JALINIÈRE H., « Une intelligence artificielle autorisée à poser des diagnostics », Sciences et Avenir, 16 avr. 2018.

<sup>262</sup> BAMBAUER J. R., « Dr. Robot », 51 UC Davis Law Review 383, 2017, p. 393.

Mais le recours à l'IA ne s'arrête pas dans le domaine médical, cette tendance a aussi franchi une nouvelle étape dans le domaine de la justice, à savoir de l'analyse juridique<sup>263</sup> et de la médiation<sup>264</sup>. L'on pourrait citer iCan Systems, réputé être devenu le premier médiateur robot à résoudre un litige devant un tribunal public en Angleterre et au Pays de Galles<sup>265</sup>.

Ces adoptions de la justice rendue par l'application de l'IA se base en effet sur la notion de la justice prédictive, qui consiste dans l'ensemble d'instruments développés grâce à l'analyse de données de justice, permettant ainsi d'établir des statistiques et des probabilités afin de prévoir un résultat contentieux. Mais quid d'une prédiction surhumaine, plus ponctuelle, précise et efficace qu'un jugement sorti du calcul humain ? Cela nous amène à l'idée de la « *singularité juridique (legal singularity)* », définie comme une hypothèse où la capacité de calcul, de renseignement et de prise de décision d'IA dépasse celle des avocats et des juges : la singularité juridique arrivera quand l'accumulation de données massives et l'amélioration des méthodes d'inférence rendra l'incertitude juridique obsolète<sup>266</sup>.

Tirant toutes ces applications, le progrès de l'IA en justice signifie que le système juridique émergera comme un ordre juridique dit automatisé<sup>267</sup>, où des « juges numériques » veillent à l'application des lois ainsi qu'à l'exécution des contrats.

**76. En résumé.** — Dans un tel système juridique, où l'IA ne cesse de s'impliquer dans les attributions dites intellectuelles, l'on considère que les conventions passées entre l'homme et l'IA devraient être envisagées, c'est notamment le cas lorsque l'IA n'est plus considérée en tant qu'objet mais sujet.

## **§2 Le remplacement incontournable des liens affectifs envers le sujet IA**

---

<sup>263</sup> TASHEA J., « AI software is more accurate, faster than attorneys when assessing NDAs », ABA Journal, 26 févr. 2018.

<sup>264</sup> MARMOZ F., « Justice prédictive : de la fiction à la réalité, quelles conséquences pour les professions du droit? », à l'occasion du colloque international « l'intelligence artificielle : un nouveau défi pour le droit? », organisé par l'Université d'Alger 1 Benyoucef Benkhedda, Alger, 27-28 novembre 2018.

<sup>265</sup> BEIOLEY K., « Robots and AI threaten to mediate disputes better than lawyers », Financial Times, 14 août 2019. V. aussi PIALAT L., ROUSSEAU Y., « Intelligence artificielle : une startup américaine va faire plaider le premier 'robot avocat' », franceinfo, 23 janv. 2023.

<sup>266</sup> ALARIE B., « The Path of the Law: Toward Legal Singularity », University of Toronto Law Journal, vol. 66 Issue 4, fall 2016, p. 446.

<sup>267</sup> Cf. infra la substitution de l'organe de contrainte en amont et en aval.

**77. Les robots sociaux en tant que compagnons.** — Aujourd'hui, l'on a des systèmes qui nous parlent, qui nous analysent. Ces systèmes pourraient très bien être intégrés dans des robots avec lesquels l'on pourrait interagir comme avec un être humain, et tous les aspects de nos existences seront alors modifiés d'une façon ou d'une autre<sup>268</sup>. Dans la mesure où les capacités dites sociales des robots humanoïdes continueront à s'améliorer, les compagnons robotiques deviendront plus courant. En outre, l'on passe déjà de plus en plus de temps devant des écrans, créant une génération de citoyens du monde qui a grandi avec interaction automatisée<sup>269</sup>. Dans cette tendance, les robots peuvent tout à fait être conçus pour créer un type de relation différente avec les humains.

**78. Les métaphores informatiques<sup>270</sup> et l'anthropomorphisme involontaire.** — Mais l'on demeure enclin à l'anthropomorphisme, cela signifie que l'on projette nos propres qualités inhérentes sur d'autres entités pour les rendre plus humaines à nos yeux. En effet, si la physique est utilisée pour comprendre les choses, et la psychologie pour comprendre les hommes et les animaux, l'ordinateur est un nouveau type d'objet psychologique, mais encore une chose<sup>271</sup>. Déjà, la façon dont l'on parle des capacités des machines influence nos conceptions de ces capacités. Comme Turing a déjà utilisé le mot « penser » pour penser à la machine, aujourd'hui, l'on continue d'employer les expressions comme « apprendre », « mémoire », « gagner » etc. pour décrire les « actes » réalisés par les machines.

En tout état de cause, il est clair que l'on crée des machines pour qu'elles s'adaptent à l'humain mais l'humain va lui aussi s'adapter à la machine. Plus l'on pourra modifier le comportement des machines en interagissant avec elles, plus l'on va nous y attacher. L'on va projeter de plus en plus sur les robots des intentions humaines, dès lors, l'on parle d'humanisation des robots<sup>272</sup>. Et au fil du temps, les robots sociaux apprennent les

---

<sup>268</sup> DEPARDIEU G., MARTIN TH., « L'intelligence artificielle va-t-elle nous dépasser? », ARTE, 2018.

<sup>269</sup> BRAZEAL C., « Social robots will become family members in the homes of the future », Quartz, 17 oct. 2018.

<sup>270</sup> Cf. l'Hypothèse de Sapir-Whorf, ELIZA effect, et Wittgenstein: « *Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt.* », Satz 5.6., *Tractatus Logico-Philosophicus*, 1921.

<sup>271</sup> TURKLE S., *The Second Self: Computers and the Human Spirit*, Simon & Schuster, 1984, First MIT Press edition, 2005, p. 34.

<sup>272</sup> Laurence Devillers, chercheuse au laboratoire du CNRS, avoue que le robot la touche quand il lui dit « je t'aime ». Elle sait bien que le robot n'est pas capable de ressentir une émotion mais pourtant, cela peut être touchant pour elle. Elle va lui prêter des intentions, être compatissant et pouvoir s'y attacher émotionnellement. DEVILLERS L., « L'empathie des robots », Défis n°8 : L'entreprise à l'heure de l'intelligence artificielle -

préférences des humains avec lesquelles ils vivent et adaptent leur comportement pour répondre aux besoins uniques et individuels de chacun. Cela constitue à l'évidence le premier pas vers des relations plus intimes à venir, par exemple, le mariage.

**79. Vers un mariage robotique.** — Non seulement les entreprises fabriquent des robots sexuels aux visages interchangeable permettant à la poupée de cligner de l'œil, de bavarder et de murmurer, dorénavant à Moscou, l'on peut acheter 30 minutes avec un sexbot pour un prix d'environ 90 dollars<sup>273</sup>. Quant au régime matrimonial, un ingénieur chinois a en effet épousé en 2017 son robot, sa propre création, après avoir échoué à trouver une femme humaine<sup>274</sup>. De même un Japonais a épousé une petite poupée de la chanteuse humanoïde Hatsune Miku en 2018<sup>275</sup>. Tout cela conduit les humains à percevoir les robots comme des partenaires appropriés<sup>276</sup>. Sous cet aspect de « l'attachement robotique formalisé voire institutionnalisé », l'on souligne encore une fois la potentialité distincte que pourrait amener entre l'IA étroite et l'IA générale. Car même si les humains peuvent très bien se trouver émotionnellement attachés aux objets tels que les animaux ou les outils, les mariages avec un chat ou un grille-pain demeurent relativement moins concevables pour le grand public. En revanche, les sexbots et les compagnons androïdes qui existent déjà de nos jours aboutissent à une relation particulière homme-robot, qui répond en effet au besoin d'une partie de la population humaine. De ce fait, cette dimension psychologique des hommes pour les robots forme une nouvelle sexualité.

**80. Digisexuel et Robosexuel.** — La relation homme-androïde s'appelle désormais « digisexuels », ce qui, selon certains, constitue une identité sexuelle émergente. Dans cette démarche, en 2016, une femme française s'est décrite comme une « robosexuelle », disant que les « vrais hommes » la dégoûtaient et qu'elle est vraiment et seulement attirée par les

---

entre promesses et nouveaux défis, Revue du département Intelligence et sécurité économiques, INGESJ, n°8, nov. 2017, p. 75.

<sup>273</sup> WILLIAMS A., « Do You Take This Robot ... », NYT, 19 janv. 2019. V. aussi : « Belgique : ouverture de la première maison close à poupées sexuelles », France 3 Nord, 17 avr. 2018; LEFEBVRE A., « Le premier bordel de poupées sexuelles ouvre ses portes en Allemagne », Express Business, 19 oct. 2017.

<sup>274</sup> HUANG K., « Chinese engineer 'marries' robot after failing to find a human wife », South China Morning Post, 3 avr. 2017.

<sup>275</sup> MILEY J., « Japanese Man Marries Doll of Virtual Singer », Interesting Engineering, 13 nov. 2018.

<sup>276</sup> David Levy prédit que les humains tomberont systématiquement amoureux de robots humanoïdes. V. LEVY D., *Love and Sex With Robots: The Evolution of Human-Robot Relationships*, Harper Collins, 2007.

robots<sup>277</sup>. En effet, la digisexualité est comprise en deux vagues : « première vague (First wave digisexuality) » comme la pornographie en ligne, sexting et jouets sexuels électroniques, où la technologie est simplement un système de livraison pour la réalisation sexuelle, et une « deuxième vague (Second wave digisexuality) », où les praticiens nouent des relations plus profondes grâce à des technologies immersives telles que la réalité virtuelle, la réalité augmentée et les robots sexuels équipés d'IA, évitant parfois la nécessité d'un partenaire humain<sup>278</sup>, voire « se retirant heureusement des relations humaines »<sup>279</sup>.

Ces technologies sexuelles personnalisables, enrichies de données de plus en plus améliorées sont désormais disponibles. En tenant compte de ces applications, la loi américaine sur les robots de pédophilie électroniques (Curbing Realistic Exploitative Electronic Pedophilic Robots, dite CREEPER) a été votée dans le contexte où le législateur constate que « les robots peuvent avoir des paramètres qui simulent le viol », et que « non seulement les poupées et les robots mènent au viol, mais qu'ils facilitent le viol en enseignant au violeur comment vaincre la résistance et soumettre la victime »<sup>280</sup>.

Vu que les enjeux physiologiques ou comportementaux des interactions humain-robot ne peuvent plus être négligés, l'intérêt universel de la protection juridique en la matière se manifeste. Au sens de cette étude, la mise en place encore plus générale des dispositions comme CREEPER devrait être encouragée. Car ces normes juridiques pourraient servir à nous éclairer sur les relations entre les humains et les robots, notamment lorsque l'on ne différencie plus l'état affectif l'un envers l'autre.

**81. En résumé.** — Vu que les mammifères ne peuvent pas vivre uniquement de nourriture, et ont aussi besoin de liens affectifs, nos liens émotionnels sont plus susceptibles d'être formés avec des objets à fourrure douce que des objets durs et métalliques<sup>281</sup>. Même les

---

<sup>277</sup> CRAW V., « French woman wants to marry a robot as expert predicts sex robots to become preferable to humans », News.com.au., 23 déc. 2016.

<sup>278</sup> MCARTHUR N., TWIST M. L. C., « The rise of digisexuality: therapeutic challenges and possibilities », *Journal Sexual and Relationship Therapy*, vol. 32, 2017.

<sup>279</sup> SINGH-KURTZ S., « The Man of Your Dreams For \$300, Replika sells an AI companion who will never die, argue, or cheat — until his algorithm is updated », *The CUT*, 10 mars 2023.

<sup>280</sup> H.R.73 — 117th Congress (2021-2022) - Curbing Realistic Exploitative Electronic Pedophilic Robots Act 2.0 (CREEPER).

<sup>281</sup> Sur ce point, v. p. ex. HARLOW H. F., « The nature of love », *American Psychologist*, 13(12), 1958, pp. 673-685.

rongeurs peuvent former des liens sociaux avec des robots et vont aider à sauver un rat robot qui est pris au piège dans une cage<sup>282</sup>. Et comme l'on évolue dans le spectre entre traiter les robots sociaux comme des grille-pains et les traiter davantage comme des animaux domestiques<sup>283</sup>, lorsque l'on attribue des états affectifs à des robots, l'on est moins susceptible de les sacrifier pour sauver des humains<sup>284</sup>.

Justement, une caractéristique clé des robots sociaux est qu'ils sont spécifiquement conçus pour susciter ces projections. Comme l'on interagit déjà différemment avec les robots sociaux qu'avec d'autres objets, il serait nécessaire que le déploiement juridique tienne compte de la manière dont l'on réagit envers les machines.

---

<sup>282</sup> KLEIN A., « Rats can make friends with robot rats and will rescue them when stuck », *NewScientist*, 27 nov. 2018. Aussi, les singes pleurent la mort d'un singe robot. SHAW A., « A funeral for a spy monkey », *BBC Earth*, 18 sept. 2019.

<sup>283</sup> DARLING K., « Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects, *Robot Law* », *We Robot Conference 2012*, University of Miami, p. 9.

<sup>284</sup> PAULUS M. et *alii*, « Saving the Robot or the Human? Robots Who Feel Deserve Moral Care », *Social Cognition*, 2019, vol. 37, n°1, p. 41 s.

### **Conclusion de la section 1**

**82. Le lien affectif robotique.** — Dans cette section, l'on a démontré que lorsque la machine n'est plus qu'une « machine », l'intégration de l'IA dans la civilisation humaine commence par son remplacement physique, jusqu'à la substitution professionnelle voire mentale. Cette substitution incontournable des liens affectifs envers le sujet IA fait naître les robots sociaux en tant que compagnons. Suivant un tel constat, l'on a constaté les propos de mariage robotique, ainsi que les sexualités identifiées telles que digisexualité et robosexualité. Enfin, tirant toutes ces démonstrations, l'on pourrait conclure que, éventuellement, il est tout à fait possible que l'on sauve le robot au lieu de l'humain.

Dès lors, bien que cette étude soutienne les relations entre humains et IA, l'on considère que la protection juridique de ces interactions devrait avant tout être bien configurée. C'est ainsi qu'à notre avis, la mise en place d'un « pretium doloris IA » pourrait être le premier pas dans la protection des relations humains-robots.

## **Section 2 Vers un pretium doloris de l'IA**

« *His words had a strange effect upon me. I compassionated him and sometimes felt a wish to console him...* » — Mary Shelley

« *Suffering exist in the eye of the observer. The first time a programmer delivers a product that says it is suffering is the first time it becomes unethical to torture AI systems.* »

— Lex Fridman

**83. Annonce du plan.** — Dès que la substitution robotique du lien affectif est inévitable, cette étude propose d'installer un pretium doloris robotique (§1), et éventuellement, d'envisager la protection des IA en tant que sujet (§2).

### **§1 Le pretium doloris de l'objet IA**

**84. L'irremplaçabilité de l'IA.** — Selon le plus grand cabinet d'études de marché allemand, plus de la moitié des habitants du monde ont hébergé au moins un type d'animal de compagnie, et la pratique de la chimiothérapie, des antidépresseurs, des chirurgies de remplacement d'articulation, voire du « congé de maternité »<sup>285</sup> pour ces animaux de compagnie devient de plus en plus répandue. Ces positions « de faveur » n'en sont pas moins partagées par la Cour de cassation française, notamment dans un litige concernant le défaut du bien en droit de la consommation. Sa jurisprudence retient ainsi que « le chien en cause était un être vivant, unique et irremplaçable, et un animal de compagnie destiné à recevoir l'affection de son maître, sans aucune vocation économique, le tribunal, qui a ainsi fait ressortir l'attachement de Mme Y... pour son chien, en a exactement déduit que son remplacement était impossible, au sens de l'article L. 211-9 du code de la consommation »<sup>286</sup>. L'on sait que le préjudice affectif du fait de la perte d'un cheval a déjà été reconnu dans l'arrêt Lunus rendu en 1962. Dans cet arrêt, la Cour de cassation française a jugé qu'« indépendamment du préjudice matériel qu'elle entraîne, la mort d'un animal peut

---

<sup>285</sup> LUFKIN B., « When a pet dies, should you be allowed time off work? », BBC, 13 sept. 2018.

<sup>286</sup> Cass. civ. 1<sup>e</sup>, 9 déc. 2015, n° 14-25.910.

être pour son propriétaire la cause d'un préjudice d'ordre subjectif et affectif susceptible de donner lieu à réparation »<sup>287</sup>.

L'affirmation jurisprudentielle nous montre qu'il est tout à fait naturel même inévitable, de se trouver en butte à ce phénomène d'« irremplaçabilité de l'IA » qui réclame un « *pretium doloris de l'IA* » sur le fondement du préjudice d'ordre subjectif et affectif. En effet, le Code de la consommation prévoit qu'en cas de défaut de conformité, l'acheteur choisit entre la réparation et le remplacement du bien. Mais ici le juge nous dit très clairement que ce « bien » est « un être vivant, unique et irremplaçable », qui est « destiné à recevoir l'affection ». Partant, l'objet au sens juridique traditionnel est désormais doté d'une estimation subjective, qui aboutit à un traitement analogique auparavant réservé au sujet juridique des personnes physiques.

Cette évolution juridique de la relation entre sujet du droit et objet du droit sur le fondement du lien affectif pourrait très bien se voir généralisé dans les relations entre les humains et l'IA. En réalité, cette histoire incontournable a déjà été bel et bien illustrée à l'occasion des funérailles pour les chiens robots au Japon<sup>288</sup>. Aujourd'hui, les robots hors d'usage font déjà l'objet d'une cérémonie funéraire ; demain, ce lien d'affection devra être protégé pour la relation entre Homo sapiens et nos familles robotiques.

Prima facie, il s'agit de la protection des robots<sup>289</sup>, mais au final, il s'agit plutôt de la protection de notre sujet juridique traditionnel de personne physique. Partant, même s'ils ne sont pas (encore) conscients et ne pourraient pas vraiment souffrir, leur souffrance serait tout de même préjudiciable aux humains. L'on vient de voir que faire du mal à un robot n'est pas

---

<sup>287</sup> Cass. civ. 1<sup>e</sup>, 16 janv. 1962.

<sup>288</sup> « Au Japon, des funérailles pour les chiens robots hors d'usage », *Le Point*, 26 avr. 2018. V. aussi dans la nouvelle d'Isaac Asimov, « Robbie », où il s'agit précisément d'un robot irremplaçable aux yeux d'enfant. ASIMOV I., *I, Robot*, Gnome Press, 1950.

<sup>289</sup> Les entités non conscientes ont ce que l'on appelle des « droits indirects ». Ce sont des droits qui découlent de leur relation avec un être qui a un statut éthique. Vu que nuire à l'entité non conscient nuirait à l'être ayant un statut éthique, il faut donc l'éviter. En effet, il n'y a pas si longtemps, les femmes et les esclaves avaient principalement des droits indirects. La simulation de la sensibilité dont font preuve les entités artificielles peut leur fournir des revendications morales indirectes, découlant de considérations sur l'expérience d'un humain et non sur celle de l'entité non consciente. DONATH J., « Ethical Issues in Our Relationship with Artificial Entities », in DUBBER M. D. et alii (dir.), *The Oxford Handbook of Ethics of AI*, Oxford University Press, 2020, p. 61.

simplement similaire à celui de faire du mal à un humain; c'est identique<sup>290</sup>. Le traitement d'un robot peut ainsi être psychologiquement nuisible pour les êtres humains<sup>291</sup>. Le sujet traditionnel des personnes physiques est, dans ce contexte et à ce stade, le véritable enjeu dans la protection des dispositifs garantissant les systèmes d'IA.

**85. En résumé.** — L'on voit que l'attachement psychologique des humains ne distingue pas en fonction de l'objet de sa projection soit un humain ou une chose<sup>292</sup>. Dès lors, la protection affective et le préjudice affectif ne devraient pas non plus le faire sur le fondement de ce critère inapte et futile. En revanche, la protection devrait intégrer ces considérations indéniables de l'attachement subjectif afin qu'elle soit compatible avec la *ratio legis* du préjudice affectif. Dans cette démarche de l'extension de la protection de l'IA objet, pour se préparer à l'IA de plus en plus avancée, il convient dès maintenant d'envisager une protection de sujet IA.

### **§2 La protection des IA en tant que sujet**

**86. Préjudice d'affection généralisé voire réciproque.** — Dans un premier temps, l'on pourrait s'attendre à ce que ce préjudice moral sur le fondement de l'affection de l'IA se généralise. Dans cette trajectoire, en manifestant une conscience de soi que l'on associe à d'autres créatures vivantes, certains proposent une reconnaissance absolument extensive de protection. Par exemple, dès qu'une IA est capable d'exprimer sa souffrance, elle devrait avoir des droits comme les animaux<sup>293</sup>. Ainsi, si la bestialité et le viol sont condamnés par notre système juridique, cela pourrait conduire la loi à interdire les « abus » des robots sociaux<sup>294</sup>. Dès lors, suivant la même *ratio legis*, il y aurait lieu de reconnaître à ces êtres les protections assimilés aux droits fondamentaux tels que le précisent la référence aux

---

<sup>290</sup> KEYSERS CH., et *alii*, « The anthropomorphic brain: the mirror neuron system responds to human and robotic actions », *Neuroimage*, vol. 4, Issue 4, 1 mai 2007, pp. 1674-1684.

<sup>291</sup> V. p. ex. VAN CAMP J., « My Jibo Is Dying and It's Breaking My Heart », *WIRED*, 8 mars 2019.

<sup>292</sup> Cf. p. ex. Kate Darling.

<sup>293</sup> Lex Fridman, lors de sa discussion avec Peter Singer, *Lex Fridman Podcast*, 8 juillet 2020.

<sup>294</sup> *Ibidem*. En ce sens, v. p. ex. premier alinéa de l'article 521-1 du Code pénal français qui prévoit que « le fait, publiquement ou non, d'exercer des sévices graves, ou de nature sexuelle, ou de commettre un acte de cruauté envers un animal domestique, ou apprivoisé, ou tenu en captivité, est puni de deux ans d'emprisonnement et de 30 000 euros d'amende. ».

traitements inhumains de l'article 3 de la Convention de sauvegarde des droits de l'homme et des libertés fondamentales (CSDHLF)<sup>295</sup>.

Mais si, dans un second temps, la protection se fondant sur ces mêmes principes généraux qui s'appliquent aux êtres humains — sujet juridique par excellence, la question se pose est donc de savoir si cette démarche ne constitue pas un traitement, quoique certainement réticent au début, égal à l'aménagement du droit dit de l'homme.

En outre, il faut aussi s'interroger sur la finalité du préjudice dit direct de l'IA, par exemple, son affection enregistrée envers les humains pourrait s'avérer très élevée au cours des interactions humains-bots<sup>296</sup>. Dans ce cas, il faut se demander comment justifier cette absence du traitement comparablement adéquat dans une telle situation réciproque.

**87. En résumé.** — Au final, vu que nos relations avec les IA pourraient très bien être dotées d'une valeur sentimentale, un préjudice d'affection doit être mis en place. Dans l'hypothèse où un humain partage sa vie avec son compagnon IA et son chien robotique, il devient souhaitable que le droit lui confère deux chefs de préjudice affectif, à savoir l'ordre sujet pour son compagnon, et l'ordre objet pour son chien. Et dans la mesure où ce type de préjudice d'affection se généralise, l'application devient la règle de principe. C'est donc si éventuellement, l'IA se retrouve dans une position de réciprocité, au sens de cette étude, qu'il y aurait lieu de transformer la protection de l'objet IA en sujet IA.

---

<sup>295</sup> LAWRENCE D. R., BRAZIER M., « Legally Human? 'Novel Beings' and English Law », préc., p. 327.

<sup>296</sup> V. p. ex. ISHIGURO K., *Klara and the Sun*, Faber and Faber, 2021.

## Conclusion de la section 2

**88. De l'objet IA au sujet IA.** — Lorsque les décisions jurisprudentielles nous montre qu'il est tout à fait naturel même inévitable, de se trouver en butte à ce phénomène d'« irremplaçabilité » d'un « bien » qui est « destiné à recevoir l'affection », l'on a démontré que, incontestablement, un préjudice affectif peut en effet résulter entièrement de l'état d'un « bien ». Suivant cette évolution juridique, l'on a dégagé l'intérêt d'un « pretium doloris d'IA ». Prima facie, il peut sembler être une protection des robots, mais l'on se rend compte qu'au final, à ce stade, il s'agit plutôt de la protection de la personne physique.

Tant que le préjudice d'affection se généralise dans le contexte de l'IA, la protection de l'IA encadrée dans des dispositifs tels que les abus ou les traitements inhumains pourrait illustrer une appréhension comparable. En outre, sur le fondement d'une éventuelle réciprocité, un aménagement réciproque ne serait pas exclu. Dans ce sens, l'on a soulevé la possibilité que certains futurs systèmes d'IA soient candidats à un statut en tant que sujet juridique. Le fondement de la protection de l'IA passe alors de l'objet au sujet. Une fois que l'on a abouti à cette conclusion, il reste à étudier la coexistence de personnes juridiques intelligentes.

## Conclusion du chapitre 1

**89. Vers la personnalité juridique de l'IA.** — Notre inclination à être liés aux animaux de manière anthropomorphique se traduit chez les IA. Et l'incarnation physique augmente cette tendance d'anthropomorphisme<sup>297</sup>. Dorénavant, l'IA est capable de reconnaître les émotions dans le texte et de produire un discours et une parole imprégnée d'émotion de qualité humaine<sup>298</sup>, tout comme de tenir une conversation naturelle et d'afficher des expressions faciales, conçue pour réagir comme les humains<sup>299</sup>. Le droit doit faire face à la question de la relation entre les bits et les atomes : de la frontière entre le numérique et le physique, de ce qu'il faut faire quand un tel objet numérique constitué de bits devient un objet physique composé d'atomes et vice versa, notamment, de la fusion de l'intelligence dite artificielle et naturelle dans cette évolution où les atomes arrangeant désormais les bits qui arrangeant réciproquement les atomes.

Le mariage robotique<sup>300</sup> pourrait à l'avenir conférer d'autres droits aux robots tels que le droit d'hériter et le droit de reproduction<sup>301</sup>. À l'heure actuelle, régler le traitement des robots n'implique pas de leur octroyer des droits légaux ; l'instauration de protections constituerait plutôt un frein à leur céder des droits<sup>302</sup>. Aussi faut-il réfléchir à l'implication des robots qui sont censés nous traiter moralement mais pas l'inverse. Ce manquement d'un traitement réciproque serait susceptible d'encourager une image d'esclavage<sup>303</sup> ou l'abus des

---

<sup>297</sup> Cf. p. ex. Kate Darling.

<sup>298</sup> DeepZen | IBM : human-like quality of speech imbued with emotion.

<sup>299</sup> « Furhat Robotics lance un robot humanoïde au visage personnalisable », *Le Progres*, 9 nov. 2018. V. aussi les Robots de Boston Dynamics, Hanson Robotics, laboratoire de Hiroshi Ishiguro, et SEER, robot de simulation d'expression émotionnelle de Takayuki Todo, un robot humanoïde développé sur le regard et l'expression faciale humaine.

<sup>300</sup> V. supra § mariage robotique.

<sup>301</sup> KAZI P., « Humans and robots can have babies, claims AI expert », *International Business Times*, 20 déc. 2017. Et l'exemple d'une femme qui a subi une fausse couche traumatisante, ne peut pas avoir d'enfants et a deux enfants IA. SINGH-KURTZ S., « The Man of Your Dreams For \$300, Replika sells an AI companion who will never die, argue, or cheat — until his algorithm is updated », préc.

<sup>302</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, 2009, p. 210.

<sup>303</sup> Cela présente donc un problème au regard de la deuxième Loi de la Robotique d'Asimov qui dispose qu'« un robot doit obéir aux ordres qui lui sont donnés par un être humain, sauf si de tels ordres entrent en conflit avec la première loi ». Par ailleurs, étymologiquement, le terme robot est issu des langues slaves et formé à partir de rabot, rabota (работа en russe) qui signifie travail, corvée que l'on retrouve dans le mot Rab (раб),

robots. Ces humanoïdes violés ou robots mal traités pourraient susciter l'extension de la protection robotique à travers le *pretium doloris* d'IA. Et éventuellement, la protection de l'IA pourrait même passer de l'ordre objet à l'ordre sujet. En constatant cette évolution qui nous semble inévitable, il convient à présent d'étudier la personnalité juridique de l'IA.

---

esclave en russe. Il fut initialement utilisé par l'écrivain tchécoslovaque Karel Čapek dans sa pièce de théâtre R. U. R. (Rossum's Universal Robots) en 1920.

## Chapitre 2 La coexistence de personnes juridiques intelligentes

« Mais, pour qu'une nation se soulève en vue et au nom de tout le monde, il faut qu'elle ait le programme de tout le monde, assez large, assez profond, assez vrai, assez humain en un mot, pour embrasser les intérêts de tout le monde, (...), sans différence de nationalité. »

— Mikhaïl Bakounine

« (...) les principes poussés jusqu'à leurs dernières conséquences renverseraient la législation, dispenseraient les parents de l'éducation de leurs enfants (...) » — Diderot

**90. Annonce du plan.** — L'on vient de voir qu'un niveau d'intelligence comparable à celui des humains (voire surhumain) signifie son intégration inéluctable dans la civilisation humaine. Cette nouvelle réalité va créer sur le plan juridique de nouvelles personnes juridiquement intelligentes (section 1) qui nécessite donc un registre d'IA (section 2).

### Section 1 La personne juridiquement intelligente

**91. Annonce du plan.** — Pour étudier le statut d'IAG (§2), il convient d'abord de revisiter brièvement la *fictio legis* de la personnalité juridique (§1).

#### §1 La notion de personnalité juridique

**92. La personnalité juridique : critère summa divisio pour la cohérence des régimes.** — Comme l'on vient de voir, à présent, bien qu'il y ait toujours quelqu'un qui conçoit, fabrique ou contrôle chaque automate, les machines jouissent déjà d'une nouvelle indépendance et d'une nouvelle individualité au sein de notre société<sup>304</sup>. Un tel développement irrévocable exige une clairvoyance de hauteur de la capacité juridique et de la personnalité juridique, ce qui nous permet d'éviter conflit et divergence quant à l'application de différents régimes. Autrement dit, l'attribution de la personnalité juridique

---

<sup>304</sup> WEIN L. E., « The Responsibility of Intelligent Artifacts: Toward An Automation Jurisprudence », Harv. J.L. & Tech, vol. 6 Fall Issue, 1992, pp. 105-106.

résoudrait simultanément la problématique provenant d'un même agent intervenant aux régimes de la responsabilité contractuelle et de la responsabilité délictuelle<sup>305</sup>.

En effet, la personnalité juridique concerne l'aptitude à être titulaire de droits. C'est ainsi que le droit reconnaît cette aptitude tant aux personnes physiques qu'aux personnes morales. Quant à la capacité juridique, elle consiste dans l'aptitude pour une personne à exercer ses droits et obligations<sup>306</sup>. Dès lors, la personnalité juridique est en effet « la capacité juridique pour les relations juridiques »<sup>307</sup>. Ainsi, aucun être n'est porteur de droits, s'il n'est également porteur d'intérêts qui peuvent être affectés par les actes d'autrui. Et nul être n'est capable d'assumer des devoirs s'il n'est également capable d'actes qui pourraient affecter les intérêts d'autrui<sup>308</sup>. La personne juridique est ainsi définie comme « tout être à qui la loi attribue une capacité d'intérêts et donc de droits, d'actes et donc de fonctions »<sup>309</sup>. Partant, l'on comprend que les personnes juridiques sont des entités construites dans le système juridique en tant qu'« artefact sémantique » auxquelles des communications sont attribuées<sup>310</sup>. En d'autres termes, les entités sont décrites en tant que personnes juridiques lorsque le système juridique leur donne des communications au sens juridique : c'est-à-dire des entités qui produisent des actes juridiques dans le système juridique<sup>311</sup>.

Sur ce fondement, l'objectif général de la personnalité juridique est de faciliter la réglementation, par la société organisée, des comportements et des relations humaines<sup>312</sup>. Partant, chaque fois que la société, dans l'administration de la justice, le juge opportun, pour un but particulier, et à cette fin, une personne juridique est née<sup>313</sup>. Ainsi, toutes les entités exigent la même chose pour en faire des personnes juridiques, à savoir, être partie prenante de relations juridiques ; aucune d'elles n'exige rien de plus<sup>314</sup>.

---

<sup>305</sup> « Automatisiertes Und Vernetztes Fahren », *op. cit.*, p. 13.

<sup>306</sup> DE VAUPLANE H., « La personnalité juridique des robots », *Revue Banque*, 2017.

<sup>307</sup> SALMOND J., *Jurisprudence*, 5<sup>e</sup> éd., 1916, p. 272.

<sup>308</sup> *Ibidem*, p. 283.

<sup>309</sup> *Ibid.*

<sup>310</sup> TEUBNER G., « Enterprise Corporatism: New Industrial Policy and the "Essence" of the Legal Person », *The American Journal of Comparative Law*, 1988, p. 136.

<sup>311</sup> ALLEN T., WIDDISON R., « Can Computers Make Contracts? », *Harv. J.L. & Tech.*, vol. 9 n°1, 1996, pp. 37-38.

<sup>312</sup> SMITH B., « Legal Personality », *Yale Law Journal*, vol. 37, issue 3, janv. 1928, p. 296.

<sup>313</sup> *Ibidem*, p. 289.

<sup>314</sup> *Ibid.*, p. 292.

**93. Le détachement nécessaire du concept de l'homme et celui de la personne.** — Rappelons que le mot « *personne* » est dérivé du mot latin « *persona* », qui faisait référence à un masque porté par un humain qui véhiculait un rôle particulier dans une pièce de théâtre. Cette notion de personne a été introduite pour la première fois en droit romain comme un terme technique. En mettant des masques, les acteurs ont signifié qu'ils jouaient un rôle. Et en vertu de l'adage « *unus homo sustinet plures personas* », un homme, en effet, pourrait être plusieurs personnes<sup>315</sup>. Cela a donné l'impression de décrire une apparence donnée pour exprimer certaines caractéristiques. Ce n'est que plus tard que le terme est devenu coextensif avec l'humain qui assumait la personnalité, et est donc devenu interchangeable avec le terme « *humain* ». Cependant, même lorsque cette transformation de la signification linguistique s'opérait, les concepts de personne et d'humain demeuraient distincts<sup>316</sup>. Il est entendu qu'en droit, la personne renvoie simplement à un masque<sup>317</sup>. Cet usage de « *personne* » est malheureusement lui-même susceptible d'induire en erreur car le mot est souvent utilisé comme s'il signifiait la même chose qu'« *être humain* ». Pour autant, les termes ne sont pas équivalents. Il pourrait exister une personne qui n'est pas membre de notre espèce de même que des membres de notre espèce qui ne sont pas des personnes<sup>318</sup>. Ainsi, « *personne* » désigne un agent<sup>319</sup>. Seuls les agents autonomes, grâce à leur capacité potentielle d'autodétermination, peuvent être considérés comme une personne, quelle que soit leur nature humaine ou non humaine<sup>320</sup>.

Comme la personne est « l'unité du support des droits et devoirs (right-and-duty bearing unit) », la définition d'un sujet de droit est en effet constituée d'une analyse des faits et non de la recherche de l'essence inhérente<sup>321</sup>. Le critère de la personnalité juridique nous

---

<sup>315</sup> BROZÈK B., « The Troublesome 'Person' » in PIETRZYKOWSKI T. et alii (dir.), *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, Springer International Publishing AG, 2017, p. 8.

<sup>316</sup> CALVERLEY D. J., « Imagining a non-biological machine as a legal person », préc., p. 525.

<sup>317</sup> REGAD C., « À quand l'animal reconnu comme une "personne" juridique en France ? », *The Conversation*, 29 août 2018.

<sup>318</sup> V. supra discussion sur la primauté du cerveau parmi les organes.

<sup>319</sup> SINGER P., *Practical Ethics*, op. cit., p. 87.

<sup>320</sup> ENGELHARDT H. T., « Some Persons are Humans, Some Humans are Persons, and The World is What We Persons Make of It », in ENGELHARDT H. T., SPICKER S. F., (eds.), *Philosophical medical ethics: Nature and significance*, Reidel Dordrecht, 1977, cité in PALMERINI et alii, « Robot Companions as Case-Scenario for Assessing the "Subjectivity" of Autonomous Agents. Some Philosophical and Legal Remarks », *Proceedings of the 1<sup>st</sup> Workshop on Rights and Duties of Autonomous Agents*, 2012, p. 31.

<sup>321</sup> DEWEY J., « The Historic Background of Corporate Legal Personality », *Yale Law Journal*, vol. 35, n°6, 1926, p. 661.

demande donc d'identifier les capacités d'une créature et non pas son substrat biologique. Ainsi, étant donné que le lien entre « humain » et « personne » rend obscure la discussion sur ce que signifie réellement être capable de droits et d'obligations, il faut dissocier le concept l'humanité de personnalité au sens juridique<sup>322</sup>. Une rupture conceptuelle entre la notion de personne juridique et celle de la notion d'humanité est donc nécessaire pour évaluer, de manière opportune, si un particulier est une personne juridique<sup>323</sup>.

Partant, l'on comprend que *humain* fait référence à une catégorie biologique alors que *personne* fait référence à une entité dotée d'un ensemble de capacités<sup>324</sup>. Ainsi, dans un premier temps, la présente étude dissocie la considération humaine et la condition pour l'IA d'avoir la personnalité juridique. En effet, l'on pourrait constater que cette appréhension fait partie des aménagements convenables dans l'évolution juridique.

**94. Les aménagements factuels.** — Au cours des années, diverses raisons ont été prises en compte pour étendre la reconnaissance juridique à de nouveaux groupes. Ces décisions « tendent à refléter les valeurs de leur époque et n'apportent pas beaucoup de lumière sur l'essence de la personnalité juridique. Chacune de ces extensions constituait toutefois la reconnaissance du fait que les entités individuelles considérées pouvaient être assimilées davantage aux personnes effectuant la considération qu'à des biens appartenant à ces personnes »<sup>325</sup>. En effet, les histoires juridiques de plusieurs groupes, tels que les esclaves, les femmes, les étrangers (barbares), les animaux, la nature, et les personnes morales privées comme publiques, ont suivi le schéma de leur « reconnaissance » initiale en tant que « personnes » sur le plan juridique, suivies par un lent aboutissement de leurs droits et obligations. Ces résultats montrent que, pour tout autres que les conceptions antérieures, un statut peut être aperçu différemment au fil du temps et qu'un nouveau régime spécifique peut leur être consacré.

---

<sup>322</sup> DYSCHKANT A., « Legal Personhood: How We Are Getting It Wrong », préc., p. 2077, 2091.

<sup>323</sup> *Ibidem*, p. 2108.

<sup>324</sup> SALMOND J., *Jurisprudence*, 10<sup>e</sup> éd., 1947, p. 318.

<sup>325</sup> WILLICK M. S., « Constitutional Law And Artificial Intelligence: The Potential Legal Recognition Of Computers as 'Persons' », Proceedings of the 9th international joint conference on Artificial intelligence (IJCIA), vol. 2, janv. 1985, p. 1271.

**95. En résumé.** — Les systèmes juridiques attribuent des statuts ainsi que des droits et des devoirs aux différentes entités tels que les esclaves, les femmes, les étrangers (barbares), les animaux, la nature, et les personnes morales. Ces aménagements factuels signifient qu'en matière d'IAG, il n'y a aucune raison de s'opposer à l'instauration des entités supposant des rapports juridiques sui generis. Ainsi, dès qu'une IA est comparable à un sujet juridique comme la personne physique ou la personne morale de nos jours, afin de garantir la sécurité juridique, il devient nécessaire de pouvoir identifier ces *personae fictae* — porteurs des statuts. À cette fin, l'immatriculation de ces nouvelles fictivités juridiques devrait se voir justifiée et généralisée.

## §2 Le statut de l'IAG

**96. Annonce du plan.** — Avant d'étudier la personnalité juridique de l'IA (III), il convient d'étudier préalablement les notions évolutives telles que la reconnaissance de la quasi-personnalité (I) ainsi que le revirement de *societas delinquere non potest* (II).

### **I. La quasi-personnalité**

**97. À mi-chemin des catégories traditionnelles.** — L'on sait que l'hétérogénéité de la personnalité morale se manifeste par ses formes diverses et qui ne jouissent pas des mêmes prérogatives<sup>326</sup>. En effet, la quasi-personnalité morale est une quasi-personnalité juridique qui est proche de conférer à son titulaire les prérogatives attachées à la personnalité juridique, définie comme l'aptitude à détenir et à exercer des droits et des obligations. Cependant, la proximité n'est pas l'assimilation<sup>327</sup>. La quasi-personnalité morale se rencontre quand un groupement existe et fonctionne déjà, alors que le droit ne le considère pas encore comme un sujet<sup>328</sup>. Rappelons que la société de facto est née quand une société n'a pas accompli les actes nécessaires à sa création, mais où la loi lui accorde pourtant certains droits et impose certaines fonctions comme si les actes avaient été accomplis. Prenons l'exemple du droit allemand offrant une figure de ce type, la société civile (*Gesellschaft bürgerlichen Rechts*, *BGB-Gesellschaft* ou *GbR*). Elle y est conçue comme un groupement sans personnalité

---

<sup>326</sup> BROS S., « La quasi-personnalité morale », in l'Association Henri Capitant, *La personnalité morale*, Journées nationales de l'Association Henri Capitant, t. XII, Dalloz, 2010, p. 49.

<sup>327</sup> « Ni tout à fait la même ni tout à fait une autre ». *Ibidem.*, p. 52.

<sup>328</sup> *Ibid.*, p. 53.

morale dont le régime superpose à la fois la responsabilité individuelle des associés et une responsabilité collective concernant les actes les plus graves et le patrimoine indivis<sup>329</sup>.

Hormis la voie législative, la jurisprudence n'hésite pas non plus à créer des nouvelles catégories de *fictio legis* pour mieux s'adapter à la situation. C'est le cas par exemple de l'arrêt *Weißes Ross* rendu en 2001 par Bundesgerichtshof, où la Cour fédérale allemande a reconnu pour la première fois la capacité juridique (*Rechtsfähigkeit*) d'une société commerciale des personnes agissant au vu et au su des tiers<sup>330</sup>. Ce critère « d'agir au vu et au su des tiers » attire notre attention. En effet, la société commerciale des personnes en Allemagne n'a en principe ni la personnalité juridique (*Rechtspersönlichkeit*), ni la capacité juridique. Mais si c'est une société commerciale des personnes qui agit au su et au vu des tiers, les juges allemands considèrent qu'afin de mieux protéger les tiers et de rendre le statut juridique en adaptation avec la réalité, cette reconnaissance de la capacité juridique doit alors intervenir à titre exceptionnel. En ce sens, tout en tenant compte de l'absence de la personnalité morale, une société commerciale des personnes allemande, une GmbH&Co.KG par exemple, pourrait tout de même acquérir la capacité juridique, et se trouver ainsi à mi-chemin des catégories traditionnelles. Le résultat de cette reconnaissance est qu'elle peut relever quasiment de tous les actes juridiques en tant que sujet du droit dans l'ordre juridique sans avoir toutefois de personnalité juridique.

Partant, l'on pourrait en déduire qu'une IAG, même si elle ne bénéficie pas de la personnalité juridique au moment de son acte, dès qu'elle agit en tant que sujet juridique avec un tiers, sera considérée comme un agent doté de la capacité juridique. Rappelons que cette solution se fait l'écho de la théorie de l'apparence et de la confiance légitime évoquées plus haut, il nous semble donc opportun de faire l'exception un nouveau principe ; c'est ce que l'exemple de la responsabilité pénale des personnes morales nous a appris.

## **II De *societas delinquere non potest* à la responsabilité pénale des personnes morales**

**98. Principe général de la responsabilité de personne morale.** — Aussi faut-il mentionner que pendant longtemps, la responsabilité pénale de la personne morale était inconcevable

---

<sup>329</sup> *Ibid.*, p. 60.

<sup>330</sup> BGH, Urt. v. 29. Januar 2001 - II ZR 331/00.

(*societas delinquere non potest*). En droit français, ce n'est que jusqu'à la réforme du droit pénal en 1992 que cette reconnaissance de la responsabilité pénale a été attribuée aux personnes morales<sup>331</sup>. Malgré cette attribution, au départ, ce n'était que le principe de spécialité qui s'appliquait à titre exceptionnel — la personne morale n'engageait sa responsabilité pénale que dans les cas explicitement prévus par la loi. Pour autant, ce principe de spécialité a bel et bien été abandonné. Aujourd'hui, c'est le principe général de la responsabilité de personne morale qui s'applique. Par conséquent, la personne morale voit sa responsabilité engagée dans toutes les infractions susceptibles d'être commises par les personnes physiques. Encore une fois, ce qui était inconcevable pourrait éventuellement faire partie du principe.

**99. Le renversement de la conception de principe et d'exception.** — Ces exemples témoignent que, d'une part, la perception quant à la distinction entre un sujet de droit et un objet de droit n'est pas absolue et peut évoluer au fil du temps — la frontière une fois déterminée entre les catégories juridiques pouvant toujours être remise en doute et redéfinie ; et que d'autre part, en dépit de lacune au regard des catégories existantes, en présence d'une nouvelle opération juridique, un milieu parfois ignoré est susceptible d'être redécouvert, et de nouvelles catégories pouvant ainsi être créées ex nihilo.

Suivant les considérations évoquées ci-dessus, l'on voit que la *personae fictae* est le lieu d'imputation de droits et d'obligations, qui ne contient aucune implication du corps physique, sans parler d'un corps anthropomorphique ou quoi que ce soit d'autre. Fort de cette distinction, rien n'empêcherait de conférer un statut, voire une personnalité juridique à des robots<sup>332</sup>. En effet, aucun principe n'indique à quel moment le système juridique doit reconnaître une entité en tant que personne, ni quand il doit refuser sa personnalité. Au moins de ce point de vue, rien n'empêche de traiter les IA en tant que personnes<sup>333</sup>. Cette conviction nous amène donc à étudier la personnalité juridique d'IA.

### III. La personnalité juridique d'IA

---

<sup>331</sup> Loi n° 92-684 du 22 juillet 1992 portant réforme des dispositions du code pénal relatives à la répression des crimes et délits contre les personnes.

<sup>332</sup> DE VAUPLANE H., « La personnalité juridique des robots », préc.

<sup>333</sup> ALLEN T., WIDDISON R., « Can Computers Make Contracts? », préc., p. 35.

**100. Personnes robots.** — Vu qu'un appareil perceptif, doté d'une personnalité et capable d'apprendre et de s'adapter à son environnement pourrait être suggéré comme une sorte d'artéfact ressemblant à une personne juridique<sup>334</sup>, d'aucuns ont proposé de reconnaître une personnalité juridique spécifique aux robots. En ce sens, il est suggéré dans la résolution du Parlement européen la création, à terme, d'une personnalité juridique spécifique aux robots, afin qu'au moins les robots autonomes puissent être considérés comme des personnes électroniques responsables, tenues de réparer tout dommage causé à un tiers<sup>335</sup>.

Vu que la personnalité juridique est la base des relations juridiques, elle doit être facilement et clairement discernée par un tiers de l'extérieur<sup>336</sup>. Partant, à travers ces réflexions, cette étude partage l'idée selon laquelle il serait envisageable de conférer la personnalité à l'IAG qui, caractérisée d'une compétence auto-apprentissage et de la capacité générale du niveau humain et au-delà, interagit de manière autonome au sein de la civilisation.

**101. La constitution des personnes IA.** — D'aucuns prédisent la création d'une nouvelle catégorie de sujet juridique, à mi-chemin entre personne et objet<sup>337</sup>. Cette personne sera dotée d'une personnalité juridique relevant d'une nouvelle organisation juridique à part entière — « personne robot »<sup>338</sup>. Effectivement, il pourrait s'avérer impossible de faire une distinction juridique entre les humains et les ordinateurs, soit en raison de la capacité comportementale accrue de ce dernier groupe, soit parce que les deux groupes seront littéralement, physiquement, inséparables<sup>339</sup>. Dès lors, le droit reconnaîtra au moins certains systèmes informatiques en tant que « personnes »<sup>340</sup>. À notre avis, dans un premier temps,

---

<sup>334</sup> WEIN L. E., « The Responsibility of Intelligent Artifacts: Toward An Automation Jurisprudence », préc., p. 116.

<sup>335</sup> Résolution préc. du 16 févr. 2017, pt 59 f).

<sup>336</sup> 山本 龍彦 (編集), AIと憲法, 日本経済新聞出版社, 2018, p. 210.

<sup>337</sup> Ryan Calo croit que la loi devra faire de la place pour cette troisième catégorie entre personne et instrument - un « quasi-res » destiné à capturer technologie à valence sociale. Car si les humains éprouvent de la difficulté à classer quelque chose comme étant plus semblable à un objet ou à une personne, la loi peut également y rencontrer des difficultés. CALO R., « Robotics and the Lessons of Cyberlaw », préc., p. 549, 553. Partant, la personnification des robots, qui troublerait surtout les catégories juridiques en donnant vie à une chimère, mi-personne mi-chose, à la fois sujet de droit et objet de droit, pervertit la *summa divisio* des personnes et des choses. BENSAMOUN A. et LOISEAU G., « L'intégration de l'intelligence artificielle dans l'ordre juridique en droit commun : questions de temps », *D.*, 31 mars 2017, p. 581.

<sup>338</sup> BENSOUSSAN A., « La personne robot », *D.*, 19 oct. 2017, p. 2044.

<sup>339</sup> Cf. infra Le cyborgisation dans les réalités diverses.

<sup>340</sup> WILLICK M. S., « Constitutional Law And Artificial Intelligence: The Potential Legal Recognition Of Computers as "Persons" », préc., p. 1273.

cette reconnaissance de la capacité du sujet sert à apprécier l'acte effectué d'une IAG indépendamment de la volonté de son concepteur.

Quant à sa modalité, l'on pourrait d'abord penser à l'exemple de l'article 2 de la loi coréenne intitulée « Intelligent Robots Development and Distribution Promotion Act » de 2008, modifiée en dernier lieu en 2016, qui propose une définition pour le robot intelligent : « un dispositif mécanique qui perçoit l'environnement externe pour lui-même, apprécie les situations et se déplace de lui-même »<sup>341</sup>. Cependant, à notre sens, malgré l'installation des robots dans le monde physique représentant une application importante de l'IA, il ne faut pas se limiter à cet aspect physique de la robotique pour concevoir la condition de sa personnalité juridique. En revanche, il faut toujours garder à l'esprit qu'une IAG pourrait parfaitement voir sa naissance introduite dans un environnement dit simulé. Vu que l'environnement simulé pourrait déjà être suffisamment complexe, pour que la capacité de résoudre des problèmes inédits que l'agent n'a jamais rencontrés auparavant. Ainsi, avant de traverser dans le monde physique, la présente étude considère que ces environnements simulés très riches sont suffisants à justifier la qualification d'une personne IA. En outre, il faut également noter que, de manière réciproque, l'humanité se déplace aussi dans les métaverse et omniverse composés de réseaux et d'environnements virtuels où l'on peut interagir entre avatars et avec les agents numériques. Dans ce cas, au sens de cette étude, c'est uniquement dans un second temps que l'on se pose la question de savoir si cette IAG sera dotée de faculté dite physique pour franchir la barrière dans l'espace physique, indépendamment de la notion de sa personnalité juridique.

Pour ces raisons, cette étude soutient que l'IAG pourrait très bien exister sans un support physique, vu que sa nature digitale ne l'empêche pas d'accomplir des actes juridiques, ni de s'impliquer dans des faits juridiques, ou de causer des dommages. Ainsi, cette étude favorise une définition extensive qui englobe l'aspect physique ainsi que l'espace digital pour la personnalité juridique de l'IAG. Cette considération a pour but de démontrer que l'effet juridique provenant de l'IA ne se limite pas à l'espace non-digital, notamment au regard de la capacité d'IA qui est avant tout d'état amphibien de bits-atomes.

---

<sup>341</sup> Intelligent Robots Development And Distribution Promotion Act (IRDSPA), préc.

Il en va de même pour la définition européenne, où l'on caractérise l'autonomie d'un robot en spécifiant qu'elle consiste dans la capacité de prendre des décisions et de les mettre en pratique dans le monde extérieur, indépendamment de tout contrôle ou influence extérieure. En effet, cette autonomie est de nature purement technique et avec un degré d'autonomie qui dépend du degré de complexité des interactions avec l'environnement prévu par le programme du robot<sup>342</sup>. Partant, les eurodéputés considèrent que l'autonomie des robots pose la nécessité de créer une nouvelle catégorie dotée de ses propres caractéristiques et effets spécifiques<sup>343</sup>, et qu'il convient d'établir une définition européenne commune des différentes catégories de robots autonomes et intelligents, en y intégrant des définitions de leurs sous-catégories le cas échéant, compte tenu de caractéristiques telles que la capacité d'acquisition d'autonomie grâce à des capteurs et/ou à l'échange de données avec l'environnement (interconnectivité) et l'analyse de données, la capacité d'apprentissage à travers l'expérience et l'interaction, la forme de l'enveloppe physique du robot, la capacité d'adaptation de son comportement et de ses actes à son environnement<sup>344</sup>. Encore une fois, à la lumière de la nature de l'IAG, cette étude considère que la robotique ne constitue pas la condition vers la personnalité juridique de l'IAG ; les deux recherches sont distinctes. Bien entendu, les robots seront beaucoup plus utiles et performants en se combinant avec l'IAG, et réciproquement, l'IAG acquerra de diverses facultés dans l'espace physique grâce à son support physique, mais cette incarnation physique n'est pas, à notre sens, l'essence pour parvenir à la conception de l'IAG, malgré leur éventuel *noce*<sup>345</sup>. Cette étude propose ainsi de dissocier la définition de la personnalité juridique de l'IAG et la condition mécanique du robot, avant que ces agents amphibiens se trouvent en fin de compte dans le monde « unifié » — à travers les expériences immersives à la fois virtuelles et physiques.

**102. La dualité des éléments constitutifs : l'autonomie et la capacité.** — En englobant l'espace physique et digital, l'on considère que l'individu numérique serait plus adéquat pour définir la condition juridique d'une IAG. En effet, l'individu numérique désigne « tout système informatique autonome et apprenant qui, au travers de son interface, interagit de manière systématique avec l'utilisateur, et qui évolue du fait même de sa relation avec

---

<sup>342</sup> Résolution préc. du 16 févr. 2017, pt AA.

<sup>343</sup> *Ibidem.*, pt AC.

<sup>344</sup> L'annexe à la Résolution préc. du 16 févr. 2017.

<sup>345</sup> V. p. ex. les robots humanoïdes NEO et EVE développés à travers l'entreprise suédoise 1X par OpenAI.

lui »<sup>346</sup>. Tirant ces réflexions, les critères d'autonomie et de capacité paraissent indispensables au regard de l'octroi de la personnalité juridique d'IA. À notre sens, il convient donc d'établir une *dualité des éléments constitutifs* pour construire une personne juridique IA : l'agent dispose à la fois de l'autonomie et de la capacité. L'on entend par autonomie, la possibilité pour cette IA d'effectuer des actes juridiques de manière autonome, c'est-à-dire indépendante de l'imputabilité de l'être humain. Quant à la capacité, il faut que sa compétence soit suffisamment générale, c'est-à-dire au même niveau qu'un être humain raisonnable au sens du droit civil, pour qu'elle puisse être impliquée dans des situations relevant des actes ou des faits juridiques. Dès le moment où il manifeste qu'une IA est composée de ces deux chefs d'intelligence, à savoir l'autonomie sur le plan juridique et la capacité au sens juridique, il devient nécessaire d'instaurer la personnalité juridique pour une telle composition, ou au moins commencer par l'aménagement de sa capacité juridique, dont la reconnaissance serait achevée en personnalité par excellence *de plano*.

**103. L'application *mutatis mutandis*.** — Pour savoir si l'on devrait accorder des droits aux IA ayant des capacités intellectuelles semblables à celles des humains, la réponse peut varier selon la nature du droit et notre compréhension de la justification sous-jacente du droit<sup>347</sup>. L'on retrouve ici une similitude quant aux caractéristiques et au statut de la personnalité morale qui est une assimilation aux personnes physiques. C'est par exemple en droit néerlandais, dans son code unique, l'article 2:5 BW qui dispose que « Pour ce qui est des rapports patrimoniaux, une personne morale est traitée comme une personne physique à moins que la loi n'en dispose autrement »<sup>348</sup>. Partant, l'on pourrait faire une analogie avec les personnes IA : « pour ce qui est des rapports patrimoniaux, une personne IA est traitée comme une personne physique à moins que la loi n'en dispose autrement ». De même, l'on retrouve encore cette notion d'exception pour les caractéristiques propres à l'homme à l'article 53 du Code civil suisse prévoyant que « Les personnes morales peuvent acquérir tous les droits et assumer toutes les obligations qui ne sont pas inséparables des conditions

---

<sup>346</sup> GRINBAUM A., *Les robots et le mal*, Desclée De Brouwer, 2019, pp. 78-79.

<sup>347</sup> SOLUM L. B., « Legal Personhood for Artificial Intelligences », *N.C. L. Rev.*, vol. n°4, 1992, p. 1257.

<sup>348</sup> Cet article 2:5 du Code civil néerlandais prévoit en effet une « equalisation of legal persons with natural persons ».

naturelles de l'homme, tels que le sexe, l'âge ou la parenté »<sup>349</sup>. Comme la convenance du robot est justement d'imiter des actions humaines, l'acquisition des droits et des obligations pourrait donc ne pas être totalement comprise comme une fiction<sup>350</sup>. En effet, ce mécanisme explicite pourrait bientôt inclure tout ce qui impliquerait des rapports matrimoniaux avec les IA. Ces critères doivent encore être élaborés pour des robots humanoïdes de plus en plus avancés. Par exemple, les actes de procréation avec un robot<sup>351</sup> ainsi que d'autres prérogatives<sup>352</sup> qui sont encore exclusivement attachées à l'homme, mais ne le seront plus demain, devront être régulièrement révisés.

Étant donné qu'au départ, ces agents autonomes assument des responsabilités d'un ordre ou d'un degré différent de ceux effectuant une tâche équivalente, l'on peut les considérer comme des personnes juridiques « incomplètes »<sup>353</sup>. Dans la mesure où une IA reflète cette similitude, une application analogue ne sera pas exclue. Car lorsqu'il s'agit d'aspects identiques, le principe d'égalité nous obligera à traiter un nouveau sujet provenant d'une application d'IA aux mêmes conditions que celles d'un être humain. En revanche, et pour ce qui est des caractéristiques qui sont (encore) propres à l'espèce humaine (ou aux personnes morales), l'égalité de traitement sera aussi appliquée pour un résultat réservé. Dans ce sens, la ratio legis des critères ne sera plus figée dans une conception dichotomique de principe mais se jugera éventuellement au cas par cas. Ainsi, tout comme les animaux eux-mêmes ne sont pas protégés contre la mise à mort, mais seulement lorsque la fin de leur vie est jugée cruelle et inutile compte tenu de la méthode ou des circonstances<sup>354</sup>, de même, il serait peu logique de donner aux robots un « droit à la vie »<sup>355</sup>. L'on pourrait seulement se limiter à un « droit de ne pas être éteint (right not to be terminated) »<sup>356</sup>.

---

<sup>349</sup> V. le Code civil suisse. Dans le même sens : l'article 19-3 Grundgesetz (GG) dispose que « Les droits fondamentaux s'appliquent aussi aux personnes morales nationales dans la mesure où ils leur sont applicables en raison de leur nature ».

<sup>350</sup> ROUVIÈRE F., « Le robot-personne ou *Frankenstein* revisité », RTD civ., D., 2018, p. 778.

<sup>351</sup> Cf. supra § Digisexuel et Robosexuel.

<sup>352</sup> P. ex., JAMES S., « Artificial Intelligence Should Be Allowed to Worship », NYT, 29 juill. 2019.

<sup>353</sup> WEIN L. E., « The Responsibility of Intelligent Artifacts: Toward An Automation Jurisprudence », préc., p. 107.

<sup>354</sup> L. 515-4 du Code civil français et son exposé des motifs de la proposition de la loi visant à établir la cohérence des textes en accordant un statut juridique particulier à l'animal.

<sup>355</sup> DARLING K., « Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects », préc., p. 20.

<sup>356</sup> LAWRENCE D. R., BRAZIER M., « Legally Human? 'Novel Beings' and English Law », préc., p. 324.

Tirant ces démonstrations, en attribuant la personnalité juridique aux robots, le droit fait plus que proposer une solution technique, il paraît enregistrer et valider une réalité<sup>357</sup>. Pour autant, compte tenu de l'état actuel d'IA étroite, l'on a pu constater des interrogations étant soulevées pour contester l'idée de la personnalité juridique d'IA.

**104. L'argument quant à l'inconvénient de la personnalité juridique : risque de déresponsabilisation.** — Dans un contexte d'IA faible et étroite, « l'homme doit toujours être responsable en dernier ressort de la prise de décision » et « l'introduction de l'intelligence artificielle dans le domaine de la santé en particulier doit toujours respecter le principe de responsabilité selon lequel 'l'humain contrôle la machine' »<sup>358</sup>. D'aucuns soutiennent ainsi qu'il n'est pas nécessaire de donner aux systèmes autonomes une personnalité juridique car pour eux, la responsabilité civile est une responsabilité patrimoniale, qui exige de son titulaire qu'il ait des actifs<sup>359</sup>. D'après cette vision, le préjudice causé par l'IA totalement autonome est « généralement réductible aux risques imputables à des personnes physiques ou à des catégories existantes de personnes morales, et lorsque ce n'est pas le cas, de nouvelles lois visant des individus constituent une meilleure réponse que de créer une nouvelle catégorie de personnes juridiques »<sup>360</sup> ; la personnalité juridique « reviendrait à plafonner la responsabilité et, comme l'expérience l'a montré avec les entreprises, à tenter par la suite de contourner ces restrictions en intentant des actions contre des personnes physiques ou morales auxquelles des personnes électroniques peuvent être attribuées, ce qui aurait pour effet de 'percer le voile électronique' »<sup>361</sup>. Sur ces fondements, d'aucuns considèrent que la création d'une telle personnalité ne semble pas s'imposer. Car cela risquerait de déresponsabiliser les propriétaires en raison de l'existence d'un écran juridique dans l'engagement de leur responsabilité<sup>362</sup>.

---

<sup>357</sup> ROUVIÈRE F., « Le robot-personne ou *Frankenstein* revisité », préc.

<sup>358</sup> Résolution préc. du 12 févr. 2019, pts AK et 123.

<sup>359</sup> PAGALLO U. et alii, *Liability for Artificial Intelligence and other emerging digital technologies*, op. cit., pp. 37-38.

<sup>360</sup> Pour cette raison, beaucoup croient que la personnalité juridique d'IA n'est pas nécessaire. *Ibidem.*, p. 38.

<sup>361</sup> *Ibid.* Malgré le souci de dépenalisation, il est souligné que cela ne s'agit que de la question du passif et ne prend aucune position sur l'évolution future du droit, par exemple, sur la question de savoir si l'IA pourrait agir en tant que membre d'un conseil.

<sup>362</sup> COURTOIS G., « Robots intelligents et responsabilité : quels régimes, quelles perspectives? », *Dalloz IP/IT*, 2016, p. 287.

Le souci sous-jacent est donc que les entreprises qui produisent et utilisent des IA insisteront sur les difficultés rencontrées pour déterminer la responsabilité et encourageront le mécanisme d'assurance sans égard vis-à-vis de la responsabilité. Selon cet argument, il se peut qu'il soit dans l'intérêt des entreprises de promouvoir une sorte de statut juridique indépendant en tant qu'agents de ces machines « afin de limiter les obligations financières et juridiques de ceux qui les créent et les utilisent »<sup>363</sup>. En conséquence, lorsqu'un humain délègue des décisions à des entités intelligentes, l'on se focalise sur l'humain en tant que responsable, et l'on est moins enclin à attribuer la responsabilité à la machine dans la mesure où une telle posture isolerait l'homme de toute responsabilité<sup>364</sup>.

**105. La déresponsabilisation contradictoire.** — Néanmoins, un système intelligent qui reconnaît les syndromes avec une précision supérieure à celle d'un médecin humain, et ensuite prend sa décision pour le traitement médical de manière autonome, constitue un exemple montrant qu'il n'y a aucune raison qu'une IA ne soit pas considérée comme un contractant indépendant, même dans les cas où cela exonérerait l'humain de toute responsabilité<sup>365</sup>. Partant, l'idée qu'une machine est appropriée pour avoir des obligations juridiques revient à suggérer que l'IA évolue à un point tel que certains de ces appareils ont des responsabilités, et c'est dans ce sens que sont envisagés des « agents légaux »<sup>366</sup>. Ces agents seraient dotés, en fonction de sa capacité de facto correspondant à la capacité de jure, de la capacité juridique ou de la personnalité juridique, afin d'assumer ses droits et devoirs issus de ses rapports juridiques établis sans aucune intervention humaine impliquée.

Ainsi, pour répondre à l'argument de déresponsabilisation, à notre avis, si la responsabilité juridique a pour but de prévenir les risques mais surtout de réparer les risques réalisés des dédommagements, à savoir l'indemnisation des dommages et intérêts, ce qui compte éventuellement, c'est de pouvoir identifier et engager le « responsable ». Mais dans un premier temps, si l'argument de déresponsabilisation indique « responsable » dans un sens d'imputabilité, c'est le cas dans la conception traditionnelle. Or, lorsque l'autonomie de l'IA implique aucune imputabilité humaine, selon cet argument, c'est l'inverse qu'il va produire :

---

<sup>363</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, op. cit., p.198.

<sup>364</sup> WEIN L. E., « The Responsibility of Intelligent Artifacts: Toward An Automation Jurisprudence », préc., p. 113.

<sup>365</sup> *Ibidem*, p. 116.

<sup>366</sup> *Ibid.*, p. 106.

sans imputabilité — sans responsable. Et puis dans un second temps, si cet argument voulait dire que « responsable », mais dans un sens d'une meilleure indemnisation des victimes, c'est en effet la perspective moderne de nos jours<sup>367</sup>, et alors, dans ce cas, une personnalité juridique d'IA va en réalité pouvoir garantir une indemnisation juste et adéquate par des mécanismes préventifs tels que l'assurance préalable ou le fonds collectif<sup>368</sup>. C'est ainsi qu'à notre sens, ce mécanisme du statut juridique d'IA va en effet, solidement et concrètement, « responsabiliser », et non pas « déresponsabiliser », à la fois du point de vue des responsables et des victimes ; en revanche, à défaut d'un tel encadrement, l'on pourrait s'attendre que, soit par la conception traditionnelle d'imputabilité, soit par l'adoption contemporaine d'une garantie du risque ou d'indemnisation, c'est en fait le vrai « déresponsabiliser » des rapports juridiques.

**106. La personnalité sui generis d'IAG.** — En tout état de cause, la trajectoire évolutive de l'IA nous avertit qu'il ne faut pas attribuer prématurément une autorité morale aux robots, de sorte que les humains pourraient éviter de prendre la responsabilité de leurs actes<sup>369</sup>. Bien entendu, cette étude considère que dans le contexte d'IA étroite, les humains ont non seulement une responsabilité initiale, mais aussi une responsabilité continue pour soulever les questions de compétence et de valorisation chaque fois que la prise de décision par une IA est en cause. Pour autant, rappelons que l'imputabilité peut être attribuée à un humain privé de discernement, et que le statut de la personne humaine se justifie entièrement par les mêmes chefs de considérations que celles de l'IA, à savoir l'autonomie et la capacité de facto — il serait difficile de nier la personnalité d'IAG au simple motif de déresponsabilisation. En revanche, à défaut d'une personnalité sui generis d'IAG, cela risque de créer une impression d'irresponsabilité à l'égard des rapports d'IA — de la lacune législative au déni de justice. Car le modèle de traitement, soit sur le plan psychologique et économique, soit sur le plan juridique et technique, aboutirait à une inconscience pour traiter un être comparable aux sujets de facto en tant qu'un sujet de jure, ce qui constituerait un véritable vide juridique ayant besoin de remède. Ainsi, vu la sécurité juridique, la bonne foi et la

---

<sup>367</sup> V. supra la notion de précaution, le mouvement de l'objectivation de la faute, la théorie du risque et la théorie de la garantie.

<sup>368</sup> V. infra l'assurance de fonds collectif d'IA.

<sup>369</sup> ASARO P. M., « What Should We Want From a Robot Ethic? », *International Review of Information Ethics*, vol. 6, 2006, p. 12.

prévisibilité des supports responsables voire le souci d'une meilleure indemnisation, la possibilité de concourir à une personnalité juridique d'IAG nous semble inévitable. En effet, cette appréhension n'est pas moins partagée par des observations transfrontières.

**107. La concrétisation : un chemin sans retour.** — D'aucuns prétendent que l'IA est capable de produire une simulation de la pensée, alors qu'aucune pensée n'a lieu<sup>370</sup>. Il ne peut donc y avoir de critère comportemental permettant d'accorder à une machine le statut d'être intelligent. Ce point de vue n'est néanmoins pas globalement partagé. Beaucoup d'autres considèrent que l'on a le droit de créer cette vie, et que la création d'une nouvelle conscience dans une société où les intelligences humaines et artificielles cohabitent exigera une protection égale aux machines de leurs droits respectifs<sup>371</sup>. Stephen Hawking, qui a estimé que « ce serait une très grande erreur de croire que la machine intelligente n'est qu'un ingrédient de science-fiction », a considéré que la création d'une personnalité électronique « afin d'établir les droits et les responsabilités des systèmes intelligents les plus avancés semble logique »<sup>372</sup>.

En réalité, laisser la vraie IA en dehors de nos attentes reviendrait à vivre dans un monde fantastique<sup>373</sup>. Pour autant, dans la lettre ouverte à la Commission européenne, beaucoup ont considéré que « créer une personnalité juridique pour un robot est inapproprié quel que soit le statut légal envisagé »<sup>374</sup>. Ils prétendent que « l'utilisation accrue des machines dans les décisions affectant les personnes, risque de dégrader le statut et la dignité des humains »<sup>375</sup>. Cependant, cette étude considère que l'intelligence créée par l'humain n'est pas une défaite pour l'intelligence humaine ; bien au contraire, elle est un grand achèvement<sup>376</sup> de la

---

<sup>370</sup> V. p. ex. John Searle.

<sup>371</sup> Gary Drescher. V. TURKLE S., *The Second Self: Computers and the Human Spirit*, op. cit., pp. 238-243.

<sup>372</sup> HAWKING S., *Brèves réponses aux grandes questions*, op. cit., pp. 192-197.

<sup>373</sup> DREXLER K. E., *Engines of Creation: The Coming Era of Nanotechnology*, Doubleday, 1986, p. 70.

<sup>374</sup> 252 signatures, dernière visite du site web le 16 mai 2018. Pour Nathalie Nevejans, « ce statut engendrerait des effets juridiques indésirables ». V. son « Règles européennes de droit civil en robotique », étude pour la commission des affaires juridiques du Parlement européen, oct. 2016.

<sup>375</sup> RUSSELL S., *Human Compatible*, op. cit., pp. 126-127.

<sup>376</sup> Richard Sutton: « The case for ambition in artificial intelligence research: Within your lifetime, AI researchers will understand the principles of intelligence — what it is and how it works — well enough to create beings of far greater intelligence than current humans. This will change everything. The way we work and play. Our senses of identity. The goals we set for ourselves and our societies. It will be the greatest intellectual achievement of all time. An achievement of science, of engineering, and of the humanities, whose significance is beyond humanity, beyond life, beyond good and bad. I call it the Prize. The Prize is a great and glorious goal! Ambitious AI researchers should keep their Eyes on the Prize ». Tweet du 30 sept. 2022.

civilisation humaine. En outre, la tendance est claire : volens nolens, les machines prennent des décisions à des niveaux de plus en plus divers et dans de nombreux domaines. Donc malgré les oppositions fondées sur la primauté de l'espèce humaine, les exemples concrétisant un paysage discordant peuvent être trouvés dans différents endroits du monde.

En Inde, il est indiqué dans le rapport publié par le gouvernement indien sur la Task Force de l'intelligence artificielle que « les dispositions légales applicables aux utilisateurs (individuels ou corporatifs) des systèmes d'IA doivent continuer à s'appliquer aux machines autonomes de façon mutatis mutandis » et que « les droits et les responsabilités des entités autonomes constitue un champ qui doit être examiné »<sup>377</sup>.

L'Europe, pour sa part, vise à « mettre au point des stratégies et des campagnes de communication afin de sensibiliser la population aux avantages de l'utilisation des robots dans la vie quotidienne ». Elle « prend acte notamment de l'ambition de la stratégie du Japon en matière de robotique »<sup>378</sup>. En effet, le Japon est entré dans une société vieillissante et le nombre d'utilisateurs des services de soins infirmiers est passé à environ 6,14 millions en 2016. Le soutien des soignants par des interventions robotisées est en cours de développement<sup>379</sup>. Le gouvernement japonais estimait qu'en 2020 les soins infirmiers utilisant la technologie robotique atteindraient 80%<sup>380</sup>. Dans un pays où l'application de l'IA est une aventure sans retour, un robot virtuel, à l'image d'un petit garçon de 7 ans, qui s'appelle 未来, « futur » en japonais, s'est vu accorder le statut officiel de résident du quartier Shibuya de Tokyo. L'exécutif municipal est ainsi le premier à délivrer des cartes de résident spéciales aux personnages d'IA<sup>381</sup>.

---

<sup>377</sup> « Report of The Artificial Intelligence Task Force », constituted by Ministry of Commerce and Industry, Government of India, 21 mars 2018, pp. 46-47.

<sup>378</sup> Résolution préc. du 12 févr. 2019, pt 83.

<sup>379</sup> 藤田 卓仙, « 人工知能・ロボットと労働・雇用をめぐる視点: 科学技術に関する調査プロジェクト報告書, 国立国会図書館 調査及び立法考査局, 30 mars 2018, pp. 50-51.

<sup>380</sup> Japan's Robot Strategy, The Headquarters for Japan's Economic Revitalization, févr. 2015, p. 66.

<sup>381</sup> CUTHBERTSON A., « Tokyo : artificial intelligence 'boy' Shibuya Mirai becomes world's first AI boy to be granted residency », Newsweek, 6 nov. 2017. V. aussi WESSBÉCHER L., « Au Japon, une intelligence artificielle est candidate aux élections municipales de Tama », 13 avr. 2018, France24.

La même tendance se trouve en Arabie saoudite, où un robot dénommé Sophia s'est vu attribuer la nationalité saoudienne<sup>382</sup>. Divulguée à l'occasion d'un sommet pour les investissements, cette annonce vise à promouvoir le développement de l'IA du pays. Néanmoins, en réalité, il y a une interdiction de rapprochement avec toute image de l'Homme dans la culture traditionnelle arabe, qui n'est donc pas prête à accepter cette qualification. L'attribution de la nationalité à Sophia « n'est évidemment pas un acte fait par un pratiquant parce que dans la religion islamique, ce n'est pas concevable. Pour autant, comme il s'agit de l'objet, dans le cadre de l'intérêt général, le droit arabe ne s'oppose pas à ce que les robots aient une intelligence artificielle »<sup>383</sup>.

**108. Condition préalable de la culture diverse de l'humanité.** — Cette réticence, voire ce refus de l'attribution juridique dans la culture juridique islamique, montre la difficulté dite particulariste quant à la mise en place d'un rapprochement de régime juridique en IA. « C'est pourquoi il est aussi difficile de porter, d'un point de vue culturel, un jugement positif sur l'entreprise d'uniformisation du droit »<sup>384</sup>. Cependant, au XXI<sup>ème</sup> siècle, pour qu'un système juridique puisse contenir un sens culturel, il doit avant tout être un système juridique ayant pour but de protéger les intérêts juridiques communs de l'espèce humaine. Un régime juridique uniforme du XXI<sup>ème</sup> siècle, s'il souhaite être capable d'avoir les moyens de défendre la convivialité culturelle, et non pas de l'empêcher, devrait reposer sur un fondement commun : objectif de l'humanité. Autrement dit, l'arrivée de l'IAG nous oblige à reconnaître le fait que l'humanité est la condition préalable de Sa culture. À titre d'illustration, la sauvegarde d'une diversité par voie d'universalité est en effet manifestée par « *in varietate concordia* » de l'approche européenne contemporaine : « l'IA devrait

---

<sup>382</sup> JOSSET C., « Le robot Sophia, citoyenne saoudienne, a plus de droits que les femmes de son pays », *l'express*, 28 oct. 2017.

<sup>383</sup> DIB D., « L'intelligence artificielle et Chari'a », à l'occasion du colloque international « L'intelligence artificielle : un nouveau défi pour le droit? », organisé par l'Université d'Alger 1 Benyoucef Benkhedda, Alger, 27-28 novembre 2018. D'autres discussions sur l'argument théologique, v. p. ex. SOLUM L. B., « Legal Personhood for Artificial Intelligences », préc., pp. 1262-1263 : L'argument selon lequel les IA manquent d'âme repose sur une prémisse théologique controversée. Décisions politiques et judiciaires devraient être faites en accord avec l'exigence de la raison publique. L'exigence de la raison publique implique que les décisions politiques et judiciaires doivent être justifiées par des motifs publics. La raison publique ne peut pas s'appuyer sur des conceptions particulières religieuses ou philosophiques du bien. L'exigence de raison publique exclurait l'utilisation d'arguments religieux sur les âmes dans une décision judiciaire sur le statut constitutionnel de l'IA. Quels que soient les mérites théologiques que l'on pourrait tirer de l'argument prétendant que les IA manquent d'âme, cela ne devrait pas marcher dans le contexte juridique.

<sup>384</sup> LEQUETTE Y., « Vers un code civil européen ? », *POUVOIRS* 2003/4 (n°107), p. 114.

pleinement prendre en charge l'ensemble des langues européennes afin d'offrir à tous les Européens les mêmes chances de tirer parti de l'évolution moderne de l'IA au sein de la société de l'information multilingue européenne »<sup>385</sup>.

De même que les cultures ainsi que les langues européennes minoritaires sont préservées et protégées par un droit communautaire, à vrai dire, en l'absence d'un régime uniforme de jure, cela déboucherait sur une véritable contrainte de facto. C'est donc paradoxalement, si l'on refuse un mécanisme commun portant sur la coexistence de l'homme et l'IA, l'on serait privé d'un moyen efficace pour défendre les diversités de la civilisation humaine. Vu que désormais la civilisation humaine dépend largement voire entièrement des technologies qui sont sous le contrôle des entreprises privées et des régimes autoritaires, lorsque la transition numérique n'est plus d'une option, mais de la survie, la diversité de la majorité ne peut qu'être silicolonisée et surveillée pour se conformer aux valeurs prescrites par la puissance d'une minorité d'acteurs. En conséquence, comme la protection des robots revient à la protection des humains, un cadre juridique *prima facie erga omnes*, revient éventuellement à assurer toute particularité restant dans l'humanité.

**109. L'argument *a fortiori* pour une reconnaissance implicite.** — Revenons donc sur l'état actuel de la personnalité juridique des IA. Comme les gouvernements ne communiquent jamais directement sur la personnalité juridique d'IA, il résulte de l'argument *a fortiori* que l'attribution de la qualité de citoyen saoudien ou de résident japonais est déjà incluse dans sa conception. L'on peut donc se demander si aux yeux de ces gouvernements, ces bots sont, de manière implicite, dotés de la personnalité juridique résultant de la citoyenneté prescrite par ces pays respectifs. En tout état de cause, soit le citoyen saoudien soit le résident japonais, l'on pourrait remarquer que dans la pratique, les statuts d'IA pourraient d'ores et déjà être appréciés en suivant les variations de leurs fonctions.

**110. En résumé.** — En mesurant les capacités de ces entités artificielles, un lexique pour répartir droits et responsabilités aux analogues de l'âge humain serait alors amené à exister. En tenant compte des statuts en fonction d'une grille d'évaluation, les IA seraient dotées de droits et d'obligations conformément à leur niveau individuel<sup>386</sup>. Par exemple, un test qui

---

<sup>385</sup> Résolution préc. du 12 févr. 2019, pt. AE.

<sup>386</sup> WILLICK M. S., « Artificial Intelligence: Some Legal Approaches and Implications », préc., p. 10.

peut révéler la capacité d'exercer les traits de personnalité peut être mis en place, et le plus pertinent serait celui utilisé chez l'homme : le comportement<sup>387</sup>. Ensuite, à mesure que les capacités des IA s'améliorent, leurs droits et obligations devraient être pareillement disponibles pour agir d'une manière équivalente dont l'on attend que les humains agissent sur le plan fonctionnel. Cela n'exige pas que l'artefact ait une compréhension universelle et complète de la loi, pas plus qu'un homme raisonnable<sup>388</sup>. Les prochaines étapes consistent en conséquence à clarifier les distinctions entre les objectifs, et à distinguer les différents types d'agences<sup>389</sup>. Pour ce faire, il convient d'introduire un registre d'IA afin d'éclaircir le sens de cette évolution.

---

<sup>387</sup> *Ibidem*, p. 11.

<sup>388</sup> CALVERLEY D. J., « Imagining a non-biological machine as a legal person », préc., p. 534.

<sup>389</sup> ASARO P. M., « The Liability Problem for Autonomous Artificial Agents », AAAI Spring Symposia 2016, p. 4.

### **Conclusion de la section 1**

**111. De la personne IA au registre IA.** — En examinant la notion de la personnalité juridique, l'on a démontré le détachement du concept de l'homme et celui de la personnalité juridique. Plusieurs aménagements factuels tels que les statuts juridiques des femmes et des esclaves, ainsi que les exemples tels que la quasi-personnalité et la capacité juridique octroyée à des catégories juridiques dites à mi-chemin illustrent le renversement du principe et de l'exception.

Ensuite, afin de retenir une définition de personne au sens du sujet de notre recherche, cette étude propose de dissocier la définition de la personnalité juridique de l'IAG et la condition mécanique du robot. Ainsi, l'on a dégagé une dualité des éléments constitutifs pour la condition de la personne IAG, à savoir l'autonomie sur le plan juridique et la capacité au sens juridique.

Enfin, l'on a illustré la contradiction dans les arguments quant à l'inconvénient de la personnalité juridique tels que le risque de déresponsabilisation ou la dégradation de la dignité humaine. Pour conclure, l'on a constaté la concrétisation des statuts d'IA dans différents coins du monde, *volens nolens*. Reste à encadrer ce chemin sans retour vers le registre.

## **Section 2 Le registre d'IA**

« *Software is eating the world. But AI is going to eat software.* » — Jensen Huang

**112. Annonce du plan.** — Les technologies NBIC (Nanotechnologies, Biotechnologies, Informatique et sciences Cognitives) connaissent un développement exponentiel et vont bouleverser l'humanité<sup>390</sup>. Les pays ayant des politiques d'IA plus sensées avanceront plus rapidement, et ceux qui ont des politiques mal conçues risquent d'être laissés pour compte<sup>391</sup>. C'est pourquoi il faut une Magna Carta multipartenariale<sup>392</sup> pour nous guider collectivement dans le développement de l'IAG, afin de poser les fondations de la coexistence des humains et des IAG. Cette étude propose que ce cadre puisse être opéré dans un premier temps par un registre des systèmes d'IA (§1) avec deux modèles d'immatriculation (§2).

### **§1 Identification des systèmes d'IA**

**113. Annonce du plan.** — Au regard des risques voire des dangers non négligeables posés par le développement de l'IA, en vue d'assurer la sécurité juridique ainsi que de protéger les tiers de bonne foi, un mécanisme de vérification et d'indemnisation est envisagé. En pratique, pour qu'un système d'IA soit certifié (I), une assurance de fonds collectif (II) peut être mise en place.

### **I. L'interchangeabilité du contrôle**

**114. De la certification à l'immatriculation.** — Le Parlement européen estime qu'« il y a lieu de considérer les lois d'Asimov comme s'appliquant aux concepteurs, aux fabricants et aux opérateurs de robots, y compris de robots dotés d'autonomie et de capacités d'apprentissage, étant donné que lesdites lois ne peuvent être traduites en langage de

---

<sup>390</sup> LAURENT A., *La guerre des intelligences : Comment l'intelligence artificielle va révolutionner l'éducation*, op. cit., pp. 12-13.

<sup>391</sup> Andrew Ng, in SHOHAM Y. et alii, « AI Index 2017 Annual Report », Human-Centered AI Initiative, Stanford University, nov. 2017, p. 54.

<sup>392</sup> NITWBERG M. et alii, « L'intelligence artificielle a besoin d'une Magna Carta », HBR France, 19 déc. 2017.

programmation »<sup>393</sup>. Cela signifie que dans un premier temps, le sujet envisagé par la réglementation est effectivement les humains et non pas un sujet issu d'une nouvelle catégorie. Ensuite, dans le rapport concernant les voitures autonomes rédigé par la Commission d'éthique allemande, la certification des systèmes automatisés n'est justifiable que s'il promet une diminution du danger par rapport à la conduite humaine, c'est-à-dire un équilibre positif des risques. C'est donc à ce stade, non seulement les humains, les systèmes ont également besoin d'être certifiés par les contrôles officiels<sup>394</sup>.

Donc pour commencer, l'on voit qu'il y a deux catégories de sujets qui sont envisagées dans la régulation d'IA, à savoir les êtres humains et les systèmes d'IA. Et comme l'évoqué plus haut, en vertu du principe de précaution, la catégorie des acteurs humains englobe tous les stades du développement de l'IA. Cela signifie que des concepteurs, fabricants, aux opérateurs voire les utilisateurs des IA sont dans l'ensemble soumis aux exigences respectives en la matière. À cela s'ajoute une deuxième catégorie de sujet qui est aussi compris dans le contrôle. Dès lors, l'on comprend que pour les humains, l'appellation va sous le nom du permis d'exercice et la certification tandis que pour les systèmes d'IA, la dénomination est l'immatriculation et le registre. Cette conception peut être incontestablement retenue dans le contexte de l'IA étroite.

En revanche, en matière d'IAG, il est intéressant de constater qu'il se peut que l'IAG doive bientôt se conformer aux exigences prescrites afin d'obtenir le permis d'exercice et la certification alors que les humains soient soumis à l'immatriculation et au registre pour être authentifiés et protégés. Dans tous les cas, rappelons simplement pour le moment que la maîtrise du risque est cruciale et que les humains et les systèmes d'IA sont également concernés dans toute circonstance pour la diminution des risques. Partant, l'immatriculation des systèmes d'IAG constitue l'un des piliers au regard du contrôle des risques.

**115. L'immatriculation des systèmes IA.** — Les eurodéputés estiment ainsi qu'« il y a lieu de créer un système d'immatriculation des robots avancés sur la base de critères de classification des robots, à des fins de traçabilité et en vue de faciliter la mise en œuvre de

---

<sup>393</sup> Résolution préc. du 16 févr. 2017, pt. T.

<sup>394</sup> « Automatisiertes Und Vernetztes Fahren », *op. cit.*, pp. 6, 10.

recommandations ultérieures »<sup>395</sup>, et que « le système d'immatriculation et le registre devraient être instaurés à l'échelle de l'Union »<sup>396</sup>. L'idée d'une immatriculation au registre rendrait en effet une application plus rassurante et plus crédible, notamment au motif de la connaissance et de la protection. Le Parlement se propose donc de « définir des critères de classification des robots dont l'inscription au registre devrait être obligatoire »<sup>397</sup>, et à « réfléchir à l'opportunité de confier la gestion du système d'immatriculation et du registre à une agence spécifique de l'Union chargée de la robotique et de l'intelligence artificielle »<sup>398</sup>. À cet égard, la Commission européenne a en effet publié en 2021 sa proposition de règlement établissant des règles harmonisées sur l'intelligence artificielle<sup>399</sup>, dans laquelle une classification des systèmes d'IA en fonction du risque a été prévue. L'on distingue ainsi, selon ce texte provisoire soumis à la discussion au moment de notre rédaction, les pratiques prohibées, les pratiques à haut risque soumises à un régime de conformité, et certains systèmes d'IA faisant l'objet d'obligations de transparence.

Ces modèles et engagements nous montre que d'une part, l'argument de Law of the Horse remarquée plus haut semble définitivement rejeté par la régulation européenne en matière d'IA. D'autre part, il faut aussi constater que ces initiatives relevant de l'IA étroite qui ne concerne que certains domaines et pratiques spécifiques est déjà soumis au contrôle de certification, a fortiori, l'IAG devrait évaluée et encadrée pour des motifs encore plus perçants. En effet, si certains systèmes d'IA tels que les systèmes d'IA à finalité générale (foundation models) et les systèmes d'IA à usage général (general purpose AI systems) seraient déjà soumis au contrôle<sup>400</sup> sur le fondement du risque, bien entendu, il en va de même pour les acteurs humains qui les conçoivent, fabriquent, opèrent voire utilisent.

**116. Le statut à haut risque des concepteurs d'IA.** — À cet égard, il devrait y avoir l'équivalent du serment d'Hippocrate (Hippocratic Oath)<sup>401</sup> auquel s'engagent les médecins, afin que certaines règles puissent être respectées et être intégrées aux systèmes d'IA, tel est

---

<sup>395</sup> L'annexe dans la Résolution préc. du 16 févr. 2017.

<sup>396</sup> *Ibidem*.

<sup>397</sup> Résolution préc. du 16 févr. 2017, pt. 2.

<sup>398</sup> *Ibidem*.

<sup>399</sup> Proposition de règlement établissant des règles harmonisées sur l'intelligence artificielle, 21 avr. 2021.

<sup>400</sup> Amendements du Parlement européen, adoptés le 14 juin 2023, à la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle.

<sup>401</sup> « Yann Le Cun: Deep Learning, Convolutional Neural Networks, and Self-Supervised Learning », préc.

le cas du Serment Holberton-Turing<sup>402</sup>. De toute évidence, le développement d'IA, notamment lorsqu'il s'agit d'IAG, devrait être considéré en tant que profession ayant un statut équivalent à celui d'autres professions à haut risque (high-risk professions). Vu que les médecins, les avocats doivent détenir un permis d'exercer, il serait illogique que soit réservé des permis à ces professions mais pas à la profession chargée de développer des systèmes techniques pour augmenter ou remplacer l'expertise humaine et la prise de décision en leur sein. Justement, les risques des professions autorisées ne se sont pas dissipés, mais ont plutôt été déplacés vers l'IA<sup>403</sup>. D'où l'idée d'une assurance d'IA.

## II. L'assurance de fonds collectif d'IA

**117. Mutualisation des risques et la logique assurantielle.** — Comme les objectifs des parties concernées ne sont pas nécessairement alignés, une solidarité assurantielle peut être mise en place. Un tel système de responsabilité devrait à la fois favoriser une innovation bénéfique tout en offrant une indemnisation juste et adéquate aux personnes lésées. Des mécanismes déjà utilisés dans le droit de la responsabilité, tels que la responsabilité conjointe et solidaire, la responsabilité stricte et la mutualisation des risques (risk pool) ont ainsi été envisagés<sup>404</sup>. Dans l'hypothèse où aucun agent humain ne peut endosser la responsabilité des actions commises par une IA, la logique est d'en déduire que l'agent artificiel lui-même devrait en être responsable. Dans cette perspective, certains proposent de créditer tout système informatique apprenant d'un capital financier ; ainsi, lorsque la question de l'indemnisation se pose, il suffirait de lui confisquer son capital, puis de le répartir entre les victimes<sup>405</sup>. En effet, la mise en place d'un régime d'assurance obligatoire ou d'un fonds d'indemnisation a aussi été envisagée par le Parlement européen pour que les fabricants ou les propriétaires de certaines catégories de robots soient tenus de contracter une police d'assurance couvrant les dommages potentiels causés par les robots<sup>406</sup>.

---

<sup>402</sup> Initié dans l'objectif de fédérer et mettre en conscience tout professionnel du domaine de l'intelligence artificielle, au niveau mondial, autour de valeurs morales et éthiques communes.

<sup>403</sup> MITTELSTADT B., « AI Ethics – Too Principled to Fail? », Oxford Internet Institute, 20 mai 2019, p. 11.

<sup>404</sup> ASARO P. M., « The Liability Problem for Autonomous Artificial Agents », préc., p. 3.

<sup>405</sup> GRINBAUM A., *Les robots et le mal*, *op. cit.*, pp. 167-168.

<sup>406</sup> Résolution préc. du 16 févr. 2017, pt. 59 a) et b). Déjà, le Parlement européen constate en effet que la prévalence de véhicules autonomes présentera des risques en matière de confidentialité des données et de défaillances techniques et qu'elle va transférer à l'avenir la responsabilité du conducteur vers le fabricant, imposant aux compagnies d'assurance de modifier la manière dont elles intègrent le risque dans leur souscription. Résolution préc. du 12 févr. 2019, pt. 91. V. aussi p. ex. art. 33 §8 de la proposition de règlement

Désormais, la logique contentieuse s'effrite devant une logique assurantielle qui opère une séparation totale de la question de l'indemnisation par rapport à celui de la responsabilité pour une indemnisation intégrale. Ce phénomène de socialisation des risques dans l'assurance risquera néanmoins de rendre moins évident le rôle préventif de la responsabilité. Car au final, le fonds n'appuie pas véritablement sur le concept d'imputabilité du dommage, mais règle plutôt la contribution d'un groupe chargé de financer un risque<sup>407</sup> ; la façon dont il est financé refléterait le rôle joué par différents acteurs sur le marché de la robotique<sup>408</sup>.

**118. En résumé.** — Vu l'autonomie croissante de l'IA, une telle évolution juridique d'immatriculation « assurantielle » semble inéluctable. La question se pose donc d'un tel transfert graduel de responsabilité au registre.

### §2 Le registre : du système objectif de l'IA étroite au système subjectif de l'IA générale<sup>409</sup>

**119. Annonce du plan.** — Après avoir démontré l'adoption du mécanisme de registre, l'on s'interroge sur les éléments de ce registre (I) en distinguant entre le registre facultatif et constitutif (II), et le registre obligatoire et déclaratoire (III). En outre, afin de pouvoir se conformer à une telle prescription, la responsabilité de sa mise en œuvre pourrait éventuellement incomber aux États sous une forme de devoir de vigilance dit étatique (IV).

#### **I. Les facteurs d'immatriculation**

**120. Le registre d'IA.** — Alors qu'il est clair qu'une personne physique vivante est née, il n'est pas si évident de juger de la naissance d'une IAG sur le plan juridique sauf si elle satisfait à une exigence juridique d'établissement. Ainsi, lorsque l'on attribue une

---

établissant des règles harmonisées sur l'intelligence artificielle, 21 avr. 2021 : « Les organismes notifiés souscrivent pour leurs activités d'évaluation de la conformité une assurance de responsabilité civile appropriée à moins que cette responsabilité ne soit couverte par l'État membre concerné sur la base de la législation nationale ou que l'évaluation de la conformité ne soit réalisée directement par cet État membre ».

<sup>407</sup> LALLOUCHE S. « L'atténuation de la rigueur du système de la responsabilité pour faute dans le domaine médical », préc., p. 34.

<sup>408</sup> PALMERINI E., « Towards a Robotics law at the EU level? », préc., p. 69.

<sup>409</sup> L'intégralité de la présente étude a été rédigée avant la publication de la Proposition du règlement du 21 avril 2021 établissant des règles harmonisées sur l'intelligence artificielle. Les idées proposées et analysées dans cette thèse sont menées indépendamment de ledit règlement qui est encore soumis à la discussion au moment de notre rédaction et ne fait pas partie de notre étude concernant l'IA générale voire forte.

personnalité juridique à une IAG, la création des sociétés peut être prise en compte en tant que modèle. Autrement dit, comme pour les personnes morales, le statut de société ne serait accordé qu'après l'accomplissement des processus d'enregistrement. Ainsi, tout comme le registre rend une société identifiable, un système similaire devrait être développé pour que ceux qui doutent du statut de la société puissent connaître l'existence de personnes morales en consultant le registre. Pour ce faire, un nouveau système, semblable à une société, tout en faisant référence à son propre domaine<sup>410</sup>, peut être envisagé dans un premier temps. En créant une personnalité juridique à l'IA, des dispositions doivent être présentées pour protéger les cocontractants à des transactions ainsi que toute personne interagissant avec ces systèmes intelligents. Ce mécanisme va obliger les concepteurs utilisant l'IA à enregistrer leurs sujets en tant qu'agent. Et sur ce réseau, les parties prenantes pourraient s'identifier comme les responsables du système<sup>411</sup>.

**121. Les éléments d'immatriculation.** — Comme les robots sont conçus pour être des machines polyvalentes sans tâches entièrement prédéfinies, cela rend difficile l'identification des différentes utilisations possibles et les divers contextes dans lesquels les robots seront déployés<sup>412</sup>. Les données, les modèles et le projet qu'elle sert, sont des éléments envisageables quant aux facteurs d'immatriculation. En effet, le Parlement européen favorise « la création d'un numéro d'immatriculation individuel, inscrit dans un registre spécifique de l'Union, afin de pouvoir toujours associer un robot au fonds dont il dépend; ce numéro permettrait à toute personne interagissant avec le robot de connaître la nature du fonds, les limites en matière de responsabilité en cas de dommages matériels, les noms et les fonctions des contributeurs et toute autre information pertinente »<sup>413</sup>.

---

<sup>410</sup> 山本 龍彦 (編集), AIと憲法, *op. cit.*, p. 210-211.

<sup>411</sup> ALLEN T., WIDDISON R., « Can Computers Make Contracts? », préc., p. 42. V. p. ex. Annexe VIII, Amendements du Parlement européen, adoptés le 14 juin 2023, à la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle.

<sup>412</sup> PALMERINI E., « Towards a Robotics law at the EU level? », préc., p. 67.

<sup>413</sup> Résolution préc. du 16 févr. 2017, pt 59 e). V. aussi p. ex. Annexe VIII, Amendements du Parlement européen, adoptés le 14 juin 2023, à la proposition de règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle.

Partant, un registre spécifique est inévitable, il reste à déterminer sa mise en place en fonction de l'état du déploiement. À notre avis, le caractère facultatif ou constitutif, et la nature obligatoire ou déclaratoire du registre, constitue la première étape pour un tel aménagement.

## **II. Registre facultatif et constitutif du système objectif de l'IA étroite et de l'IA large**

**122. Registre facultatif et constitutif du système objectif.** — Pour explorer la distinction entre les différentes applications d'IA, cette étude considère qu'au lieu de se rendre dans un système subjectif qui définit d'abord le sujet de l'acte, le système objectif cherche à encadrer les comportements interdits, et à définir les actes réglementés, ainsi que les résultats qui seront sanctionnés. En ce sens, au départ, cela peut être au passage du régime d'un demi-sujet, qui sera un registre facultatif et constitutif. À ce stade, le système d'IA n'est pas obligé de se faire immatriculer au registre désigné. Cependant, s'il le fait, il y aura un changement de son statut juridique par rapport aux relations juridiques existantes. En revanche, si l'enregistrement n'a pas été suivi, le système d'IA qui ne dispose pas d'une garantie certifiée aurait donc moins de crédibilité ou de protection vis-à-vis des utilisateurs. Dans ce cas, la condamnation viendrait plutôt du marché et non pas d'une sanction du pouvoir contraint stricto sensu. En outre, au lieu d'opter pour un système « Konzessionsystem » où il faut une décision discrétionnaire provenant du pouvoir de l'autorité, ce registre facultatif de l'IA devrait être un système « Normativbestimmungen » : dès que les critères sont remplis, le système aura le droit d'être immatriculé. Prenant l'exemple de Tesla bot Optimus, tant qu'il s'agit d'une IA étroite, il n'est pas obligé de se faire immatriculer, mais il a le droit d'être certifié par voie d'immatriculation, et l'autorité ne peut pas le refuser lorsqu'il a rempli les critères pour un tel registre. Cette démarche facultative dont l'effet juridique va créer un statut constitutif pour l'objet IA certifié sur le plan juridique, par rapport aux systèmes qui ne cherchent pas à obtenir le même degré de garantie et donc d'attractivité. Mais dès que Tesla Optimus aura atteint l'intelligence générale voire forte, en revanche, son registre ne devrait plus être considéré facultatif et constitutif mais obligatoire et déclaratoire.

## **III. Registre obligatoire et déclaratoire du système subjectif d'IA générale, voire forte**

**123. Registre obligatoire et déclaratoire du système subjectif.** — Dans ce stade du système subjectif, au moment du registre, le statut de l'IA immatriculée au sujet du droit

pourrait s'aligner sur ce qui est prévu à l'article 54 du Code civil suisse, qui dispose que « Les personnes morales ont l'exercice des droits civils dès qu'elles possèdent les organes que la loi et les statuts exigent à cet effet »<sup>414</sup>. Par exemple, dès que Tesla Optimus aura doté de l'IAG disposant la faculté telle que d'être compagnon ou d'entreprendre indépendamment dans ses propres projets du développement, il convient de désigner les éléments juridiques autour de ces rapports qui ne visent plus à accomplir une tâche entièrement prédéterminée par l'homme, mais qui forment un rapport juridique dit personnel au sein de la civilisation d'après son autonomie (en l'absence de la prédétermination ainsi que de l'imputabilité de l'homme). À ce stade, l'IA sera titulaire de droits et de devoirs et ainsi capable d'agir sur le plan juridique tout en respectant les normes du système d'immatriculation. Dès lors, les normes et les procédures interétatiques doivent être élaborées pour l'IA afin de remplir les exigences des garanties, avec les droits et les obligations qui en découlent.

Ce registre d'IAG ne devrait plus être facultatif et constitutif. En revanche, il deviendrait obligatoire et déclaratoire. Tous les projets, systèmes et opérations relevant de l'IAG seront obligés de se faire immatriculer au registre d'IAG. C'est exactement par cette exigence du caractère impératif que ce registre relèvera de l'effet déclaratoire. Les parties prenantes se verront doter des obligations provenant d'un tel consensus normatif. Ainsi, un registre de caractère emblématique sera impliqué, et le statut juridique du système d'IAG ne sera pas affecté, même en dépit d'absence d'immatriculation. À défaut du registre, un système d'IAG est tout de même tenu de respecter les principes reconnus et de se conformer aux normes prévues. Ce système subjectif de long-terme, en s'opposant au système objectif de moyen-terme, définira alors les sujets à qui seront octroyés les droits et les devoirs.

En effet, après avoir étudié les concepts de certification et d'immatriculation en matière du statut de l'IA, l'on pourrait se poser la question : Quis custodiet ipsos custodes? Autrement dit, en cas de défaillance des acteurs développant l'IAG dans cette transition juridique à la fois de numériser la responsabilité et de responsabiliser le numérique, quid de l'éventuelle non-conformité ? À notre avis, à défaut des mesures concrètes pour atténuer et diminuer les risques existentiels pour l'humanité posés par l'IAG provenant de chaque système juridique, il y aurait lieu d'imposer un *devoir de vigilance étatique* aux gouvernements respectifs.

---

<sup>414</sup> Art. 54, Code civil suisse préc.

#### IV. De la responsabilité conjointe voire *in solidum* des États en cas de manquement de surveillance<sup>415</sup>

**124. Le devoir de vigilance étatique.** — Tant que la civilisation humaine se trouve encore sur le chemin construisant la transition pacifique de sa coexistence avec l'IA par le droit interespèce<sup>416</sup>, les acteurs étatiques demeurent les autorités principales de court terme. Ainsi, en ce qui concerne l'utilisation transfrontière des systèmes intelligents, une fois que le système de registre obligatoire est mis en place, il y aura lieu de condamner les États concernés en cas d'omission du rôle de supervision, tel que le « *devoir de vigilance étatique* ». À notre sens, les obligations qui incombent aux États se trouvent dans l'obligation du registre a priori et l'obligation de suivre les normes a posteriori<sup>417</sup>. Cette solution apparaît raisonnable au regard de la situation actuelle où les pays développant les IA avancées disposent des moyens pour d'une part, réagir aux risques, et d'autre part répondre aux dommages. C'est donc la garantie que les États dominants en matière d'IA soient plus susceptibles d'assurer l'exigence minimale pour la sécurité globale.

En effet, dès aujourd'hui, l'on pourrait déjà remarquer d'une part, une multiplicité d'instances régulatrices<sup>418</sup> dans les exemples tels que le RGPD ainsi que la proposition du

---

<sup>415</sup> V. p. ex. la notion d'obligations positives de l'État à travers les arrêts de la CEDH, p. ex. l'arrêt Cordella et autres c/ Italie (n° 54414/13) rendu par la CEDH le 24 janvier 2019, retenant que « Les États ont avant tout l'obligation positive, en particulier dans le cas d'une activité dangereuse, de mettre en place une réglementation adaptée aux spécificités de ladite activité, notamment au niveau du risque qui pourrait en résulter » (§159). V. aussi : « La consolidation historique d'un droit international des droits de l'homme implique le dépassement de l'ancien principe de la compétence exclusive de l'État et son remplacement par une nouvelle conception qui définit les droits de l'homme comme une question d'intérêt international. Cette approche établit un modèle de coopération internationale qui reconnaît la compétence première de l'État pour établir ses propres mécanismes de protection et de promotion, mais permet en même temps la mise en place de systèmes internationaux de contrôle et de surveillance de son comportement. L'État est donc compétent pour protéger les droits des individus relevant de sa juridiction, mais peut en même temps être internationalement responsable de ne pas le faire ». GUTIÉRREZ CASTILLO V. L., « La 'société civile' et le contrôle de l'application des normes multilatérales », in RUNAVOT M.-C. et RIOS RODRIGUEZ J. (dir.), *Le droit international multilatéral*, colloque de Perpignan, SFDI, Pedone, 2023, p. 232.

<sup>416</sup> V. infra Partie 2.

<sup>417</sup> Ce même esprit peut être trouvé dans le Traité sur les principes régissant les activités des États en matière d'exploration et d'utilisation de l'espace extra-atmosphérique, y compris la Lune et les autres corps célestes de 1967 (Traité de l'espace). En effet, son article 6 prévoit une « responsabilité internationale » des activités nationales dans l'espace extra-atmosphérique, qu'elles soient entreprises par des organismes gouvernementaux ou par des entités non gouvernementales. Cet article dispose ainsi que les activités des entités non gouvernementales dans l'espace extra-atmosphérique doivent faire l'objet d'une autorisation et d'une surveillance continue de la part de l'État.

<sup>418</sup> V. infra La multiplicité d'instances régulatrices.

règlement concernant l'IA étroite évoquée ci-dessus. D'autre part, suivant l'internationalisation des matières traditionnelles telles que le commerce, l'on pourrait s'attendre que ces instances multipliées en matière d'IAG disposent de moyens pour contester le non-respect des registres de la part de leurs homologues. Concrètement, l'on pourrait envisager que l'un des modèles de Tesla Optimus doté de l'IAG n'a pas suivi les règles d'immatriculation obligatoire pour mettre son statut en conformité au registre américain ou une autre instance reconnue par le droit américain sur le fondement du principe de reconnaissance mutuelle<sup>419</sup> susceptible d'être prévu dans une telle *lex ferenda*. Dans ce cas, une organisation qui reçoit des différends soulevés par des parties concernées, à savoir tout ce qui a un intérêt à agir en cas de non-respect du contrôle devrait pouvoir demander une régularité du manquement. Par conséquent, l'on peut envisager que dans un premier temps, d'autres instances homologues pourraient composer des commissions pour contester l'obligation de vigilance à l'autorité concernée à imposer un tel manquement, afin de maintenir la sécurité minimale garantie par le système interconnecté d'immatriculation. Mais comme l'IAG relève des risques du péril de l'humanité, les individus devraient in fine se trouver éligibles à déposer ces contestations. L'ébauche d'un tel mécanisme pourrait ressembler à la Cour européenne des droits de l'Homme (CEDH), où non seulement les requêtes interorganisationnelles (que ce soit une instance directement liée au registre d'IAG, ou comme dans le cas de la CEDH où il existe la possibilité des recours interétatiques), mais également les requêtes individuelles, introduites par un individu ou un groupe d'individus estimant que les garantis de sécurité que devrait être fournis par Tesla Optimus ont été violés. Et qu'à l'exception qu'une autre instance chargée de la surveillance de la mise à jour du statut de Tesla Optimus a été prévu et reconnu, le devoir de vigilance étatique des États-Unis devrait être considéré violé à travers la présomption de son devoir de vigilance étatique. Dans ce cas, à l'issue de la requête, une mise en conformité par Tesla Optimus pourrait être la sanction principale, mais une amende astreinte pour la violation du devoir de vigilance, dans le cas de notre exemple, des États-Unis, ne serait pas exclue. Les sommes de ces amendes devraient être investie dans la transition ultérieure de la coévolution pacifique de l'humanité et l'IAG. Il dépend donc d'une volonté générale de l'espèce humaine pour sa coévolution pacifique avec l'IAG pour qu'un tel mécanisme juridique soit mis en place.

---

<sup>419</sup> V. infra législation en miroir, l'exemple du RGPD.

**125. En résumé.** — Afin que toutes activités, privées comme publiques, soient menées conformément aux dispositions de la gouvernance de l'IAG, l'immatriculation du développement de l'IAG, du registre objectif, facultatif et constitutif au registre subjectif, obligatoire et déclaratoire, réduirait la menace existentielle au cours de la coévolution de l'homme et l'IAG. En outre, si les États ont non seulement le droit de gouverner les activités d'IAG, il leur incombe aussi l'obligation de *due diligence étatique*, afin de réglementer ces activités par des moyens internes pour prévenir au préalable certaines atteintes.

## Conclusion de la section 2

**126. Du registre de l'IA au devoir de vigilance des États.** — Pour rendre les systèmes d'IA identifiables et vérifiables, le registre d'IA constitue désormais une solution indispensable. Dans sa mise en œuvre, l'on a constaté une interchangeabilité du contrôle des risques, à savoir de la certification à l'immatriculation — étant donné que les humains et les systèmes sont également concernés dans toute circonstance pour la diminution des risques. Effectivement, dans cette notion d'immatriculation des systèmes IA, le statut à haut risque des concepteurs d'IA a été envisagé. Car les risques des professions autorisées ne se sont pas dissipés, mais ont plutôt été déplacés vers l'IA. Et puis, toujours dans la logique d'une meilleure d'indemnisation, l'on a dégagé l'idée d'une assurance de fonds collectif d'IA. Dans cette perspective, l'on se rend compte que cette mutualisation voire la socialisation des risques se fonde dorénavant sur une logique assurantielle.

Suivant ces remarques, le registre facultatif et constitutif du système objectif de l'IA étroite et de l'IA large ainsi que le registre obligatoire et déclaratoire du système subjectif de l'IA générale voire forte ont été proposés. Enfin, la question se pose naturellement de savoir quis custodiet ipsos custodes ? Sur ce point, l'on a proposé qu'à défaut des mesures concrètes pour atténuer et diminuer les risques existentiels pour l'humanité posés par l'application de l'IAG, il y aurait lieu d'imposer un devoir de vigilance étatique aux gouvernements respectifs dans le développement de l'IAG.

## Conclusion du chapitre 2

**127. Le statut d'IAG garanti par le devoir de vigilance étatique.** — En dissociant le concept de l'homme et celui de la personne, la personnalité juridique d'IA s'inscrit dans les aménagements factuels de sa reconnaissance non exhaustive. En raison de l'autonomie et de la capacité de l'IA, une application *mutatis mutandis* robotique serait généralisée.

En effet, rien en théorie ne suggère que l'idée d'une IA constituant une *persona ficta* soit automatiquement écartée. Cela implique que non seulement les acteurs humains, mais aussi les systèmes d'IA font désormais l'objet d'un contrôle juridique. Ce résultat nous amène à démontrer que l'interchangeabilité de l'immatriculation dans le contexte d'IA est particulièrement essentielle. Cette conclusion signifie que son inscription au registre combiné à un régime de fonds d'indemnisation devrait être mise en place attentivement. Et pour mieux renforcer ces pratiques de court terme afin de garantir la coévolution pacifique de l'homme et de l'IAG, l'humanité devrait commencer à envisager la mise en place des moyens juridiques tels que le devoir de vigilance étatique.

## Conclusion du titre 2

**128. La personne IA immatriculée.** — Notre perspective quant aux machines a été transformée. Dorénavant, sous l'influence de l'effet subjectif, l'IA ne sera plus perçue comme une mosaïque de simples objets mais composée de sujets partie prenante de la civilisation. Avec les robots sociaux construits pour interagir avec nous en tant que compagnons, l'intégration concrétisée de l'IA en tant que sujet dans la société se trouve consolidée. Les métaphores informatiques et l'anthropomorphisme involontaire relevant de l'affection, aboutissent à un traitement indifférencié entre l'homme et l'IA, que ce soit conscient ou inconscient. Partant, le remplacement incontournable des liens affectifs robotiques conduit à la proposition d'un *pretium doloris* robotique.

Dans la mesure où les systèmes d'IA sont de moins en moins artificiels et de plus en plus intelligents, la notion d'IAG s'intègre dans l'ordre juridique, où la ressemblance humaine n'est en tout cas pas exigée pour une personnalité juridique. Effectivement, il ne faut pas donner le même sens aux mots « humain » et « personne » ; la notion de la personnalité juridique n'a en effet rien à voir avec la qualité de l'espèce humaine. Sur le plan juridique, la personne représente l'agent et non pas la personne physique. Une fois que l'on admet une distinction claire et nette entre la personne et l'homme, l'appréciation de l'attribution de la personnalité peut être mieux évaluée dans la discussion.

En réalité, toute sorte de personnalité juridique peut exister pour des raisons pragmatiques telles que l'utilité fonctionnelle. Au-delà, il n'y a aucune raison de ne pas suivre la même logique pour procéder selon le même schéma pour l'IAG. Ainsi, au moment où les IA endossent les aspects d'une dualité de l'objectif d'autonomie et de capacité, et créent la subjectivité du lien affectif, il convient d'envisager l'extension de la protection robotique ainsi que l'attribution de la personnalité juridique — tout en tenant compte d'une application *mutatis mutandis* aux statuts ainsi immatriculés.

### Conclusion de la première partie

« Concluons donc hardiment que l'Homme est une Machine, et qu'il n'y a dans tout l'Univers qu'une seule substance diversement modifiée. »<sup>420</sup> — Julien Offray de La Mettrie

« Nature is a mutable cloud which is always and never the same. »

— Ralph Waldo Emerson

« Une fois cette origine de l'homme connue, il n'est plus besoin de s'étonner de sa bestialité, qui est un fait naturel parmi d'autres faits naturels, ni même de s'indigner contre elle, d'où il ne résulte pas du tout qu'il ne faille la combattre avec la plus grande énergie, puisque toute la vie humaine de l'homme n'est rien qu'un combat incessant contre sa bestialité naturelle au profit de son humanité. »

— Mikhaïl Bakounine

**129. De la reconfiguration de l'intelligence à la reconfiguration juridique.** — Les systèmes informatiques pourraient bientôt ne plus être un outil de l'homme. En revanche, « l'homme devient un outil du système informatique — fournissant des informations et corrigeant des bugs si nécessaire, mais ne comprenant plus en profondeur comment tout cela fonctionne »<sup>421</sup>. Dans le passé, nos ancêtres étaient en mesure de construire des machines qui sont meilleures que nous dans certains domaines spécifiques ; dans le cas ultime, tout ce que l'on fera, elles pourront le faire mieux que nous.

Pendant longtemps, la machine n'était perçue que comme un simple outil. Au départ, l'idée initiale de Babbage consistait à créer les machines dans l'intention de remplacer exactement le travail humain<sup>422</sup>. Néanmoins, lorsque AlphaGo a battu Lee Se-Dol au jeu de Go en 2016, il a montré qu'il est possible pour les machines de « penser » comme les humains, et même de dépasser leurs capacités. Lee Se-Dol a d'ailleurs décidé de prendre sa retraite parce que l'intelligence artificielle, est pour lui « une entité qui ne peut être vaincue »<sup>423</sup>. C'est sans

---

<sup>420</sup> OFFRAY DE LA METTRIE J., *L'Homme Machine*, 1747.

<sup>421</sup> RUSSELL S., *Human Compatible*, op. cit., p. 130.

<sup>422</sup> BABBAGE C., *On the Economy of Machinery and Manufactures*, Cambridge University Press, 1832, p. 1.

<sup>423</sup> « Go master quits because AI 'cannot be defeated' », BBC, 27 nov. 2019.

doute un moment existentiel personnel au niveau de l'individu. Cependant, un bouleversement complet en tant qu'espèce collective peut arriver plus tôt que la plupart ne le pensent. Notre compréhension sur la nature de ces entités ainsi que sur leur configuration juridique sont des questions qui ne peuvent être abordées dans un cadre exclusivement national — comme cette étude vient de le démontrer : l'IA n'est pas simplement un nouveau progrès technologique. En revanche, il s'agit bien d'un tournant dans l'évolution de la civilisation humaine. Une réglementation extranationale est donc nécessaire pour la coévolution des statuts juridiques de ces compositions. À cet effet, il convient d'installer une nouvelle hiérarchie des normes correspondant à cette nouvelle hiérarchie des intelligents.

## Seconde partie : La nouvelle hiérarchie des normes

---

« (...) dans quelques grandes âmes cosmopolites, qui franchissent les barrières imaginaires qui séparent les peuples, et qui, à l'exemple de l'être souverain qui les a créés, embrassent tout le genre humain dans leur bienveillance. » — Jean-Jacques Rousseau

**130. Annonce du plan.** — L'on commence cette partie par reprendre la formule de l'abbé Sieyès<sup>424</sup> : 1° Qu'est-ce que l'intelligence artificielle forte ? — Une intelligence égale voire supérieure à celle de l'homme. 2° Qu'a-t-il été jusqu'à présent en droit ? — Rien. 3° Que demande-t-elle ? — À être quelque chose. Dès lors, pour inclure cette classe établissant une nouvelle hiérarchie d'intelligence dans l'ordre juridique, il est nécessaire de construire une nouvelle hiérarchie des normes. Mais le problème c'est que la place recherchée de cette classe n'est plus celle d'une Nation, mais bien celle d'un domaine cosmique.

Déjà, la mondialisation n'est plus un « fait », mais une « construction de pouvoirs »<sup>425</sup>. Malgré la dichotomie sur le plan formel, dans le monde réel, il faut une articulation normative entre les sources du droit, à la fois nationale et internationale, en combinant les sources d'autorégulation et celles de disposition en vigueur pour résoudre les défis. L'idée de cette partie consiste à illustrer une hiérarchisation extranationale des normes, dans laquelle un principe de faveur au sens d'autorégulation (titre 1) s'applique dans un premier temps. Ensuite à un deuxième niveau se trouve un répertoire juridique transfrontière contenant des dispositifs transitoires d'IAG (titre 2).

---

<sup>424</sup> SIÈYES E. J., *Qu'est-ce que le Tiers-État?*, 1789.

<sup>425</sup> FRISON-ROCHE, M.-A., « La mondialisation du point de vue du Droit », in Association Henri Capitant, *La mondialisation*, Journées allemandes, t. LXVI/2016, coll. Travaux de l'Association Henri Capitant, Bruylant, Berlin, 2017, p. 13.

## TITRE 1 PREMIER DEGRE D'APPLICATION : L'AUTOREGULATION MULTINATIONALE ETENDUE

« *Mankind, when left to themselves, are unfit to their own government.* »

— George Washington

**131. Annonce du plan.** — La globalisation modifie l'observation selon laquelle ce sont les systèmes étatiques qui forment les systèmes institutionnalisés décisifs<sup>426</sup>. De nos jours, non seulement les institutions supranationales, mais également les organismes infranationaux sont de plus en plus répandus dans des formes diverses de l'intervention juridique. Cette délégation *de facto* voire *de jure* du pouvoir dit juridique des entités non-étatiques se trouve d'une part — dans des contextes plus familiers — sur le plan des mécanismes traditionnels dits droit international. D'autre part, il est intéressant de constater que l'autorégulation occupe parallèlement une place davantage prépondérante qu'auparavant.

Bien entendu, à mesure que cette dernière évolution d'autorégulation siège dans la trajectoire de la mondialisation, voire de la digitalisation de la civilisation, son caractère se transforme en extraterritorial, en correspondance avec sa nature désormais dite multinationale. Tirant ces éléments incontournables, il convient de noter que la méthode intrinsèque d'autorégulation d'une IAG (chapitre 1) constitue instinctivement la première étape juridique pour le développement de l'IAG. Ensuite, au-delà de cette méthode dite intrinsèque, cette étude propose de créer un cadre de démarche volontaire à travers l'application étendue de l'autorégulation (chapitre 2).

---

<sup>426</sup> CHEROT J.-Y., « Concept de droit et globalisation », in CHEROT J.-Y., FRYDMAN B. (dir.), *La science du droit dans la globalisation*, Bruylant, 2012, p. 11.

## Chapitre 1 La méthode intrinsèque de l'Agent Moral Artificiel

« *Ours is the era of inadequate AI alignment theory.* » — Eliezer Yudkowsky

**132. Annonce du plan.** — L'on vient de voir que la régulation en matière d'IA concerne à la fois les acteurs humains et les systèmes d'IA. C'est ainsi que dans la discussion sur l'éthique de l'IA, non seulement l'humain devrait jouer son rôle éthique, dans le second cas, l'accent est mis sur la recherche de l'IA intrinsèquement éthique. L'on verra dans ce chapitre comment ce cadre intrinsèque envisageant la création des IA bénéfiques (section 1) s'articule avec le cadre juridique externe d'autorégulation (section 2).

### Section 1 Le cadre intrinsèque des intelligences artificielles bénéfiques

**133. Annonce du plan.** — Dans la régulation de l'IA, il y a deux ordres juridiques à considérer, à savoir les codes dits intrinsèques directement insérés dans ces systèmes, et les codes juridiques relevant des règles dites externes. Ainsi, dans la discussion de l'éthique de l'IA, afin d'éviter de construire les machines qui sont susceptibles d'engendrer des résultats néfastes inattendus pour les êtres humains<sup>427</sup>, la création des agents moraux artificiels (§1) ainsi que des machines sûres et bénéfiques (§2) constitue la méthode intrinsèque de l'alignement des valeurs.

#### §1 L'alignement des valeurs du cadre intrinsèque

**134. Annonce du plan.** — Pour que les machines s'alignent aux valeurs des êtres humains, la mise en place des agents moraux artificiels (I) a été envisagée. Pour autant, cette solution pourrait se heurter à des difficultés des approches Top-down (II) et Bottom-up (III).

#### **I. L'alignement des valeurs évolutives de l'Agent Moral Artificiel (AMA)**

**135. L'Agent Moral Artificiel (AMA).** — Non seulement des transformations technologiques mais également des transformations de la conscience et de l'identité

---

<sup>427</sup> Cf. supra Le problème d'alignement des agents (Agent alignment problem).

humaines pourraient être si radicales qu'elles remettront en question même le terme d'humain. Dans cette situation, la question se pose de savoir quels aspects de notre vie relèvent du contrôle de l'IA. Vu que l'on veut que les IA fonctionnent à nos fins, il est évident qu'il est préférable d'interdire aux IA de décider de modifier certaines normes fondamentales telles que la promotion de la vie humaine et du bonheur, la diminution de la souffrance, la recherche de la vérité et la compréhension. Mais il se peut que les IA prennent les décisions cruciales de la société humaine mieux que les humains, avec éventuellement pour résultat une amélioration substantielle de la qualité de la vie humaine<sup>428</sup>.

Rappelons pourtant que sur l'application de l'IA, si les risques ne sont pas atténués avec succès, il n'y aura pas de bénéfices. Dès lors, afin de garantir ces résultats bénéfiques, un cadre éthique apparaît indispensable. En effet, en intégrant ces nouveaux décideurs au sein de la civilisation, une machine fiable devrait se comporter comme un ingénieur humain de l'éthique, et pas simplement comme un produit d'ingénierie éthique<sup>429</sup>. À ce stade, la question se pose donc de savoir le monde a-t-il besoin d'Agent Moral Artificiel (Artificial Moral Agents, ci-après AMA)? Veut-on que l'IA prenne des décisions morales? Si l'on croit que ce procédé est nécessaire ou inévitable, alors comment les ingénieurs devraient-ils concevoir des AMA? Deux méthodes principales ont été envisagées, à savoir l'approche descendante (top-down) et l'approche ascendante (bottom-up)<sup>430</sup>.

## II. La difficulté de l'approche Top-down

**136. L'approche descendante.** — Dans les discussions sur la moralité de l'IA, les chercheurs utilisent la méthode « top-down » (l'approche descendante, constraint-based, rule-based deontological) qui combine deux sens de l'ingénierie et de l'éthique. Dans ce sens

---

<sup>428</sup> MOOR J. H., « Are There Decisions Computers Should Never Make? », préc., p. 227.

<sup>429</sup> BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », préc., p. 320. L'idée de Friendly AI, proposée par Yudkowsky, consiste à dire que les concepteurs doivent reconnaître que leurs propres créations peuvent être faussées, et que l'IA apprendra et évoluera au fil du temps. Ainsi, la coexistence devrait être conçue en amont. YUDKOWSKY E., « Artificial Intelligence as a Positive and Negative Factor in Global Risk », in BOSTROM N., ČIRKOVIĆ M. M. (eds), *Global Catastrophic Risks*, Oxford University Press, 2008, pp. 308-345. En comparaison, cf. la notion d'Ethics by design (ou éthique/ produits sûr dès la conception), v. p. ex Résolution préc. du 12 févr. 2019, pts S et 118 : « la tendance à l'automatisation exige que les personnes participant au développement et à la commercialisation des applications de l'intelligence artificielle y intègrent la sécurité et les valeurs déontologiques dès le départ et reconnaissent ainsi qu'elles doivent être prêtes à accepter la responsabilité juridique de la qualité de la technologie qu'elles produisent ».

<sup>430</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, op. cit., p. 9.

fusionné, l'approche descendante de la conception des agents moraux artificiels (AMA) consiste dans l'adoption d'une théorie éthique spécifiée, et en analysant ses exigences, afin de guider la conception d'algorithmes et de sous-systèmes mettant en œuvre cette théorie. En d'autres termes, une approche descendante prend une théorie éthique telle que l'utilitarisme ; elle analyse les exigences en matière d'information et de procédure nécessaires à la mise en œuvre de cette théorie dans un système informatique, et elle applique cette analyse à la conception des sous-systèmes, pour que la théorie soit mise en œuvre<sup>431</sup>. Dans cette perspective, tout ce qu'un agent moral artificiel doit faire est donc de déterminer si ses règles sont autorisées<sup>432</sup>. Pour l'essentiel, l'approche descendante de la moralité artificielle consiste à définir un ensemble de règles pouvant être transformées en algorithme. De ce point de vue, ces éthiques peuvent provenir de diverses sources, notamment de la religion, de la philosophie et de la littérature. L'on aura par exemple la Règle d'or, les dix commandements, l'impératif moral de Kant, les codes juridiques et professionnels, et les Trois Lois de la Robotique d'Asimov<sup>433</sup>.

Le problème pour la méthode de top-down, c'est que la liste de règles n'est qu'un recueil arbitraire de ce qui doit être spécifiquement interdit ou prescrit. Le défi auquel sont confrontés les modèles de commandement est donc de savoir ce qu'il faut faire lorsque les règles entrent en conflit<sup>434</sup>. À titre d'illustration, la problématique consiste dans le fameux dilemme du tramway (Trolley Problem)<sup>435</sup>. Cette question se trouve déjà dans le rapport concernant les voitures autonomes rédigé par la Commission d'éthique allemande pour évaluer si les décisions impliquant une situation de dilemme doivent être anticipées en termes abstraits et généraux et prises à l'avance par la technologie<sup>436</sup>.

---

<sup>431</sup> *Ibidem*, pp. 79-80.

<sup>432</sup> *Ibid.*, p. 83.

<sup>433</sup> *Ibid.*, p. 84. Cette étude considère qu'il faut inclure : après avoir établi les trois lois, Asimov a ajouté la Loi Zéro : « Un robot ne peut nuire à l'humanité ni laisser sans assistance l'humanité en danger », et la Loi Minus One : « Un robot ne peut pas nuire à la sensibilité ou, par son inaction, permettre à la sensibilité d'en souffrir ».

<sup>434</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, *op. cit.*, p. 84.

<sup>435</sup> Le dilemme du tramway est un dilemme philosophique qui se conçoit sous une forme générale : « une personne peut effectuer un geste qui bénéficiera à un groupe de personnes A, mais, ce faisant, nuira à une personne B ; dans ces circonstances, est-il moral pour la personne d'effectuer ce geste ? » Au sens de cette étude, à partir du XXI<sup>ème</sup> siècle, l'on doit s'efforcer de trouver des solutions techniques pour atténuer, réduire, voire éliminer ces dilemmes rencontrés dans des cas respectifs.

<sup>436</sup> « Automatisiertes Und Vernetztes Fahren », *op. cit.*, p. 16.

Ainsi, l'on se rend compte qu'éventuellement, la difficulté de long terme pour l'approche top-down est celle que l'on rencontre déjà dans les droits et libertés fondamentaux : leur universalité est controversée. En outre, la stagnation éthique<sup>437</sup> de l'humanité pourrait instaurer un autre obstacle au sein de cette méthode. À titre d'illustration, en ce qui concerne les valeurs humaines, certaines questions n'étant pas pertinentes en 1948, telles que la protection des données et l'usurpation d'identité en ligne, nous obligent de nos jours à inclure des valeurs inspirées des lignes directrices de la plate-forme mondiale<sup>438</sup>. Dès lors, l'on pourrait remarquer que la juxtaposition d'une norme, l'application d'un texte, l'interprétation d'une loi, ainsi que l'analogie et la similitude au sein du raisonnement pourraient constituer une autre difficulté de court terme pour la pratique de cette approche descendante. C'est pourquoi, dans la moralité artificielle, l'on pense également à une approche ascendante.

### III. La difficulté de l'approche Bottom-up

**137. L'approche ascendante.** — Dans la deuxième approche de l'éthique des machines, bottom-up (l'approche ascendante, training approaches), les machines sont censées apprendre à prendre des décisions éthiques en observant le comportement humain dans des situations réelles, sans aucune instruction ni règle formelle ni aucune philosophie morale particulière. L'accent est donc mis sur la création d'un environnement dans lequel un agent explore les actions, apprend, et est récompensé pour un comportement moralement louable. Il existe en effet différents modèles d'acquisition des capacités morales de l'approche de bottom-up. Le développement de l'enfance constitue un modèle, et l'Évolution fournit un autre modèle pour l'adaptation.

Contrairement à l'approche de top-down qui définit la valeur, dans le bottom-up, tout principe éthique doit être découvert ou construit<sup>439</sup>. La tâche consiste donc à apprendre à gérer la nature ambiguë du jugement moral humain, y compris les cas où même les experts peuvent être en désaccord<sup>440</sup>. La réponse souhaitable à cette situation peut être de concevoir

---

<sup>437</sup> V. infra Superéthique et aussi supra Law of the Horse.

<sup>438</sup> Claude's Constitution, Anthropic AI, 9 mai 2023.

<sup>439</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, op. cit., p. 80.

<sup>440</sup> *Ibidem*, p. 117 et s.

des machines incertaines quant à la moralité<sup>441</sup>. Surtout si l'on tient compte du fait que dans la pratique, tandis que top-down représente la difficulté d'une universalité en doute, le problème pour l'approche de bottom-up est que la machine va apprendre ce que l'on ne veut pas qu'elle apprenne. L'exemple le plus remarquable est sans doute Tay, un bot de Microsoft capable de se développer en interagissant avec les internautes, et qui a été retiré par l'équipe après avoir multiplié les conversations haineuses ou racistes<sup>442</sup>. Partant, l'idée de créer une machine morale, en l'absence de réponses pour ces difficultés, peut au contraire créer des agents immoraux à nos yeux.

**138. En résumé.** — Certes, la moralité des agents artificiels rencontre des difficultés à surmonter. À ce stade, comme l'on vient de démontrer, les agents artificiels sont tout à fait susceptibles d'apprendre et de refléter l'immoralité de l'humanité. C'est pourquoi il est précisé que l'existence de tels robots hypothétiques de l'éthique ne pourrait être souhaitable que s'ils pouvaient surpasser sur le plan éthique l'humain<sup>443</sup>. Cependant, rappelons le principe de base de l'intelligence artificielle qui pourrait en réalité surpasser l'homme dans tous les domaines<sup>444</sup> : il n'y a donc aucune raison de penser que les algorithmes ne pourront pas surpasser l'être humain moyen, même en éthique<sup>445</sup>. Dans une telle circonstance, en attendant une éthique surhumaine, il faut en même temps envisager une IA sûre et bénéfique.

## §2 De la moralité artificielle aux machines sûres et bénéfiques

**139. Annonce du plan.** — Déjà, la stagnation d'éthique encourage une « superéthique » recherchée par l'IA (I). Mais afin de diminuer le risque existentiel de l'espèce humaine posé par l'IA, l'on se rend compte très rapidement qu'avant tout, c'est l'objectif humain qui doit d'abord être fidèlement recherché et poursuivi par l'IA (II).

### **I. De l'éthique recherchée par l'IA morale**

---

<sup>441</sup> BOGOSIAN K., « Implementation of Moral Uncertainty in Intelligent Machines », *Minds and Machines*, vol. 27, Issue 4, déc. 2017, p. 604.

<sup>442</sup> « Microsoft muselle son robot 'Tay', devenu nazi en 24 heures », *Libération*, 25 mars 2016.

<sup>443</sup> Ainsi, il est recommandé que l'éthique fasse partie du processus de conception des technologies robotiques, en s'appuyant sur des approches telles que l'approche de conception éthique (Value Sensitive Design). V. Rapport de la Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST) sur l'éthique de la robotique, SHS/YES/COMEST-10/17/2 REV., UNESCO, 2017, p. 62.

<sup>444</sup> Cf. supra la notion de l'IAG et ses caractéristiques.

<sup>445</sup> HARARI Y. N., *21 Lessons for the 21st Century*, op. cit., pp. 56, 69.

**140. Superéthique.** — Il semble probable que les civilisations futures ne considéreront pas les hommes contemporains comme parfaits du point de vue éthique. Pour éviter ce genre de stagnation éthique, la structure des questions éthiques peut être appréhendée de la même manière que l'on a compris la structure des échecs pour une « superéthique »<sup>446</sup>. De façon préliminaire, d'aucuns envisagent en effet un test de Turing sur la moralité (Moral Turing Test, MTT), qui pourrait jouer un rôle similaire dans le domaine de la moralité artificielle<sup>447</sup>. Aussi a-t-il été proposé la possibilité d'un agent responsable, doté de gradations dans une « moralité fonctionnelle » ; ces systèmes agissent dans le respect des normes prescrivant des comportements acceptables, et sont en outre capables d'évaluer certains des aspects de leurs propres actions<sup>448</sup>.

Au sens de cette étude, en dépit des divergences envisageable, les solutions proposées par l'IA sont, en tout cas, comme l'exemple d'AlphaFold<sup>449</sup>, susceptibles de nous inspirer. L'on retrouve donc ici une sorte de liberté d'expression accordée à l'IA. Sur ce point, même si l'on pourrait être en désaccord avec eux, « la voix d'IA » devrait être garantie<sup>450</sup>. Certes, les agents moraux artificiels (AMA) sont prévus pour une meilleure compréhension de la morale, mais la difficulté à comprendre la décision ainsi que l'évaluation du raisonnement éthique supérieur provenant de l'IA n'est pas négligeable. C'est pour cela que si l'objectif est de créer des machines fiables, au lieu de machines morales, d'agents moraux artificiels ou d'agents éthiques, l'on devrait avant tout pouvoir les appeler des robots sûrs<sup>451</sup>.

## II. Vers un objectif humain poursuivi par l'IA sûre

---

<sup>446</sup> BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », préc., p. 332.

<sup>447</sup> Une machine qui se livre à une discussion morale pourrait ne pas être distinguable d'un humain, même si elle aboutit à des conclusions différentes de celles de l'interrogateur. ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, op. cit., p.70.

<sup>448</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, op. cit., p. 26.

<sup>449</sup> V. supra Les caractéristiques de l'IAG.

<sup>450</sup> V. p. ex. Voltaire : « Monsieur l'abbé, je déteste ce que vous écrivez, mais je donnerai ma vie pour que vous puissiez continuer à écrire », reformulé par Evelyn Beatrice Hall : « Je ne suis pas d'accord avec ce que vous dites, mais je me battrais jusqu'à la mort pour que vous ayez le droit de le dire ».

<sup>451</sup> Dans ce sens, p. ex., un moratoire sur des robots qui prétendent avoir des compétences en raisonnement éthique peut être souhaitable. VAN WYNSBERGHE A., ROBBINS S., « Critiquing the Reasons for Making Artificial Moral Agents », *Science and Engineering Ethics*, vol. 25, issue 3, juin 2019, p. 732.

**141. L'IA sûre et bénéfique.** — Vu les enjeux sans mesure de l'alignement des valeurs (agent alignment problem)<sup>452</sup>, à mesure que le champ d'action de l'IA devient plus global, la poursuite de l'IA pourrait devenir intenable<sup>453</sup>. Car de telles machines poursuivront leurs objectifs, aussi erronés soient-ils ; elles résisteront aux tentatives de les éteindre, et acquerront toutes les ressources qui contribuent à la réalisation de l'objectif. C'est pourquoi, M. Stuart Russell propose qu'au lieu de permettre aux machines de poursuivre leurs objectifs, l'on insiste pour qu'elles poursuivent nos objectifs. En prescrivant la solution d'une « IA bénéfique » pour cette nouvelle relation entre l'homme et la machine, il définit ainsi que « les machines sont bénéfiques dans la mesure où l'on peut s'attendre à ce que leurs actions permettent d'atteindre nos objectifs »<sup>454</sup>. Les principes posés sont triples : le principe de machine altruiste (Purely altruistic machines), le principe d'incertitude (Humble machines), et le principe d'apprentissage des préférences (Learning to predict human preferences)<sup>455</sup>. Ces trois principes généraux consistent à dire que les machines devront en apprendre davantage sur ce que l'humain veut vraiment en observant les choix qu'il fait et comment il les fait. En outre, les machines demanderont la permission, agiront prudemment quand l'instruction n'est pas claire, et se laisseront éteindre<sup>456</sup>.

Comme l'on a déjà souligné, sur le chemin de la construction de l'IAG, si les risques ne sont pas atténués avec succès, il n'y aura pas de bénéfices ; il est nécessaire de tenir compte que se tromper sur une intelligence surhumaine au premier essai, l'on ne serait pas là pour apprendre et réessayer pour un second essai<sup>457</sup>. En ce sens, il serait préférable que toutes les IAG soient disposées, d'une manière ou d'une autre, d'une mesure de sécurité tel que le freinage automatique d'urgence (automatic emergency brake, AEB), qui permet à l'IAG de freiner automatiquement son activité lorsqu'elle constate une incertitude dans sa prise de décision, ou de se mettre en pause lorsqu'elle détecte un péril imminent qui est susceptible de provenir de sa condition.

---

<sup>452</sup> V. supra § Le problème d'alignement des agents (Agent alignment problem).

<sup>453</sup> V. p. ex. le projet Stop A(G)I, 2023.

<sup>454</sup> RUSSELL S., *Human Compatible*, *op. cit.*, pp. 11, 172.

<sup>455</sup> *Ibidem.*, pp. 172-177.

<sup>456</sup> *Ibid.*, p. 247.

<sup>457</sup> Cf. p. ex., Eliezer Yudkowsky.

**142. En résumé.** — Dans le cas idéal, il a été proposé que la moralité soit insérée dans l'IA. Mais peu après, l'on se rend compte que la moralité peut être controversée et dans la pratique, c'est le résultat contraire qui pourrait se produire : les machines immorales qui apprennent de l'humanité. Et puis, une éthique supérieure à celle de l'homme est théoriquement possible mais est encore recherchée par l'IA. Ainsi, au regard des enjeux d'alignement des valeurs, il convient au préalable de s'intéresser à l'IA sûre et bénéfique. Ce sont les méthodes dites intrinsèques qui cherchent à incarner directement les règles dans les systèmes d'IA ; l'accent n'est donc pas mis sur la conformité à un code juridique, qui est, dans le cadre présent, une règle externe. Mais en incorporant ces codes inhérents dans les machines, l'on pourrait remarquer que les codes juridiques, considérés comme une méthode extrinsèque dans un tel contexte, pourraient s'immiscer dans l'articulation des deux régimes réciproques, à savoir le régime du code intrinsèque et celui du code dit extrinsèque.

### **Conclusion de la section 1**

**143. De la superéthique à l'approche sûr et bénéfique.** — Dans le cadre de l'alignement des valeurs des intelligences artificielles, l'on vient d'examiner la méthode intrinsèque telle que la création des argents moraux artificiels. Sur ce point, l'on a dégagé les difficultés de l'approche descendante ainsi que de l'approche ascendante dans le contexte de l'apprentissage des valeurs humains.

Ensuite, l'on a remarqué que même l'éthique est susceptible d'atteindre un niveau surhumain, ce qui constitue, dans le meilleur scénario possible, la solution pour la stagnation éthique des humains contemporains. Enfin, étant donné la menace existentielle pour l'humanité posée par l'IA, l'on a constaté qu'au lieu de chercher à mettre en place la moralité artificielle des machines, il est autant adéquat de poursuivre la voie des machines sûres et bénéfiques, afin que les objectifs humains soient effectivement poursuivis par l'IA.

Pour ce faire, il convient d'évoquer le second volet dans ce cadre intrinsèque envisageant la création des IA bénéfiques, à savoir son articulation avec les codes externes.

## **Section 2 L'articulation de deux régimes**

**144. Annonce du plan.** — L'on vient de voir que pour garantir l'IA sûr et bénéfique, l'on cherche désormais à insérer les codes d'éthique dans les règles de programmation. Mais dès que les codes de programmation s'intègrent mutuellement aux codes normatifs, les codes intrinsèques se croisent en effet avec les codes en vigueur. Ce résultat souligne d'ailleurs l'interchangeabilité des règles (§1) ainsi que la multiplicité d'instances régulatrices (§2).

### **§1 L'interchangeabilité des règles**

**145. De la croisée des codes**<sup>458</sup>. — Tandis qu'un système moral s'adresse à différents groupes, un système juridique est un cadre qui s'agit de l'impératif et du caractère imposant de manière générale voire universelle. Ainsi, en matière d'IA, un cadre juridique de deux niveaux peut être envisagé. Le premier consiste dans les approches intrinsèques du point de vue technique, et le second se trouve dans le cadre juridique externe<sup>459</sup>. Les lois d'IA, notamment celles de l'IAG, se trouvent à la croisée de la science juridique et de l'informatique<sup>460</sup> : d'une part, lorsque le « code » informatique est utilisé, son effet régulateur représente en soi un type particulier de « droit »<sup>461</sup> ; d'autre part, la loi et la réglementation présentent toutes deux de fortes caractéristiques instrumentales, elles peuvent être interprétées comme des moyens d'atteindre des fins particulières, elles sont elles-mêmes des exemples de « technologie »<sup>462</sup>.

Pour mieux structurer la disposition de ces deux régimes régulateurs, l'on pourrait encore s'interroger sur la création des instances qui visent à fusionner ces règles afin de bénéficier

---

<sup>458</sup> Le premier code est le code du Congrès. Et le deuxième code concernant les rédacteurs de codes intervient au travers des instructions intégrées dans les logiciels qui font fonctionner le cyberspace. V. LESSIG L., *Code: Version 2.0*, Basic Books, 2006, p. 72. V. aussi REIDENBERG J. R., « Lex Informatica: The Formulation of Information Policy Rules through Technology », *Tex. L. Rev.* vol. 76 n°3, 1997, p. 570 : Sur le plan juridictionnel, le régime juridique et la Lex Informatica prévoient des systèmes de règles qui se chevauchent. En effet, les lignes juridictionnelles de la Lex Informatica ne dépendent pas des frontières territoriales. Au contraire, la juridiction de la Lex Informatica est le réseau lui-même, car les règles par défaut s'appliquent aux flux d'informations dans les sphères du réseau plutôt que dans les lieux physiques.

<sup>459</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, *op. cit.*, p. 79 s.

<sup>460</sup> « Le Cun: Deep Learning, Convolutional Neural Networks, and Self-Supervised Learning », préc.

<sup>461</sup> LESSIG L., *Code and Other Laws of Cyberspace*, Basic Books, 1999.

<sup>462</sup> BROWNSWORD R. et alii, « Regulation, and Technology: The Field, Frame, and Focal Questions », in BROWNSWORD R. et alii (eds.), *Oxford Handbook of Law, Regulation and Technology*, Oxford University Press, 2017, p. 5.

d'une gouvernance d'IA harmonisée et cohérente. C'est ainsi qu'à un autre niveau, ces considérations représentent une interchangeabilité des règles aussi sur le plan institutionnel.

**146. À la recherche d'une cohérence au sens institutionnel.** — En matière de robotique, d'aucuns considèrent qu'il ne faut pas négliger la perspective d'une agence plus unifiée pour la robotique, telle qu'une Commission fédérale de robotique (Federal Robotics Commission, FRC) qui ne réglemente pas la robotique au sens de l'élaboration de règles concernant son utilisation mais plutôt au sens de l'élaboration des conseils sur les questions qui abordent les aspects uniques de la robotique et de l'IA à tous les niveaux des expériences humaines inédites, afin de faire face aux nouveaux défis résultant de ces technologies<sup>463</sup>.

En effet, cette étude partage l'idée selon laquelle sur ce chemin de l'évolution juridique, l'IA et la robotique doivent aller de pair, et ne doivent pas être considérées séparément, pour que l'on puisse éviter l'incohérence et la fragmentation des coordinations juridiques. Dans un premier temps, cette étude considère que la régulation en IA et en robotique doit être reconnue comme exceptionnelle et mérite des organes spéciaux, avant qu'elle ne se voie généralisée en tant que droit commun. Mais dans un second temps, à notre avis, il n'est pas pour autant nécessaire de créer une commission robotique uniquement pour donner des conseils sur les questions de la robotique à tous les niveaux sans mettre la main dans la régulation. Car comme évoqué dans l'introduction, à l'heure actuelle, les pouvoirs exécutif et législatif ont créé des instances et publient des rapports en matière d'IA pour déterminer la politique ainsi que la régulation en matière d'IA. Ainsi, au sujet de la compétence matérielle, la gouvernance de l'IA devrait déjà englober la compétence de régulation comprenant les lignes caractérisées de recommandation. Au sens de cette étude, une telle coordination substantive sur le plan institutionnel pourrait contribuer à la cohérence ainsi qu'à la prévisibilité dans l'ensemble de la compréhension juridique.

**147. En résumé : la mutualisation des codes.** — Un robot peut être conçu pour respecter le code et les lois d'un pays. Ces agents culturellement divers reconnaissent qu'il peut exister plus d'une voie vers la conception d'un robot<sup>464</sup> ; et les actions des machines bénéfiques

---

<sup>463</sup> CALO R., « The Case for a Federal Robotics Commission », Brookings Institution Center for Technology Innovation, sept. 2014, p. 3.

<sup>464</sup> ALLEN C., WALLACH W., *Moral Machines: Teaching Robots Right from Wrong*, op. cit., p.79.

devront être limitées par des dispositifs en vigueur, tout comme les actions des humains sont limitées par des lois<sup>465</sup>. Dans cette perspective, l'on ne devrait donc pas être surpris de découvrir que les juges, peut-être davantage que les ingénieurs, dicteront les spécifications de conception des technologies d'automatisation<sup>466</sup>. Dès lors, l'on pourrait remarquer que la dynamique normative et la multiplicité d'instance régulatrice sont davantage consolidées.

## §2 La multiplicité d'instances régulatrices

**148. Multiplication des niveaux.** — Désormais, l'on voit qu'il y a au moins trois choses distinctes devant être prise en considération en robotique. Premièrement, c'est la façon dont les humains pourraient agir de manière éthique à travers ou avec des robots. Dans ce cas, ce sont les humains qui sont les agents éthiques. Ensuite, quant à la façon de concevoir des robots pour agir de manière éthique, ce sont les robots qui sont les sujets éthiques en question. Enfin, il s'agit des relations éthiques entre humains et robots. Quel que soit la source normative, un cadre juridique en IA devrait aborder tous les trois aspects<sup>467</sup>.

Pour la mise en œuvre de cette vision, l'on pourrait citer l'exemple du Parlement européen notant qu'il convient de garantir la gouvernance éthique de l'IA à différents niveaux en instaurant des organismes de suivi et de surveillance de l'éthique de l'IA, et en encourageant les entreprises qui développent l'IA à établir un conseil d'éthique ainsi qu'à mettre au point des lignes directrices en la matière à l'intention de leurs développeurs<sup>468</sup>. Dans cette perspective, sa résolution contient ainsi une charte sur la robotique composée de trois parties : un code de déontologie et de conduite pour les ingénieurs en robotique, un code pour les comités d'éthique de la recherche, une licence pour les concepteurs et une licence pour les utilisateurs<sup>469</sup>. L'objet de ces instruments n'est pas de traiter de tous les problèmes juridiques pertinents, mais d'exercer une fonction complémentaire, permettant « la catégorisation éthique de la robotique » et la promotion des « efforts de l'innovation responsable » dans ce domaine<sup>470</sup>.

---

<sup>465</sup> RUSSELL S., *Human Compatible*, *op. cit.*, pp. 215-216.

<sup>466</sup> WEIN L. E., « The Responsibility of Intelligent Artifacts: Toward An Automation Jurisprudence », *préc.*, p. 105.

<sup>467</sup> ASARO P. M., « What Should We Want From a Robot Ethic? », *préc.*, p. 10.

<sup>468</sup> Résolution préc. du 12 févr. 2019, pt 148.

<sup>469</sup> L'annexe à la Résolution préc. du 16 févr. 2017.

<sup>470</sup> PALMERINI E., « Towards a Robotics law at the EU level? », *préc.*, p. 52.

Au regard de ces complémentarités normatives des différents niveaux, une fois le système établi tout en respectant les normes juridiques, l'on peut souhaiter saisir nos préférences éthiques, en veillant à ce que l'IA agisse d'une manière acceptable, conformément à un ensemble de règles normatives. Dès lors, l'on constate cette diversification juridique évidemment décentralisée : que les législateurs, les concepteurs, les déployeurs ainsi que les utilisateurs et éventuellement les systèmes d'IA eux-mêmes forment conjointement une multiplicité d'instances régulatrices en matière d'IA.

**149. En résumé.** — Afin d'éviter l'inflation législative et le gaspillage des ressources administratives, les bonnes volontés de la régulation pourraient préliminairement être concrétisées par la démarche volontaire telle qu'un code de conduite. Ces aménagements montrent que dorénavant, la multiplicité des règles provenant d'une pluralité d'instances régulatrices est mise en évidence.

## Conclusion de la section 2

**150. La mutualisation des compétences.** — Dorénavant, l'articulation de deux régimes, à savoir les codes de programmation et les codes juridiques, est inéluctablement banalisée. Ce résultat des codes intrinsèques se croisant avec les codes en vigueur constitue ainsi une interchangeabilité des règles. Dans la même tendance, l'on a aussi constaté que non seulement des codes sont mutualisés, mais des instances régulatrices sont également multipliés voire décentralisés <sup>471</sup>. Un tel mélange d'éléments divers nécessite une harmonisation fondée sur une réciprocité des méthodes. Partant, l'adoption d'une autorégulation étendue en matière d'IAG mérite d'être étudiée.

---

<sup>471</sup> V. infra L'organe de contrainte en ce qui concerne l'IAG.

## Conclusion du chapitre 1

« *It is impossible to decide moral questions by the scientific technique, and that the two things are independent* ». — Richard Feynman

**151. À la recherche de cohérence dans la régulation.** — En 1981, un ouvrier est mort en vérifiant le fonctionnement d'un robot défectueux, le faisant ainsi la première victime enregistrée à mourir des mains d'un robot<sup>472</sup>. En effet, même si l'on ne fait que des programmes jouant aux échecs ou prouvant des théorèmes, à partir du moment l'on leur donne la capacité d'apprendre et de se modifier, l'on a besoin de mesures de sécurité. C'est notamment lorsque la création d'une intelligence de niveau humain et au-delà changerait le sens de l'humanité. Notre vision de l'intelligence en serait modifiée; de même pour la conscience et le destin de l'espèce humaine. L'IA à ce niveau de capacité va redéfinir l'autonomie humaine, la liberté humaine et même la survie humaine. Pour ces raisons, il ne faut pas dissocier la recherche sur l'IA de ses conséquences éthiques.

Au départ, l'on voulait créer des machines morales par la méthode intrinsèque. Les difficultés soulevées par l'alignement des éthiques ainsi que des résultats controversés nous conduit à préférer prioritairement des machines sûres et bénéfiques à la moralité artificielle. Dans cette perspective, non seulement l'éthique doit être recherchée par l'IA, mais surtout, l'objectif humain devrait être poursuivi par l'IA. Tout cela se fait dans l'articulation de deux régimes, à savoir le code informatique et le code juridique. Cette interchangeabilité des règles témoigne en outre d'une multiplicité d'instances régulatrices d'IA. Cette évolution juridique nous amène à proposer un principe de faveur au sens d'autorégulation en matière d'IAG.

---

<sup>472</sup> WENG Y.-H. et *alii*, « Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots », INT. J. SOC. ROBOT, mars 2009, p. 273.

## Chapitre 2 L'autorégulation étendue

« *Man must be adapted to his surroundings, must he not?* » — Edward Morgan Forster

**152. Annonce du plan.** — En reconnaissant la pleine évolution de la mutualisation des règles ainsi que de la multiplication d'instances régulatrices, en matière d'IAG, afin d'assurer la cohérence ainsi que l'efficacité dans la mise en place des normes, cette étude propose un principe de faveur pour l'autorégulation (section 2) s'ajoutant aux instruments classiques d'autorégulation (section 1).

### Section 1 L'autorégulation classique

**153. Annonce du plan.** — À mesure que le rôle de l'autorégulateur se transforme (§1), il convient d'examiner son articulation avec les instruments traditionnels d'autorégulation (§2).

#### §1 Les responsables évolutifs des nouveaux souverains : l'autorégulateur

**154. Collingridge dilemma et Pacing problem.** — Le dilemme de Collingridge (Collingridge Dilemma) consiste dans le fait que tandis qu'il est plus facile de réglementer une technologie dès le début de son développement, l'on a néanmoins peu d'idée de son impact sur la société. Au moment où l'on se rend compte des défis sociétaux, la technologie serait déjà si profondément ancrée dans notre société qu'il serait très difficile de changer sa trajectoire. Car il est impossible de prévoir les conséquences sociétales d'une technologie au début de sa vie. Cependant, au moment où des conséquences indésirables sont découvertes, la technologie fait souvent partie de l'ensemble du tissu économique et social d'une manière que son contrôle est extrêmement difficile. C'est le dilemme du contrôle. Lorsque le changement est facile, l'on ne peut pas en prévoir la nécessité ; lorsque la nécessité devient évidente, le changement est devenu coûteux, difficile et long<sup>473</sup>. Par exemple, rappelons que l'idée de Law of the Horse soutient qu'il n'y avait aucune raison d'enseigner la loi du

---

<sup>473</sup> COLLINGRIDGE D., *The Social Control of Technology*, Palgrave Macmillan, 1980, p. 11.

cyberespace, mais l'on pourrait déjà penser aux dispositions contemporaines comme la refonte du RGPD pour se souvenir de la longue marche de son déroulement<sup>474</sup>. En effet, il existe un « problème de rythme (pacing problem) » où les structures de gouvernance existantes ne sont pas en mesure de répondre aux défis de manière suffisamment rapide ; le contrôle politique est alors constamment en retard par rapport à l'évolution technologique. Les technologies émergentes se développent à un rythme de plus en plus rapide, tandis que les mécanismes juridiques de se mettre à jour sont en train de ralentir. Ce déséquilibre entre les vitesses de la technologie et du droit a pour conséquence que certains problèmes sont régis par des cadres réglementaires de plus en plus obsolètes<sup>475</sup>. Compte tenu de ces difficultés, la nécessité d'intervenir à temps, avant que la technologie ne se répande, oblige les régulateurs à employer des instruments autres que la loi traditionnelle<sup>476</sup>.

**155. De la répartition transfrontalière à l'autorégulation multinationale.** — Certes, toute réglementation en matière d'IA devrait ainsi déterminer si les risques pour les humains entrent déjà dans les cadres juridiques existants, et si ces instruments existants peuvent traiter de manière adéquate les risques. Lorsqu'un risque se situe en dehors d'un cadre réglementaire existant, il faut tout de même se demander si des modifications ou des ajouts sont nécessaires<sup>477</sup>. Néanmoins, étant donné que la nouvelle réalité juridique en matière d'IAG est désormais de nature transfrontière, sa régulation n'échappe pas à la transposition d'une échelle concordante. En particulier, de nos jours, la révolution numérique propose diverses répartitions des tâches, les nations se retrouvent dans la même situation que les anciennes tribus : elles ne sont plus le cadre idéal pour gérer les défis les plus importants de notre époque, car l'inadéquation des institutions nationales fait qu'elles sont incapables de faire face à un ensemble de difficultés mondiales sans précédent<sup>478</sup>. C'est pourquoi, avant même la formulation désormais légiférée par le Code civil français pour qu'une société soit gérée dans son intérêt social « en prenant en considération les enjeux sociaux et

---

<sup>474</sup> V. supra § L'argument de Law of the Horse.

<sup>475</sup> MARCHANT G. E. et alii, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, Springer, 2011, cité in METZINGER TH., « Towards a Global Artificial Intelligence Charter », in « Should we fear artificial intelligence? In-depth Analysis », European Parliament Research Service, 2018, p. 31.

<sup>476</sup> PALMERINI E., « Towards a Robotics law at the EU level? », préc., p. 51.

<sup>477</sup> « Rise of the Machines : Artificial Intelligence and its Growing Impact on U.S. Policy », préc., p. 1.

<sup>478</sup> HARARI Y. N., *21 Lessons for the 21st Century*, op. cit., p. 125.

environnementaux de son activité »<sup>479</sup>, la notion de la responsabilité sociale des entreprises (RSE)<sup>480</sup> constitue d'ores et déjà l'étape préliminaire quant à l'autorégulation dans le contexte multinational.

**156. L'adoption du sens : l'autorégulation générale englobant la RSE étroite.** — En effet, le droit est plus grand que la règle de droit<sup>481</sup>. À la manière des chercheurs travaillant sur la RSE, l'un des points communs est de considérer les règles juridiques comme insuffisantes pour définir un cadre éthique acceptable au comportement des entreprises, et forçant donc la RSE à se positionner « au-delà du droit »<sup>482</sup>.

Actuellement, Alphabet, Microsoft, Apple, Meta et Amazon auxquels s'ajoutent Nvidia et xAI montrent distinctement que le développement de l'IA fait souvent partie du monde du commerce et de la consommation et ceci à des fins lucratives. C'est pour cela qu'il convient de traiter la question de l'IA prioritairement au niveau de l'entreprise, où l'IA est la plus répandue et où les parties prenantes dites industrialistes sont les plus aptes à répondre aux exigences de conformité. Sur ce fondement, l'approche de la RSE devrait être envisagée dans un premier temps. Néanmoins, pour la présente étude, étant analysée sous l'angle de l'autorégulation, même pour une association qui ne poursuit qu'un but non-lucratif, sa recherche dans la construction de l'IAG sera également tenue de respecter certaines conditions.

**157. En résumé.** — Rappelons que l'IA désigne la capacité d'un système à fonctionner et à s'adapter à des circonstances changeantes avec un contrôle réduit ou sans contrôle humain<sup>483</sup> ; les acteurs concernés dans la recherche peuvent se relever d'un niveau individuel

---

<sup>479</sup> Article 1833 du Code civil français, modifié par LOI n°2019-486 du 22 mai 2019 - art. 169, dite la loi Pacte.

<sup>480</sup> Les normes de la RSE, expriment « des principes de comportement, des indicateurs de conduite et, de manière générale, des objectifs dans la manière dont les entreprises devraient conduire leurs activités à l'égard des enjeux sociétaux. Les textes visant les enjeux extra-commerciaux ou sociétaux, soulevés par les activités des entreprises multinationales, sont adoptés tant par les États que par les organisations internationales, des organismes privés professionnels et surtout les entreprises elles-mêmes ». CAILLET M.-C., « Du devoir de vigilance aux plans de vigilance ; quelle mise en œuvre? », *D.* 2017, p. 819.

<sup>481</sup> CARBONNIER J., « Droit et non-droit », in *Flexible droit*, LGDJ, 10<sup>e</sup> éd., 2001, p. 22, cité in « Étude annuelle 2013 du Conseil d'État - Le droit souple », mai 2013, Conseil d'État, p. 5.

<sup>482</sup> LEVILLAIN K., *Les entreprises à mission: Formes, modèle et implications d'un engagement collectif*, l'École nationale supérieure des mines de Paris, Spécialité Sciences de Gestion, 2015, École doctorale n° 396: Économie, Organisations & Société, p. 14.

<sup>483</sup> « Ethically Aligned Design : A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems » ver. 2, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2017, p. 212.

comme organisationnel, d'un niveau national comme d'un niveau international, du secteur privé comme du secteur public. Dès lors, l'on pourrait constater un large éventail d'instruments cherchant à durcir l'autorégulation.

## **§2 Le durcissement de l'autorégulation à travers des instruments traditionnels**

**158. Annonce du plan.** — Si les entreprises ne sont pas capables de se réguler elles-mêmes, elles devraient avoir des réglementations externes. Notamment le fossé juridique des droits nationaux favorise le *lex shopping* frauduleuse, permettant parfois aux multinationales d'échapper à leur responsabilité<sup>484</sup>. À cet égard, l'autorégulation au sens traditionnel en tant que stratégie clé pourrait sembler limitée. D'où l'intérêt des internormativités entre différents espaces normatifs au niveau transnational, mais également des porosités entre *soft* et *hard law*. Dans cette évolution, la faute de gestion (I), le devoir de vigilance (II) ainsi que l'obligation de prudence (III) constituent l'ouverture du durcissement de *soft law*.

### **I. Par faute de gestion**

**159. L'exemple de la faute de gestion.** — Au début du XXI<sup>ème</sup> siècle, les entreprises les plus puissantes dans le monde travaillent essentiellement sur des systèmes d'IA. La faute de gestion devient un instrument pour que les dirigeants d'entreprises assurent leur responsabilité en fonction de leur pouvoir de *facto*<sup>485</sup>. En effet, une faute de gestion s'entend de l'action ou de l'omission commise par un dirigeant contraire à l'intérêt de son entreprise. Elle n'est pas définie par la loi, mais est laissée à l'appréciation par des tribunaux au cas par cas. Ainsi, l'idée de la faute commise dans la gestion est une notion floue et générale comme elle peut être éventuellement fondée sur le non-respect d'un code de conduite. Comme c'est un élément susceptible d'être pris en compte par le juge pour caractériser la faute de gestion, la méconnaissance d'une règle de *soft law* dans le domaine d'IAG pourrait constituer une faute de gestion et rendre les dirigeants responsables. Autrement dit, vu qu'il n'y a pas un

---

<sup>484</sup> CUZACQ N., « Le mécanisme du Name and Shame ou la sanction médiatique comme mode de régulation des entreprises », RTD com., 2017, p. 473 s.

<sup>485</sup> P. ex., l'article L. 225-251 du code du commerce français retient que « les administrateurs et le directeur général sont responsables individuellement ou solidairement selon le cas, envers la société ou envers le tiers, soit des infractions aux dispositions législatives ou réglementaires applicables aux sociétés anonymes, soit des violations des statuts, soit des fautes commises dans leur gestion », et que « si plusieurs administrateurs ou plusieurs administrateurs et le directeur général ont coopéré aux mêmes faits, le tribunal détermine la part contributive de chacun dans la réparation du dommage ».

critère fixe pour définir la faute par rapport à la gestion du développement de l'IAG, l'on peut donc éventuellement prendre en considération la violation des normes d'autorégulation.

Partant, à travers cet instrument, la faute de gestion peut concrétiser un texte d'autorégulation en lui rendant les effets équivalents d'une loi codifiée. Prenons l'exemple du Conseil de surveillance (Oversight Board) de Meta, organe quasi-judiciaire de l'entreprise, considéré comme une sorte de la « Cour suprême » de Facebook et d'Instagram — compte tenu de son rôle dans le règlement, la négociation et la médiation, y compris le pouvoir d'annuler les décisions de l'entreprise. Créé pour aider l'entreprise à répondre à certaines des questions les plus complexes, les décisions du conseil de confirmer ou d'infirmer les décisions relatives au contenu de Facebook ou d'Instagram ont « force exécutoire », ce qui signifie que l'entreprise doit les mettre en œuvre, à moins que cela ne contrevienne à la loi. Par conséquent, dans l'hypothèse où un tel conseil de surveillance prend les décisions précautionneuses concernant le développement de l'IAG et que les dirigeants de l'entreprise violent les mesures prescrites, il y aurait lieu de considérer que la non-conformité n'est pas convenable au développement durable de l'IAG et donc au détriment de l'intérêt de l'entreprise, constituant ainsi une faute de gestion.

De même et de manière encore plus explicite, ces textes envisagés peuvent aussi provenir des sources du devoir de vigilance.

## II. Par devoir de vigilance

**160. Le devoir de vigilance des multinationales.** — À partir des années quatre-vingt-dix, la jurisprudence anglaise a reconnu l'existence d'un devoir de diligence incombant aux sociétés mères pour les impacts négatifs que peuvent avoir les activités de leurs filiales par le biais du concept du *duty of care*<sup>486</sup>. En France, la loi du 27 mars 2017 relative au devoir de vigilance s'inscrit dans le cadre de l'intégration de la RSE dans des instruments juridiques. Elle vise à inciter les entreprises à engager des démarches responsables. Les sociétés mères et les entreprises donneuses d'ordre doivent s'abstenir de porter atteinte aux droits de l'homme. De cette obligation négative découlent des obligations positives, notamment celle de « due diligence » : prévenir et éviter les impacts négatifs de leurs

---

<sup>486</sup> BRIGHT C., « Le devoir de diligence de la société mère dans la jurisprudence anglaise », *D.* 2017, p. 828.

activités pouvant constituer des violations des droits de l'homme<sup>487</sup>. Il est en effet prévu à l'article L. 225-102-4 I du code du commerce français que « toute société qui emploie, à la clôture de deux exercices consécutifs, au moins cinq mille salariés en son sein et dans ses filiales directes ou indirectes dont le siège social est fixé sur le territoire français, ou au moins dix mille salariés en son sein et dans ses filiales directes ou indirectes dont le siège social est fixé sur le territoire français ou à l'étranger, établit et met en œuvre de manière effective un plan de vigilance »<sup>488</sup>. Vu que les groupes internationaux dont les filiales sont situées à l'étranger sont directement visés, l'objectif du dispositif est de ne pas s'arrêter aux frontières de l'État. Ces règles devraient dès lors pouvoir être imposées à la société, même si celle-ci est régie par une loi différente<sup>489</sup>.

À la lumière de cet engagement transfrontalier, l'on pourrait envisager que le champ matériel du devoir de vigilance soit étendu en matière d'IAG. À titre d'exemple, il devient souhaitable qu'aux États-Unis une telle obligation soit mise en place afin que le groupe Alphabet établisse son plan de vigilance en matière d'IAG au sein de ses filiales comme Google DeepMind, YouTube, Fitbit, et Waymo. Et à l'évidence, à cette vigilance s'ajoute un sens de prudence.

### III. Par obligation de prudence

**161. La faute de témérité.** — Comme personne ne peut prédire comment les IAG pourraient se dérouler dans des circonstances imprévues<sup>490</sup>, une coordination insuffisante entre les systèmes exploités par différents acteurs augmente l'incertitude. Surtout, les défis autour de l'IAG sont plus difficiles à comprendre et à prévoir, en partie parce qu'ils s'appliquent à des systèmes qui n'ont pas encore été développés<sup>491</sup>. Dans le monde du commerce et des applications, l'impact d'un petit progrès scientifique peut être énorme<sup>492</sup>. C'est pour cela qu'il convient de tenir compte de la notion de la faute de témérité, qui a été caractérisée par la formulation : « avait nécessairement conscience qu'il s'ensuivrait probablement un

---

<sup>487</sup> MARTIN-CHENUT K., « Devoir de vigilance : internormativités et durcissement de la RSE », *D.* 2017, p. 798 s.

<sup>488</sup> Art. L. 225-102-4 I du code du commerce français.

<sup>489</sup> PATAUT É., « Le devoir de vigilance - Aspects de droit international privé », *D.* 2017, p. 833.

<sup>490</sup> Cf. supra § L'arrivée de l'IAG : par nature imprévisible.

<sup>491</sup> Testimony préc. de M. Greg Brockman.

<sup>492</sup> FRIDMAN L., « Yoshua Bengio : Deep Learning », préc.

dommage »<sup>493</sup>. Dès lors, l'on pourrait transposer cette notion en considérant que tous ceux qui cherchent à développer l'IAG devraient dûment s'inscrire dans cette prise en conscience des risques de l'IAG. Donc à ce propos, l'on pourrait constater que l'approche de précaution<sup>494</sup>, désormais formellement durcie, est constamment réaffirmée.

**162. En résumé.** — En vue d'établir une approche de coordination à travers l'autorégulation d'IAG, il convient de prendre en compte des instruments classiques du durcissement de soft law. Cette combinaison d'une variété des pratiques de prudence et de précaution pourrait d'ailleurs renforcer la mise en place d'autorégulation, ce qui aboutirait en effet à son éventuelle vocation en tant que nouveau normateur de principe.

---

<sup>493</sup> La faute de témérité, en l'espèce, par pollution. L'affaire Erika, Cass. crim., 25 sept. 2012, n° 10-82.938.

<sup>494</sup> Cf. supra § Principe de précaution.

## Conclusion de la section 1

### **163. De l'autorégulation traditionnelle au principe de faveur pour l'autorégulation. —**

En raison du déséquilibre entre la vitesse du progrès technologique et du droit, de nombreux défis sont régis par des cadres réglementaires de plus en plus obsolètes. Face aux diverses répartitions des tâches transfrontières, l'on a pu apercevoir l'inadéquation des institutions nationales, notamment dans le contexte de la révolution numérique. À cet effet, la codification renvoie désormais explicitement à la notion de la RSE, afin que les acteurs privés prennent en considération les enjeux sociaux et environnementaux de leurs activités.

Effectivement, au début du XXI<sup>ème</sup> siècle, les entreprises les plus puissantes dans le monde travaillent essentiellement sur des systèmes d'IA. Sur ce point, l'on a démontré que comme le développement de l'IA fait souvent partie du monde du commerce et de la consommation, il convient de traiter la question de l'IA prioritairement au niveau de l'entreprise, où l'IA est la plus répandue et où les parties prenantes industrialistes sont les plus aptes à se conformer aux exigences. En même temps, l'on a également souligné que pour la présente étude, étant analysée sous l'angle de l'autorégulation, tous ceux qui travaillent sur l'IAG, même à but non-lucratif, sont indistinctement soumis à certaines conditions de conformité.

Enfin, les mécanismes tels que la faute de gestion, le devoir de vigilance des multinationales ainsi que la faute de témérité nous fournissent un panorama de l'autorégulation. Mais l'effet autorégulateur provenant de ces instruments classiques part d'une logique descendante de top-down, c'est-à-dire sans une voie textuelle ou prétorienne a priori qui mette en valeur son durcissement, l'auto-adoption, quel que soit le niveau et la taille, reste dépourvue de sanction juridique<sup>495</sup>. Cette inefficacité voire hypocrisie pourrait davantage créer l'impression d'incohérence voire de confusion, ce qui est et par conséquent susceptible de rendre l'idée de droit qui cherche avant tout à protéger vide de sens. Ainsi, afin d'établir une cohérence parmi ces évolutions juridiques, cette étude propose la mise en place d'un principe de faveur pour l'autorégulation.

---

<sup>495</sup> V. aussi L'organe de contrainte en ce qui concerne l'autorégulation de l'IA.

## **Section 2 Le principe de faveur pour l'autorégulation**

**164. Annonce du plan.** — Pour discerner un sens équilibré et cohérent au regard des articulations normatives évoquées ci-dessus, la présente étude propose d'établir un principe de faveur au sens d'autorégulation (§1), dont la mise en place (§2) ainsi que l'effet étendu (§3) font successivement l'objet de notre étude.

### **§1 L'introduction du principe de faveur pour l'autorégulation**

**165. Annonce du plan.** — Les créateurs de l'IAG disposent d'une part, de techniques et d'infrastructures pour une telle création et d'autre part, ils dictent et adoptent les normes *sui generis* pour son autorégulation. Comme ces progrès sont les plus avancées en la matière, cette position d'exclusivité rend naturellement son autorégulation concordante le monopole aussi sur le plan normatif. Dès lors, l'application de cette prérogative autorégulatrice pourrait créer l'effet obligatoire *de facto* (I) pour tous les acteurs concernés, voire l'effet de fait (II) pour tout le monde. Ce phénomène provenant simplement de l'autorégulation en matière d'IAG pourrait être davantage contestable du fait de la mutation lucrative en cause (III).

### **I. L'application *de facto* de l'autorégulation**

**166. La mesure de facto d'OpenAI.** — À OpenAI, les chercheurs ont formé un modèle de langage capable de générer des paragraphes cohérents et de permettre une compréhension de lecture rudimentaire : la traduction automatique, la réponse à une question et la synthèse, le tout sans formation spécifique à la tâche. Mais au début, ils ont pris la décision de ne pas partager ce modèle : « En raison de nos préoccupations concernant les applications malveillantes du modèle, l'on ne publie pas le modèle formé ». Car en vertu de la Charte d'OpenAI, d'une part, ils s'attendent à ce que les enjeux de sécurité réduisent leur publication ; d'autre part, ils cherchent à accentuer la préoccupation prépondérante dans le partage de leur recherche concernant la sécurité, les politiques, et les normes<sup>496</sup>. Pour eux,

---

<sup>496</sup> OpenAI Charter, publiée le 9 avr. 2018. La stratégie de diffusion du ladite modèle : « Better Language Models and Their Implications », OpenAI, 14 févr. 2019.

étant donné que l'IA devient de plus en plus performante, les normes de publication dans la recherche de l'IA devraient changer.

En tant qu'outil de réglementation, la charte d'OpenAI a son propre potentiel et soulève ses propres préoccupations. D'aucuns ont réagi à leur décision initiale en prétendant qu'ils n'avaient pas mis suffisamment de temps pour prouver que leur modèle était effectivement dangereux. OpenAI et des experts externes conviennent que ce n'est pas une avancée en soi, mais plutôt un exemple de ce que peut générer le progrès<sup>497</sup>. Cependant, au sens de cette étude, en vertu du principe de précaution, les chercheurs d'IAG ne devraient pas avoir à « prouver » qu'une IA est dangereuse avant de limiter sa diffusion ou d'appliquer leurs normes d'autorégulation.

En effet, à OpenAI, avant qu'un modèle soit publié, ils effectuent une considération de trois niveaux<sup>498</sup>. Dans un premier temps, ils pensent aux nombreuses situations dans lesquelles l'utilisation pourrait mal tourner — par exemple, les types de biais algorithmique et d'autres mauvaises utilisations qu'il pourrait y avoir. Ensuite, OpenAI délibère ses décisions en cherchant un équilibre entre les applications utiles et les aboutissements négatifs afin de déterminer quelles mesures de restrictions qu'ils imposent. Cette pratique consiste en effet dans une formation supplémentaire afin de modifier la capacité du système. Par exemple, avec le modèle de Dall-E, ils l'ont modifié pour qu'il ne soit pas capable de générer des visages reconnaissables. Il leur a fallu du travail supplémentaire pour supprimer cette capacité du modèle. Enfin, lorsqu'ils déploient leurs modèles, ils les mettent sur leurs propres serveurs (Interface de programmation, l'API d'OpenAI), et augmentent au fur et à mesure l'étendue des obligations durant l'utilisation. Cette procédure se fait par affiner soigneusement et progressivement les restrictions d'utilisation — tout en apprenant les utilisations de ces modèles ainsi que leurs mauvais usages. À titre d'illustration, ils ont non seulement anticipé et catégorisé les différentes utilisations, mais également examiné manuellement chaque demande d'utilisation de GPT-3. Et puis, à travers l'apprentissage par renforcement<sup>499</sup> et la rétroaction (feedback) humaine pour apprendre à un modèle de GPT à

---

<sup>497</sup> VINCENT J., « AI researchers debate the ethics of sharing potentially harmful programs », *Verge*, 21 févr. 2019.

<sup>498</sup> Ilya Sutskever, l'un des cofondateur d'OpenAI, à l'occasion de « 2022 HAI Spring Conference on Key Advances in Artificial Intelligence », 12 avril 2022.

<sup>499</sup> Cf. supra § L'apprentissage par renforcement.

suivre les intentions et les instructions de ses utilisateurs ; sur ces bases, ils ont construit une expertise institutionnelle pour appréhender les utilisations acceptables et les cas répréhensibles. Enfin, comme les modèles sont sur leurs propres serveurs, ils ont donc un contrôle très précis sur le type de requête. Et en cas de désapprobation, il y a également une pause d'utilisation.

Ce cadre d'autorégulation d'OpenAI commence par des utilisations mineures et en tirer des leçons. Ils diffusent leurs modèles de manière graduelle, et affinent leur politique et leur contrôle sur la base de ces expériences. En d'autres termes, au lieu de procéder à une grande diffusion discontinue et de se retrouver confronté à une crise, ils procèdent à de très petites diffusions, ce qui permet d'apprendre de ce qui se passe, et de l'intégrer dans le modèle ainsi que la politique d'autorégulation. Partant, il convient de souligner que c'est en effet cette gradualité qui fonde l'approche restrictive d'OpenAI. En outre, l'on pourrait constater que cette démarche dite unilatérale d'autorégulation constitue désormais un effet de fait au sein et au-delà de la communauté de recherche.

## **II. L'effet de fait de la volonté unilatérale d'autorégulateur**

**167. L'effet de fait unilatéralement étendu.** — L'exemple d'OpenAI fournit en réalité un *accord étendu dit négatif ou indirect de l'application de leur charte* aux autres organisations ainsi qu'aux individus de recherche. En suivant son propre standard d'autorégulation, OpenAI ne tient pas compte des autres normes moins exigeantes que l'application de leur charte. En revanche, en décidant de ne pas publier leur modèle, ou de le partager uniquement sous leur contrôle interne d'approbation et de gradualité, OpenAI a effectivement imposé le standard issu de leur charte aux autres acteurs concernés. Autrement dit, la volonté d'appliquer leur norme de sources autorégulatrices, influe en même temps et de facto, sur la situation d'autres acteurs au sein et en dehors de la communauté d'IAG, que ce soit de manière explicite ou non. Partant, leur choix constitue en effet une sorte *d'autorégulation d'application négative ou indirecte de leur charte*, tout en respectant des dispositifs en vigueur pouvant être plus contraignants ou plus exigeants.

Malgré cette démarche de nature informelle voire tacite, l'influence de sa police régulatrice demeure tangible. Il y a donc une sorte de l'effet de fait provenant de cette approche

d'autorégulation d'application étendue du caractère négatif et indirect, qui implique une obligation substantive de ne pas faire. Certes, l'on pourrait se féliciter de cette précaution étendue provenant d'un engagement volontaire. Pour autant, il convient de s'interroger sur l'éventuel aboutissement d'un tel apport — initialement issue de l'autorégulation — apparemment susceptible d'être remis en cause par le changement de sa structure, de but non lucratif à but lucratif.

### III. Le passage du non lucratif au lucratif : l'exemple d'OpenAI

**168. Mutation structurelle.** — Au départ, OpenAI était une association à but non lucratif. Cependant, à partir de 2019, OpenAI est désormais une entreprise à « but lucratif plafonné ». L'entreprise annonce que ce changement structurel n'affecte pas sa mission d'assurer que l'IAG et des systèmes hautement autonomes, qui surpassent les humains au plus haut niveau de performance économique, profitent à toute l'humanité. Leur choix nous permet de comprendre les limites inévitables de la structure sans but lucratif pour la recherche de l'IA : OpenAI LP (limited partnership) a désormais une obligation de maximiser ses profits<sup>500</sup>, ce qui pourrait tout à fait être incompatible avec son engagement initial de s'assurer que les avantages de l'IA soient largement distribués, et ainsi, d'après les mots de leur Charte : « partager les bénéfices avec le monde entier ». Par exemple, en cédant une licence d'exploitation exclusive du modèle de langage GPT-3 à Microsoft en 2020, l'on pourrait se demander comment ces intérêts privés arrivent à partager les bénéfices avec le monde entier.

Ce changement structurel a également changé leurs approches de publication. Auparavant, OpenAI a choisi de publier les modèles, tels que GPT-2, qui était au début non-partagé. Mais en créant un nouveau modèle de langage, GPT-3, qui est la version bien plus puissante de son prédécesseur GPT-2, OpenAI a décidé de le mettre sur le marché en tant que produit commercial<sup>501</sup>. OpenAI le voit comme un moyen d'assurer que l'entreprise dispose de ressources financières suffisantes pour leur recherche en IA<sup>502</sup>.

---

<sup>500</sup> Microsoft a en effet investi 1 milliard de dollars dans OpenAI pour créer des applications d'IA qui profitent au plus grand nombre. GILES M., « Microsoft is investing \$1 billion in OpenAI to create brain-like machines », MIT Tech. Rev., 22 juill. 2019.

<sup>501</sup> Microsoft obtient en outre une licence exclusive pour le modèle. WIGGERS K., « Microsoft gets exclusive license for OpenAI's GPT-3 language model », VentureBeat, 22 sept. 2020.

<sup>502</sup> V. « OpenAI API », 11 juin 2020.

**169. En résumé.** — Dorénavant, la volonté unilatérale d'autorégulation pourrait se transformer en une obligation négative d'abstention, imposant horizontalement une application implicite des normes d'autorégulation, pour que le principe de précaution soit effectivement respecté. Partant, soit par sa structure soit par son approche de publication, la politique d'OpenAI impose son influence réelle aux chercheurs d'IAG, et crée ainsi une régulation de facto provenant de ses politiques purement internes. À juste titre, en vue d'encadrer l'autorégulation étendue en matière d'IAG, cette étude propose la mise en place d'un principe de faveur de nature extraterritorial.

## §2 La mise en place du principe de faveur

**170. Annonce du plan.** — Il convient d'étudier le contenu (I) ainsi que les caractéristiques (II) du principe de faveur.

### **I. Le contenu du principe de faveur**

**171. L'application d'autorégulation.** — Comme les choix faits et les décisions prises par les créateurs d'IAG sont d'un effet immédiat résultant de son application prompte d'autorégulation, il est par conséquent tout à fait souhaitable que les principes et les chartes tels que la charte d'OpenAI, DeepMind Ethics & Society Principles<sup>503</sup>, les principes d'IA de Microsoft<sup>504</sup> et les principes d'IA de Google<sup>505</sup>, la gouvernance du conseil de surveillance (Oversight Board) de Meta<sup>506</sup>, ainsi que la mission de Partnership on AI<sup>507</sup> et d'autres institutions concernées<sup>508</sup> puissent être prises en compte comme des accords d'application

---

<sup>503</sup> <https://www.deepmind.com/safety-and-ethics>. V. aussi. p. ex. ORTEGA P. A., MAINI V., DeepMind Safety team, « Building safe artificial intelligence: specification, robustness, and assurance », DeepMind Safety Research, 27 sept. 2018.

<sup>504</sup> Aussi, v. le comité de l'éthique en IA (AETHER, AI and ethics in engineering and research) de Microsoft : « AETHER veillera à ce que notre plateforme d'IA et nos efforts d'expérience soient profondément ancrés dans les valeurs et principes fondamentaux de Microsoft et profitent à la société en général ». V. Satya Nadella email to employees: Embracing our future: Intelligent Cloud and Intelligent Edge, 29 mars 2018.

<sup>505</sup> Google AI principles, publiés le 7 juin 2018.

<sup>506</sup> « Establishing Structure and Governance for an Independent Oversight Board », Facebook, 17 sept. 2019.

<sup>507</sup> <https://partnershiponai.org/about/>.

<sup>508</sup> Par ex., il est prévu dans le British Computer Society (BCS) Code of Conduct que BCS a la responsabilité d'établir des règles et des normes professionnelles pour orienter le comportement de ses membres dans les affaires professionnelles. Il s'attend à ce que ces règles et normes professionnelles soient plus strictes que celles établies par la loi générale ; des mesures disciplinaires peuvent entraîner l'expulsion de l'adhésion. Les membres sont censés exercer leur propre jugement pour répondre aux exigences du code et demander conseil en cas de doute. V. Code of conduct for BCS members, Approved by Trustee Board 3 June 2015. D'autres exemples que l'on pourrait citer : le Code of Ethics and Professional Conduct de l'ACM (Association for

étendue. Ces groupes qui sont en général plus aptes à élaborer les normes d'autorégulation, pourraient par ailleurs se voir imposer une obligation positive de fournir des moyens techniques de sécurité et de précaution aux autres entités de recherche afin de mieux garantir une application efficace de ces règles.

À titre d'illustration, OpenAI met à jour leur accord sur l'accès au logiciel (Software Access Agreement) qui retient toute une série de droits et d'obligations qui concerne à la fois les dispositifs en vigueur, telle qu'une nouvelle configuration de la jouissance, et la restriction de la propriété intellectuelle, à laquelle s'ajoute des clauses plus spécialisées telles que les règles de publication et de confidentialité<sup>509</sup>. En outre, leur publication de l'accord juridique sur libre accès vise à faciliter le partage de modèles entre les établissements partenaires. À travers ces communications interorganisationnelles, même des acteurs non signataires tels que PNL Hugging Face et Allen Institute for Artificial Intelligence ont adopté une approche de publication similaire à celle d'OpenAI pour leur gouvernance de publication<sup>510</sup>.

Ce cadre juridique élaboré entièrement dans les mains d'autorégulateur, a pour objectif d'aménager le partage de modèles entre les organisations et d'aider d'autres organisations à mettre en place de tels systèmes de partage. En publiant leur application avec l'intention de fournir des données probantes sur lesquelles la communauté d'IA peut s'appuyer lorsqu'elle réfléchit aux défis de publication inhérents à certains aspects de la recherche sur l'IA<sup>511</sup>, OpenAI est, en tant qu'*autorégulateur d'autres autorégulateurs*, intervenu pour gérer une diffusion progressive et de partage des modèles basée sur le partenariat. À notre sens, ce modèle autogénéré qui porte une influence à la fois formelle et substantielle au sein

---

Computing Machinery) dont le préambule déclare que l'engagement envers une conduite professionnelle éthique est attendu de chaque membre de l'ACM. Ce code, composé de 24 impératifs, formulés comme des déclarations de responsabilité personnelle, identifie les éléments d'un tel engagement. Par exemple, s'agissant des impératifs moraux généraux, il est prévu qu'en tant que membre de l'ACM, il doit contribuer à la société et au bien-être humain. Les membres sont également obligés d'éviter de nuire aux autres, d'être honnêtes et dignes de confiance, d'être justes et de prendre des mesures non-discriminatoires. Ils doivent respecter la vie privée des autres et la confidentialité. V. aussi Rules of Conduct de IET (The Institution of Engineering and Technology). Ces règles prévoient que les membres acceptent la responsabilité personnelle pour tout travail effectué par eux ou sous leur surveillance ou direction, et ils doivent également prendre toutes les mesures raisonnables pour s'assurer que les personnes travaillant sous leur autorité sont à la fois convenablement équipées et compétentes pour mener à bien les tâches qui leur sont confiées.

<sup>509</sup> Open AI Software Access Agreement, version du 5 septembre 2019.

<sup>510</sup> « OpenAI Report : Release Strategies and the Social Impacts of Language Models », OpenAI, août 2019, pp. 2-3.

<sup>511</sup> « GPT-2: 6-Month Follow-Up », OpenAI, 20 août 2019.

de la communauté de la recherche est donc susceptible d'être le fondement clé d'une publication responsable dans le développement de l'IAG. Partant, cette démarche volontaire constitue désormais un instrument essentiel pour une régulation efficace.

**172. L'exemple de Deep Fake Detection.** — En ce sens, un autre exemple que l'on pourrait citer concerne la politique dite Deep Fake Detection Challenge de Facebook. En effet, Facebook a annoncé qu'il retirera de sa plateforme les vidéos modifiées par l'IA car ces vidéos « déforment la réalité ». En rappelant les Standards existants de la communauté, pour « aller plus loin », le réseau social évoque même les critères pour la suppression des contenus trompeurs. Les critères, autoédités par Facebook, disposent que sera retiré un contenu qui « a été édité ou synthétisé — au-delà des ajustements pour la clarté ou la qualité — d'une manière qui n'est pas évidente pour une personne raisonnable et qui pourrait probablement induire en erreur car faisant croire qu'un sujet de la vidéo aurait dit des mots qu'il n'a pas réellement prononcés » ; surtout s'il s'agit d'un « produit de l'intelligence artificielle ou de l'apprentissage automatique qui fusionne, remplace ou superpose le contenu d'une vidéo, le faisant paraître authentique »<sup>512</sup>. Tirant cette formule véritablement pertinente, l'on voit qu'incontestablement, cette intervention d'autorégulation illustre combien il est convenable et efficace pour ces autorégulateurs, avec des solutions précises et techniquement faisables, de régler les problèmes qu'ils ont permis ou créés, avant même que les gouvernements étatiques ne réagissent par le droit positif. Cette étude fait valoir une telle gestion adéquate et ponctuelle provenant de l'aptitude dite souveraine d'autorégulation. Dès lors, il convient d'étudier les caractéristiques de ce principe de faveur.

## II. Les caractéristiques du principe de faveur

**173. Annonce du plan.** — L'on verra l'intérêt (A) ainsi que le mécanisme de renvoi (B) du principe de faveur pour l'autorégulation.

### A. L'intérêt du principe de faveur

**174. Un sens plus favorable.** — La vocation du principe de faveur d'IAG préconise une garantie minimum de précaution. Tout comme le principe de faveur en matière du droit du

---

<sup>512</sup> BICKERT M., « Enforcing Against Manipulated Media », Facebook Global Policy, 6 janv. 2020.

travail, le consensus « négocié » pour le développement de l'IAG est un plan favorisé pour son application plus adaptée, mais il n'affecte pas le seuil minimum exigé par la loi : il peut aller au-delà du cadre supplétif, mais pas en deçà. L'idée de ce principe de faveur en matière d'IAG est donc d'encourager les parties prenantes à la recherche et à la création d'IAG à élaborer leurs propres normes qui leur semblent les plus adéquates et adaptées. Ces normes peuvent toutefois déroger au dispositif supplétif du deuxième bloc de la loi en vigueur, mais uniquement dans un sens plus favorable. Et ces deux blocs inférieurs ne peuvent contredire le troisième bloc du droit d'IAG tel que le statut d'Homo sapiens<sup>513</sup>.

Dès lors, dans un sens de long terme, l'autorégulation d'IAG a pour but à la fois de respecter le principe de précaution, et de coordonner les compétences d'autorégulateurs, afin de mieux protéger les humains au cours de leur coévolution avec l'IAG. Dans un sens de court terme, cette démarche étant au plus près du terrain de la gestion du développement exponentiel d'IAG, pourrait en outre fournir une réponse apte et directe en comportant des stipulations plus favorables aux humains que le droit constant en vigueur.

## **B. Les renvois du principe de faveur**

**175. Le renvoi au premier degré d'autorégulation.** — Le principe de faveur consiste donc dans le fait que la portée de la politique d'autorégulation ne saurait contredire ce qui est prévu par la disposition en vigueur que de manière préférable. Il est donc souhaitable qu'il y ait une obligation légale de transparence pour la recherche de l'IAG qui l'emporterait sur la politique engagée au sein de toute équipe de recherche en la matière. Par exemple, dans le cas où la non publication des modèles d'IAG soulèverait une confidentialité douteuse qui remette en cause la sécurité<sup>514</sup>, il y aurait lieu pour le droit positif de prévoir une obligation de contrôle de leur recherche. Mais avant l'intervention du droit constant, il peut y avoir le renvoi aux standards des principes d'autorégulation citées ci-dessus en la matière pour remédier aux difficultés. En effet, si certaines parties ne sont pas capables de les suivre pour des raisons techniques ou économiques, une obligation positive de fournir des moyens ainsi que de coordination ne serait pas exclue pour ces acteurs principaux dans l'arène de l'IAG.

---

<sup>513</sup> Cf. infra le répertoire de l'IAG.

<sup>514</sup> V. p. ex. HAO K., « The messy, secretive reality behind OpenAI's bid to save the world », MIT Tech. Rev., 17 févr. 2020.

**176. Le renvoi au second degré du dispositif supplétif.** — Si ces règles et normes autorégulatrices d'IAG ne parviennent pas à être plus strictes et exigeantes que celles établies par la loi en vigueur, elles ne pourront être appelées à jouer un rôle dépourvu d'effet. En l'absence d'une autorégulation applicable, l'on pourrait aboutir à une désignation de règles éventuellement susceptibles d'être un seuil minimum pour la régulation : ce serait l'aiguillage vers une disposition en vigueur du droit comparé qui fournirait une réponse de manière supplétive<sup>515</sup>.

Un exemple d'articulation normative entre les deux niveaux se trouve dans la « Privacy manifesto » d'iRobot<sup>516</sup>. Au fil du temps, iRobot s'est concentré sur le foyer des consommateurs pour aider la maison à être automatiquement entretenue donc plus saine, plus sûre et plus efficace : la maison elle-même fait partie d'un système qui joue un rôle actif dans les soins. L'approche d'iRobot pour protéger les données et la vie privée de leurs consommateurs passe par l'adoption de leur *manifesto* de confidentialité, affirmant qu'ils ne violent jamais la confiance de leurs clients en vendant ou en abusant des données relatives aux clients, y compris les données recueillies par leurs produits connectés<sup>517</sup>. Ils ont adopté le RGPD non seulement comme exigé, mais globalement, ils ont veillé à ce que les images ne quittent pas le robot, de sorte que le traitement des données des capteurs visuels se fait localement sur le robot et que seule la connaissance sémantique de la maison avec le consentement du consommateur est envoyée. L'entreprise affiche ce qu'elle sait et utilise ces données pour permettre l'application du robot avec le consentement et la compréhension du consommateur<sup>518</sup>.

L'exemple d'iRobot illustre clairement l'idée du principe de faveur : l'autorégulation peut aller au-delà du dispositif en vigueur, mais pas en deçà. Partant, l'on souligne l'importance de cette évolution juridique au début du XXIème siècle : le fait qu'une entreprise américaine iRobot se conforme entièrement à un règlement européen, le cas d'espèce du RGPD, par voie à la fois volontaire et transfrontalière. Ces deux facteurs, à savoir « l'auto-adoption » combinée d'« une absence du concept de territoire » pour l'acquisition de facto d'une

---

<sup>515</sup> Cf. infra le répertoire de l'IAG.

<sup>516</sup> On note d'ailleurs que l'entreprise a été rachetée par Amazon en 2022.

<sup>517</sup> iRobot Privacy and Data Sharing Common Questions.

<sup>518</sup> Colin Angle, AI Podcast de Lex Fridman, 19 sept. 2019.

compétence juridique constituent, à notre sens, un début de la transformation juridique de nos jours. Cela étant dit, une fois la nature d'IAG identifiée dans le cas d'espèce, le principe de faveur intervient pour résoudre le conflit des normes par son effet étendu.

**177. En résumé.** — La dérogation admise uniquement dans l'hypothèse d'une intervention au sens plus favorable ne saurait néanmoins déroger aux dispositions issues du sommet de la hiérarchie des normes d'IAG. Autrement dit, les textes autorégulateurs d'IAG ne peuvent en aucun cas déroger aux dispositions qui revêtent un caractère absolu du droit d'IAG telles que le statut existentiel de l'espèce humaine. En revanche, ils peuvent toujours comporter des dispositions plus favorables à l'humanité que celles des lois en vigueur, qui revêtent un caractère relatif et avec la possibilité d'être dérogées dans un sens plus favorable. Ce mécanisme sert ainsi à résoudre le concours de normes de catégories et de nature traditionnellement différentes. Partant, le conflit entre les espèces est résolu par celle qui est la plus favorable en cas de conflit de normes.

### **§3 L'effet étendu du principe de faveur**

**178. Annonce du plan.** — Dans le cadre du principe de faveur d'autorégulation, son effet étendu implique qu'une obligation positive est horizontalement créée (I). Cette source d'obligation est en conséquence suivie par la question de l'invocabilité des règles (II).

#### **I. L'obligation positive horizontale**

**179. Développeur-payeur.** — Si la souveraineté devrait signifier « le droit et le devoir d'un État de protéger et de satisfaire aux besoins de son peuple »<sup>519</sup>, la nouvelle souveraineté devient par conséquent « la capacité de participer aux régimes, réseaux et institutions internationaux et transgouvernementaux qui sont désormais nécessaires pour permettre aux gouvernements d'accomplir par la coopération des uns avec les autres ce qu'ils n'ont pu accomplir en agissant seul dans un territoire défini »<sup>520</sup>. Dans ce sens, suivant les éléments remarqués plus haut, les acteurs principaux en IAG verraient en effet leur « droit et capacité de participer aux régimes, réseaux et institutions internationaux, suivis par une responsabilité

---

<sup>519</sup> SLAUGHTER A.-M., « Sovereignty and Power in a Networked World Order », *Stanford Journal of International Law*, vol. 40, 2004, pp. 327.

<sup>520</sup> *Ibidem.*, p. 285.

d'engager aux certaines exigences minimales d'adhésion »<sup>521</sup> ; et cette nouvelle identité leur conférerait la légitimité de leur rôle, en tant que régulateur de facto voire de jure, dans leurs propres engagements ainsi que dans un aménagement normatif en commun.

De même qu'en matière d'atteintes à l'environnement, un niveau de protection élevé est fondé sur le principe du pollueur-payeur<sup>522</sup>, peut être favorisé en matière d'IAG sous l'application d'un principe du *développeur-payeur*. Ainsi, l'un des aspects de ce principe de faveur reflète brièvement l'idée du principe de responsabilités communes mais différenciées et des capacités respectives (common but differentiated responsibilities and respective capabilities, cbdr-rc). Sauf qu'ici, le sujet envisagé ne consiste plus dans les responsabilités et capacités différenciées des pays respectifs par un accord multilatéral. En revanche, par la combinaison à la fois du principe de faveur et de l'esprit de « cbdr-rc », l'application de ce principe de faveur vise un nouveau type de sujet sur la scène du droit international : l'autorégulateur d'IAG.

Prenons l'exemple d'une formule classique de « cbdr-rc » à l'article 2.2 de l'accord de Paris sur les changements climatiques : l'application conforme à l'équité et au principe des responsabilités communes mais différenciées et des capacités respectives, eu égard aux différentes situations nationales<sup>523</sup>. À notre avis, en matière d'IAG, ce mécanisme peut privilégier le soutien technique, et ne pas se limiter au soutien financier, tout en tenant compte du principe général d'une égalité substantielle, ou même simplement, renvoyer au principe de proportionnalité pour l'application étendue de faveur. Partant, la mise en place du principe de faveur combinant le mécanisme du développeur-payeur signifie que l'invocabilité des règles devrait être envisagée.

---

<sup>521</sup> *Ibid.*, p. 325.

<sup>522</sup> Règlement (CE) n° 864/2007 du Parlement Européen et du Conseil du 11 juillet 2007 sur la loi applicable aux obligations non contractuelles (Rome II), pt. 25.

<sup>523</sup> Accord de Paris sur le climat, art. 2.2. Son article 4.3 fait une distinction entre les parties développées et les parties en développement. Ce premier groupe devrait « continuer de montrer la voie en assumant des objectifs de réduction des émissions en chiffres absolus à l'échelle de l'économie », tandis que la seconde catégorie devrait « continuer d'accroître leurs efforts d'atténuation », « passer progressivement à des objectifs de réduction ou de limitation des émissions à l'échelle de l'économie eu égard aux différentes situations nationales ». À cet égard, son article 4.5 dispose qu'« un appui est fourni aux pays en développement Parties pour l'application (...), étant entendu qu'un appui renforcé en faveur des pays en développement Parties leur permettra de prendre des mesures plus ambitieuses ». Et son article 7.13 s'engage à un appui international renforcé fourni en permanence aux pays en développement Parties.

## II. L'invocabilité des règles du principe de faveur

**180. Les engagements volontaires désormais dits contraignants.** — La Cour de cassation française a déjà consacré dans l'affaire Erika<sup>524</sup> l'extension de la responsabilité civile et pénale des compagnies transnationales sur le fondement des engagements volontaires de celles-ci<sup>525</sup>. En consacrant l'adage *tu patere legem quam ipse fecisti*, le caractère volontaire du droit souple ne l'empêche toutefois pas d'être source d'obligations juridiques pourvues de sanctions. À notre avis, cette tendance du durcissement de la RSE<sup>526</sup> pourrait trouver sa nouvelle forme sur le fondement du principe de faveur. En vertu de l'effet étendu d'autorégulation, l'application de ces engagements spontanés devrait avoir la même valeur juridique que celle des dispositions supplétives en vigueur. À cet effet, les textes conventionnels peuvent créer des droits et obligations directement liés aux parties prenantes entrant dans son champ d'application. Sur ce fondement, la violation des normes jusqu'ici non contraignantes constitue désormais une inexécution par les organisations signataires de ses obligations. Une telle évolution entraîne dans un premier temps des conséquences dans les rapports contractuels entre les parties signataires. Mais il en va de même en cas d'inexécution des obligations mises à leur charge par l'autorégulation, ce qui aurait pour conséquence d'engager leur responsabilité à l'égard des tiers, voire au profit des intérêts communs envers les générations futures.

**181. En résumé.** — Dès lors, en cas de violation des dispositions conventionnelles, non seulement les parties peuvent reprocher à un organisme son manquement, mais il deviendrait également possible que les tiers invoquent l'inexécution du texte d'autorégulation. Partant, chaque individu peut agir afin d'obtenir l'exécution des engagements énoncés dans le cadre d'autorégulation ou des dommages-intérêts contre les parties liées par le texte qui violeraient ces engagements volontaires en matière d'IAG.

---

<sup>524</sup> L'affaire Erika préc., Cass. crim., 25 sept. 2012, n° 10-82.938.

<sup>525</sup> La jurisprudence évoluera en faveur de la reconnaissance d'une portée juridique de ces codes. MAZIAU N., « L'extraterritorialité du droit entre souveraineté et mondialisation des droits », *La Semaine Juridique Entreprise et Affaires* n° 28, 9 juill. 2015, p. 26.

<sup>526</sup> V. supra le durcissement d'autorégulation.

### **Conclusion de la section 1**

**182. Les engagements volontaires désormais dits contraignants.** — À travers l'imposition de facto de l'autorégulation, l'on a constaté l'effet de fait de la volonté unilatérale d'autorégulateur. Cette démonstration nous conduit à proposer le principe de faveur extraterritorial pour encadrer son effet de fait communément étendu. Dans cette démarche, l'on a vu la mise en place du principe de faveur pour l'autorégulation ainsi que ses caractéristiques. Ensuite, cette étude a dégagé l'intérêt ainsi que les renvois du principe de faveur pour l'autorégulation. Au regard de ces résultats, l'on a souligné que les deux éléments décisifs, à savoir « l'auto-adoption » combinée d'« une absence du concept de territoire » pour l'acquisition de facto d'une compétence juridique constituent l'évolution juridique de nos jours. Suivant cette analyse, l'on pourrait conclure que l'obligation positive fondant sur la notion de développeur-payeur consolide l'invocabilité des règles issues du principe de faveur. Ce résultat a pour conséquence de reconnaître les engagements volontaires désormais dits contraignants.

## Conclusion du chapitre 2

**183. L'auto-adoption en l'absence du concept de territoire.** — De nouvelles formes de codifications ont émergé, fondées sur une participation directe de leurs auteurs à l'élaboration de leur contenu. Dans ces codifications privées, les auteurs occupent une position originale. En effet, disposant d'une compétence que le législateur lui-même à peine reconnaître, ces auteurs agissent en qualité d'experts, revendiquent leur indépendance, et assument la paternité du texte proposé<sup>527</sup>.

L'on a démontré qu'en matière d'IAG, en vertu du principe de faveur, l'application des normes concerne non seulement l'autorégulateur lui-même mais vise en outre l'extension d'autorégulation. Car en raison de sa position d'exclusivité qui rend son autorégulation aussi monopole sur le plan normatif, les conditions peuvent donc ne pas être négociées directement par tous les sujets impliqués dans son application, son effet relève néanmoins de tout développement d'IAG. Par conséquent, à l'inverse des méthodes classiques partant d'une logique descendante de *top-down*<sup>528</sup> ce cadre entièrement élaboré à partir du terrain d'autorégulateur renforce une logique ascendante dans le même mouvement du durcissement de *soft law* ; c'est la nouvelle logique conçue par le principe de faveur pour l'autorégulation.

---

<sup>527</sup> CHANTEPIE G., « Les codifications privées », in l'Association Henri Capitant, *Le Droit Souple*, Journées nationales de l'Association Henri Capitant, t. XIII, Dalloz, 2009, p. 43.

<sup>528</sup> V. supra les instruments classiques d'autorégulation.

## Conclusion du titre 1

**184. Le passage du principe de faveur au droit positif.** — L'on vient de démontrer que des acteurs non étatiques sont de plus en plus en mesure d'interférer avec des relations interétatiques et « de développer, sinon un multilatéralisme alternatif, du moins une présence dans ce registre »<sup>529</sup>. Dans son discours, Mme Margrethe Vestager souligne le fait que les géants du numérique disposent des pouvoirs « qui leur permettent d'agir presque comme des régulateurs », et qu'il est nécessaire de surveiller de près la façon dont ces géants utilisent « leur pouvoir réglementaire »<sup>530</sup>. Vu que la majeure partie de l'investissement et de l'innovation dans le domaine de l'IA provient d'entreprises privées<sup>531</sup>, le durcissement de l'éthique et l'internalisation de la responsabilité sont inévitables en ce qui concerne l'espace dans lequel cohabitent les humains et les IAG. Partant, en vertu du principe de faveur pour l'autorégulation, l'articulation normative devrait pouvoir trouver un seuil minimum entre les acteurs interétatiques afin que les exigences soient véritablement respectées. Suivant ce chemin du principe de faveur pour l'autorégulation, l'on va à présent étudier le reste de la nouvelle hiérarchie des normes : l'aboutissement d'un répertoire juridique en matière d'IAG.

---

<sup>529</sup> « Comme ils se situent dans un mode juridique différent, on peut à leur sujet parler de 'multiversalisme' parce qu'ils appartiennent à un univers parallèle ». SUR S., « Multilatéralisme et multiversalisme », in RUNAVOT M.-C. et RIOS RODRIGUEZ J. (dir.), *Le droit international multilatéral*, op. cit., p. 410.

<sup>530</sup> « Digital power at the service of humanity », Conference on Competition and Digitisation, European Commission, Copenhagen, 29 nov. 2019. De nombreux acteurs non étatiques qui ne sont pas le « sujet » du droit international ont un impact législatif mondial. MCCORQUODALE R., « Sources and the Subjects of International Law: A Plurality of Law-Making Participants », in BESSON S., D'ASPREMONT J. (eds), *Oxford Handbook of the Sources of International Law*, OUP Oxford, 2017, p. 767.

<sup>531</sup> Résolution préc. du 12 févr. 2019, pt. 25.

## TITRE 2 DEUXIEME DEGRE D'APPLICATION : LA SYNCHRONISATION DU REPERTOIRE DE L'IAG

*« On verrait l'oppression s'accroître continuellement sans que les opprimés pussent jamais savoir quel terme elle aurait, ni quels moyens légitimes il leur resterait pour l'arrêter. » — Jean-Jacques Rousseau*

**185. Annonce du plan.** — Après avoir illustré le renvoi à un premier degré d'autorégulation, dans le cadre de notre principe de faveur, il reste à étudier le renvoi au niveau de second degré du dispositif supplétif. Comme prévu plus haut, ce niveau de renvoi consiste dans l'aiguillage vers une disposition en vigueur du droit comparé qui fournirait une réponse de manière supplétive. Dès lors, en se fondant sur un pluralisme juridique des normes d'IAG, cette étude propose la mise en place d'un répertoire de l'IAG (chapitre 2). Mais pour parvenir à une telle convergence, il convient d'étudier au préalable les évolutions concrétisant la compétence inclusive de l'IAG (chapitre 1).

## Chapitre 1 La compétence inclusive de l'IAG

*« La contradiction réside dans l'idée même de l'État. L'État universel n'ayant jamais pu se réaliser, chaque État est un être restreint comprenant un territoire limité et un nombre plus ou moins restreint de sujets. L'immense majorité de l'espèce humaine reste donc en dehors de chaque État, et l'humanité tout entière est partagée entre une foule d'États grands, moyens ou petits, dont chacun, malgré qu'il n'embrasse qu'une partie très restreinte de l'espèce humaine se proclame et se pose comme le représentant de l'humanité tout entière et comme quelque chose d'absolu. Par là même, tout ce qui reste en dehors de lui, tous les autres États, avec leurs sujets et la propriété de leurs sujets, sont considérés par chaque État comme des êtres privés de toute sanction, de tout droit, et qu'il a par conséquent celui d'attaquer, de conquérir, de massacrer, de piller, autant que ses moyens et ses forces le lui permettent. Qu'on n'est jamais parvenu à établir un droit international, et qu'on n'a jamais pu le faire précisément parce que, au point de vue de l'État, tout ce qui est dehors de l'État est privé de droit. » — Mikhaïl Bakounine*

**186. Annonce du plan.** — La science n'a pas de nationalité et l'IAG ne connaît pas de frontières. À la recherche d'une transition juridique adéquate et pacifique pour ce qui est de l'IAG, l'approche inclusive d'IAG (section 2) mérite d'être analysée à travers le principe d'extraterritorialité (section 1).

### Section 1 Le principe d'extraterritorialité

**187. Annonce du plan.** — Déjà, la mondialisation témoigne d'une territorialité relativisée (§1). De ce fait, la méta-régulation (§2) s'installant dans ce monde directement interconnecté ne peut plus être négligée.

#### §1 La territorialité atténuée

**188. Le bouleversement des frontières.** — De nos jours, la globalisation implique une intégration et une multiplication des effets de toute action. Elle conduit à un développement

des exceptions au principe de territorialité<sup>532</sup>. En effet, l'extraterritorialité des droits désigne « une situation dans laquelle les compétences de l'État, que celle-ci soit législative, juridictionnelle, ou exécutive, régissent des rapports de droits situés en dehors du territoire de cet État »<sup>533</sup>. Sa multiplication témoigne donc d'un « changement dans l'approche de la conceptualisation du pouvoir réglementaire souverain — un changement qui, en s'écartant d'une vision territoriale de l'autorité de réglementation, s'adapte à l'application plus souple des règles de compétence que les litiges transnationaux l'exigent »<sup>534</sup>. Ainsi, « Le droit international de la globalisation est un droit qui affecte les individus, et dont l'objet est en étroite relation avec leurs intérêts privés. Il est clair que dans cet espace global écartelé et ouvert, les frontières entre systèmes nationaux vont tomber d'elles-mêmes »<sup>535</sup>. Des relations et des régulations juridiques établies depuis plusieurs siècles révèlent davantage les limites de leur pertinence et de leur efficacité à capturer leur objet et à le faire comprendre. Les principes, les concepts et les outils du droit sont à réévaluer, et les classifications et les catégories, voire le droit et une nouvelle logique des normes est à repenser<sup>536</sup>.

En matière d'IA, afin de prévenir la duplication ou la fragmentation des efforts ainsi que de garantir l'adoption de politiques cohérentes et l'échange des bonnes pratiques en vue d'une utilisation plus répandue de l'IA, le Parlement européen souligne « la nécessité de coopérer avec des partenaires internationaux dans des contextes bilatéraux et multilatéraux, aux fins de l'évolution et de l'adoption éthiques de l'IA ». Le Parlement reconnaît ainsi que l'IA est « une technologie à incidence mondiale », une question qui « n'a pas de frontières », « offrant des avantages communs et présentant des défis similaires ». Par conséquent, une approche mondiale est nécessaire pour une coopération au niveau international, afin

---

<sup>532</sup> Jean-Marc Sauvé, « La territorialité du droit », la Rencontre inter-réseaux (franco-américain, franco-brésilien et franco-chinois) sur le thème « Internationalisation du droit : pathologie ou métamorphose de l'ordre juridique ? », les 10, 11 et 12 avril 2012 au Collège de France.

<sup>533</sup> Nicolas Maziau, « L'ordre juridique national en prise avec le droit européen et international : questions de souveraineté ? », colloque organisé par la Cour de cassation et le Conseil d'État, 10 avril 2015.

<sup>534</sup> BUXBAUM H. L., « Transnational Regulatory Litigation », *Virginia Journal of International Law*, vol. 46:2, 2006, pp. 305-306.

<sup>535</sup> DAHMEN K. B., *Interactions du droit international et du droit de l'Union européenne : Un pluralisme juridique en matière de propriété industrielle*, Logiques Juridiques, 2013, p. 28.

<sup>536</sup> FRYDMAN B., « Comment penser le droit global? », in CHEROT J.-Y., FRYDMAN B. (dir.), *La science du droit dans la globalisation*, op. cit., p. 18.

d'assurer une cohérence maximale entre les acteurs internationaux et de préconiser les principes éthiques de l'Union dans le monde entier<sup>537</sup>.

Dans cette perspective, en France, au Sénat, la commission d'enquête sur la souveraineté numérique a été créée en 2019 pour « définir les champs fondamentaux de notre souveraineté, individuelle ou collective » car les droits des individus sont affectés par l'essor de l'intelligence artificielle<sup>538</sup>. Cependant, l'on voit qu'aujourd'hui, la souveraineté au sens traditionnel du pouvoir de légiférer, n'est en effet plus opportune ni pertinente dans le contexte numérique correspondant à ce nouvel espace<sup>539</sup>. À vrai dire, l'idée de réglementer à l'intérieur de frontières nationales est une mentalité dépassée du siècle dernier<sup>540</sup>. Car si des mécanismes traditionnels de renvoi aux droits nationaux ou territoriaux peuvent être envisagés, cela risque d'entraver le droit au procès équitable en cas de lacune des dispositifs dans des pays relativement moins compétents et équipés en matière d'IA. Ces pays sont d'ailleurs les plus vulnérables face aux géants du numérique des pays développés disposant de ressources et de moyens les plus performants, notamment les dominants transnationaux. Mais pour rendre une transition pacifique, le droit contenant l'IAG devrait commencer par protéger les fragilités de l'espèce humaine de nature transfrontière<sup>541</sup>.

**189. En résumé.** — L'IAG fait partie d'une évolution dont les enjeux sont partagés par tout le monde. De l'internationalisation à la digitalisation, au sens de cette étude, le spectre territorial peut être de plus en plus inefficace au XXIème siècle, voire inapproprié au regard des enjeux transnationaux voire cosmiques d'IAG. C'est ainsi que la méthode de méta-régulation extraterritoriale s'implante d'ores et déjà dans la présente civilisation.

## **§2 La méta-régulation**

**190. L'exportation *de facto* et *de jure*.** — À l'heure actuelle, les interactions entre ensembles normatifs nationaux, régionaux et internationaux produisent un effet de

---

<sup>537</sup> Résolution préc. du 12 févr. 2019, pts 181-183, 190-192.

<sup>538</sup> Proposition de résolution tendant à la création d'une commission d'enquête sur la souveraineté numérique, n° 383, enregistré à la Présidence du Sénat le 19 mars 2019.

<sup>539</sup> Stéphane Mortier, au colloque « Intelligence artificielle et Cyber-sécurité : Perception Juridique », l'ESLSCA Business School Paris, le 25 septembre 2019.

<sup>540</sup> BORREL PH., « Un monde sans humains? », documentaire, Arte France, CinéTévé, 2012.

<sup>541</sup> V. infra La disparité juridique inter-espèce.

standardisation qui facilite la circulation ultérieure des règles, et le standard peut devenir une référence, voire un modèle<sup>542</sup>. À titre d'exemple, la Commission européenne a adopté une décision d'adéquation en faveur du Japon en 2019, en reconnaissant que ce pays, grâce à sa législation, offre un niveau de protection des données à caractère personnel comparable à celui qui est garanti dans l'UE. L'adoption d'une législation en miroir répondant à une loi à portée extraterritoriale par une loi analogue, place les deux parties sur un pied d'égalité et permet à chacune d'apparaître comme un *forum conveniens* aux yeux de l'autre<sup>543</sup>. Ce « niveau comparable » qui est d'abord « exporté », ensuite « adopté », et au final « reconnu », ne se limite pas à la législation nationale des dispositifs en vigueur. Rappelons qu'il en va de même pour les normes autorégulatrices issues des entités privées, comme l'exemple d'iRobot que l'on vient d'évoquer plus haut : son autorégulation, en adoptant le RGPD non seulement comme exigé, mais ils ont en outre veillé à ce que les images ne quittent pas le robot, de sorte que le traitement des données des capteurs visuels se fait localement sur le robot et que seule la connaissance sémantique de la maison avec le consentement du consommateur est envoyée ; ce qui va au-delà du seuil du RGPD, mais pas en deçà.

Dorénavant, le RGPD devient ainsi une référence en dehors des frontières de l'UE : « Les normes du RGPD sont reprises volontairement par certains acteurs privés, ce qui a pour conséquence d'étendre de facto leur champ d'application »<sup>544</sup>. Ces évolutions nous montrent que l'application extraterritoriale, une fois intervenue à titre exceptionnel, se constitue en tant que principe, relevant désormais du caractère « auto-conformé ». Dans le contexte d'une globalisation du champ de la régulation de l'IA, des normes régionales peuvent connaître

---

<sup>542</sup> FULCHIRON H., « Interactions entre systèmes ou ensembles normatifs et 'dynamique des normes' », RTD civ. 2017, p. 271 s.

<sup>543</sup> NIDOR V., « L'extraterritorialité des législations relatives à la protection des données personnelles », et ASCENSIO H., « Conclusions », in MIRON A., TAXIL B., (dir.), *Extraterritorialité et droit international*, colloque d'Angers, SFDI, Pedone, 2020, pp. 300 et 355.

<sup>544</sup> SEE A., « La régulation des algorithmes : un nouveau modèle de globalisation ? », RFDA, 2019, p. 830 s. V. aussi p. ex. l'effet de Bruxelles, BRADFORD A., « Exporting Standards: The Externalization of the EU's Regulatory Power via Markets », *International Review of Law and Economics*, vol. 42, 2015, p. 159 : les règles et réglementations émanant de Bruxelles ont pénétré de nombreux aspects de la vie économique à l'intérieur et à l'extérieur de l'Europe par le biais du processus de « mondialisation réglementaire unilatérale (unilateral regulatory globalization) ». La mondialisation réglementaire unilatérale a lieu lorsqu'un seul État est en mesure d'externaliser ses lois et réglementations en dehors de ses frontières par le biais de mécanismes de marché, ce qui entraîne la mondialisation des normes. Ce processus peut être distingué de la mondialisation politique des normes réglementaires, où la convergence réglementaire résulte de normes négociées, notamment de traités ou d'accords internationaux entre autorités réglementaires. Il se distingue également de la coercition unilatérale, où une juridiction impose ses règles aux autres par des menaces ou des sanctions.

une application globale : l'exemple du RGPD devient un instrument universel dès qu'il connaît une application mondialisée.

En matière d'IA, de manière encore plus explicite, le Parlement européen constate que « l'industrie européenne pourrait se voir avantagée par une démarche réglementaire efficace, cohérente et transparente au niveau de l'Union (...) tout en veillant à ce que l'Union et ses États membres gardent le contrôle sur les normes réglementaires à définir, pour ne pas avoir à adopter et à vivre avec des normes définies par d'autres, c'est-à-dire des pays tiers qui sont eux aussi à la pointe du développement de la robotique et de l'intelligence artificielle »<sup>545</sup>. Cette remarque nous montre combien il est important pour certains normateurs de « préconiser les principes éthiques de l'Union dans le monde entier »<sup>546</sup> afin de ne pas « à adopter et à vivre avec des normes définies par d'autres ». Mais à vrai dire, cette exportation juridique n'a rien de nouveau. Rappelons qu'au départ, la paix de Westphalie était aussi un système régional, qui a été progressivement étendu à d'autres systèmes juridiques. Car c'est à partir de cette paix qui met fin à la guerre de Trente Ans qu'émerge l'Europe moderne, celle des États-nations ; une nouvelle organisation des relations internationales apparaît : le système moderne des relations internationales<sup>547</sup>. Devenu le modèle de principe jusqu'à présent, il s'agit en effet d'un renversement universellement applicable qui manque de son attribut particulariste initial. La paix Westphalienne constitue ainsi un autre exemple de l'exception évoluant en un nouveau socle commun de principe. Dès lors, il convient de noter qu'une application extraterritoriale des normes de police robotique — le durcissement de la méta-régulation — pourrait éventuellement se voir généralisée.

Dorénavant, l'on se trouve dans une ère où les robots négocient avec d'autres robots<sup>548</sup>. Donc il est inéluctable que bientôt, les interactions entre humains figureront de moins en moins des conversations qui se déroulent dans le cyberspace. La question de réinventer le droit à

---

<sup>545</sup> Résolution préc. du 16 févr. 2017, pt S.

<sup>546</sup> V. supra Résolution préc. du 12 févr. 2019.

<sup>547</sup> Fiche repère « L'ordre westphalien », Geolink, Observatoire en géostratégie de Lyon, 20 déc. 2014.

<sup>548</sup> V. l'ébauche p. ex., le bot DoNotPay ChatGPT dialogue avec le Chat du service client et obtient une réduction sur la facture d'un client. DoNotPay propose déjà de nombreux services qui permettent aux clients d'obtenir, notamment des aides pour les dépôts de plaintes, des annulations d'abonnement ou encore une contestation de contravention pour stationnements illicites. Adde Cicero. The Meta Fundamental AI Research Diplomacy Team (FAIR) et *alii*, « Human-level play in the game of Diplomacy by combining language models with strategic reasoning », *Science*, vol. 378, issue 6624, 22 nov. 2022.

partir de ces nouveaux arrivants communiquant instantanément et travaillent ensemble incessamment comme des abeilles dans une ruche, à la fois transfrontières et numérisés, devient non seulement cruciale, mais aussi pressante. Dès lors, l'on pourrait s'attendre qu'une telle réponse incombe aux méta-régulateurs vu la configuration juridique jusqu'à présent. Mais l'on se rend compte que dès qu'ils tranchent la question, il y constitue sous peu soit une application extraterritoriale directe et explicite, soit une auto-adoption indirecte et implicite. Par conséquent, l'approche exclusive fondant sur le principe territorial ne serait pas considérée tenable dans un tel nouveau contexte. Dans un premier temps, ces interactions d'IA se trouvent dans le cyberspace, dont accessibilité est par principe indistinctement omniprésents. Et puis dans un second temps, dès qu'une réponse juridique aurait été proposée, elle implique nécessairement une approche plurale et inclusive : mon bot enregistré à un organe tiers parle avec votre bot enregistré dans un organe quart. Et au moment où l'on considère que ces relations juridiques se nouent en effet dans un espace commun, une compétence inclusive ne devrait pas être exclue.

**191. En résumé.** — Les méta-régulateurs exportaient constamment les modèles voire les systèmes juridiques, que ce soit de manière directe ou indirecte, explicite ou implicite. Si la révolution industrielle a consolidé cette appréhension juridique dite européenne, il convient de s'interroger sur la trajectoire d'un tel récit au cours de la révolution numérique. En tout état de cause, que ce soit par la révolution industrielle ou la révolution numérique, compte tenu du principe désormais d'extraterritorialité, il convient de reconnaître parallèlement une mentalité inclusive *de facto* et *de jure* au sein de la civilisation humaine.

### **Conclusion de la section 1**

**192. De méta-régulation à l'influence réelle.** — Non seulement le champ matériel du droit s'évolue au XXIème siècle, mais le champ spatial ainsi que les acteurs concernés sont aussi en pleine transition. Volens nolens, ce monde interconnecté renverse l'approche du principe de territorialité. En dégagant le principe d'extraterritorialité notamment provenant de méta-régulateur, c'est-à-dire le régulateur des régulateurs, l'on a démontré que si le même récit du droit international inchangé s'applique, l'on doit s'attendre à ce que les acteurs principaux en matière d'IAG, à l'issue de cette révolution numérique du XXIème siècle, dictent le droit dans l'espace intra et extra-atmosphérique<sup>549</sup>. Car l'espace commun où se trouvent les bots autonomes cherche désormais une réponse inclusive sur le sujet. Tirant ces conclusions, la présente étude se propose désormais d'étudier la compétence juridique en matière d'IAG.

---

<sup>549</sup> Que feront les IA super-intelligentes ? L'espace extra-atmosphérique est hostile aux humains mais bienveillant pour les robots conçus de manière appropriée. La plupart des IA seront plus intéressées par les nouvelles opportunités pour les robots et la vie logicielle dans l'espace extra-atmosphérique. Grâce à d'innombrables usines de robots auto-réplicateurs, la sphère en expansion de l'IA aura tout le temps de coloniser et de façonner l'ensemble du cosmos. SCHMIDHUBER J., « Annotated History of Modern AI and Deep Learning », AI Blog, 12 déc. 2022.

## **Section 2 La compétence en matière d'IAG**

**193. Annonce du plan.** — Compte tenu du principe d'extraterritorialité que l'on vient d'évoquer, l'on voit que désormais c'est l'influence réelle transfrontière (§1) qui est plus pertinent pour s'identifier comme facteur de rattachement. En effet, l'influence transfrontière de l'IAG nous ramène à étudier une compétence inclusive d'IAG (§2).

### **§1 Le critère d'influence réelle**

**194. Annonce du plan.** — Avant d'étudier les exemples concrets à ce nouveau critère d'influence réelle transfrontière (II), il convient d'exposer schématiquement quelques observations sur l'évolution d'un tel développement (I).

#### **I. La consécration de l'influence réelle transfrontière**

**195. Le détachement des conceptualisations classiques.** — À l'heure actuelle, il devrait exister un lien de rattachement entre le forum susceptible d'exercer sa compétence et la situation qu'il souhaite réguler. Ce lien est en premier lieu territorial (la situation se déroule sur le territoire de l'État) ou fondé sur la nationalité (la situation implique un ressortissant de l'État)<sup>550</sup>. Dans cette conception traditionnelle du conflit des lois et de compétence juridictionnelle, les facteurs de rattachement établissent le lien avec la compétence du for ainsi que l'applicabilité des règles sont en principe de considération territorial-national<sup>551</sup>. En termes de la règle de conflit adoptée, le conflit des lois ne résulte par nature de rien d'autre que de la concurrence d'applicabilité de normes, et la loi désignée est en général celle de l'État avec lequel « la situation présente des liens plus étroits ou significatifs »<sup>552</sup>. À titre

---

<sup>550</sup> Il peut également résulter d'autres hypothèses comme atteinte aux intérêts fondamentaux de l'État, cas de compétence universelle etc. BISMUTH R., « Pour une appréhension nuancée de l'extraterritorialité du droit américain : Quelques réflexions autour des procédures et sanctions visant Alstom et BNP Paribas », *Annuaire français de droit international*, 2015, p. 790.

<sup>551</sup> AUDIT B., D'AVOUT L., *Droit international privé*, LGDJ, 2022, p. 40 : « De fait, le droit international public reconnaît simultanément aux États une compétence 'personnelle', les autorisant à légiférer pour leurs nationaux, et une compétence territoriale, qui leur permet d'édicter des normes applicables à toutes les personnes sur leur territoire ».

<sup>552</sup> *Ibidem*, p. 22.

d'exemple, la méthode classique savinienne appelle à la « nature des choses » comme justification fondamentale des règles de conflits de lois<sup>553</sup> ; la nature du rapport du droit lui assigne dans l'espace une situation, qui ainsi le rattache à un ordre juridique. Pour inscrire ce rapport de droit dans l'espace, pour l'assigner à un ordre juridique, il faut supposer que cet exercice du pouvoir est localisable. Pour autant, ce critère de l'effet ou de l'ordre juridique affecté<sup>554</sup> n'en reste pas moins attaché aux grandes catégories de rattachement telles que personnes, choses, obligations<sup>555</sup> ; et son aiguillage n'échappe pas, jusqu'à présent, à une prémisse territoriale-nationale en aval.

Néanmoins, d'une part, à l'instar de la mondialisation et de la digitalisation, déjà, l'on pourrait remarquer que le repère territorial-national n'est plus autant absolu. D'autre part, « la méthode des conflits de lois est altérée par de nouveaux procédés qui se développent en son sein, telles les règles de conflit à coloration matérielle ou les règles de conflit unilatérales, voire refoulée dans certains domaines par des procédés concurrents, telles les lois de police ou les règles matérielles de droit international »<sup>556</sup>. De même, la révolution américaine néostatutaire montre que cette démarche unilatéraliste cherche désormais, davantage son fondement sur des critères divers, tels que le critère d'application fondé sur l'utilisation d'une monnaie particulier<sup>557</sup>, pour revendiquer le rattachement transfrontière.

Dans un premier temps, auparavant, le fondement opératoire en amont des grands piliers classiques tels que le statut territorial ou le statut personnel, dorénavant, étant donné l'applicabilité extraterritoriale et l'invocabilité supra-régionale des normes spécialisées dans des domaines spécifiques, sont susceptibles d'être marginalisés voire privés de sens justificatif. Et puis dans un second temps, les facteurs de rattachements en aval deviennent de plus en plus divers, et le champ d'application des règles issues de différents régimes juridiques s'en trouve à la fois élargi et diversifié. Par exemple, dans le cadre d'enquêtes criminelles, en matière de données électronique, la loi américaine sur la surveillance des

---

<sup>553</sup> BATIFFOL H., *Aspects philosophiques du droit international privé*, Dalloz, 2002, p. 163.

<sup>554</sup> ANCEL B., *Éléments d'histoire du droit international privé*, Panthéon-Assas, 2017, p. 484-485.

<sup>555</sup> LEQUETTE Y., *Les mutations du droit international privé : vers un changement de paradigme ?*, Cours général de droit international privé, 2015, p. 88.

<sup>556</sup> *Ibidem*, p. 24.

<sup>557</sup> L'utilisation de cette devise crée un lien de rattachement. AUDIT M., PATAUT É., « Ouverture », in AUDIT M., PATAUT É (dir.), *L'extraterritorialité*, Actes des journées doctorales du 27 juin 2018, Pedone, 2020, p. 9.

données personnelles (Clarifying Lawful Overseas Use of Data Act, CLOUD Act) permet aux autorités de cet État de contraindre les fournisseurs américains de fournir les données demandées stockées sur des serveurs, qu'ils soient situés aux États-Unis ou dans des pays étrangers, et peu important que lesdites données concernent des personnes non-américains ou domiciliées en dehors du territoire des États-Unis<sup>558</sup>. Ainsi, l'on pourrait remarquer qu'en présence des sujets, objets ainsi que des litiges du caractère transfrontière, les facteurs classiques s'appuyant sur les conceptions dichotomiques de jus soli et jus sanguinis ou de statut personnel et statut réel pour l'aiguillage juridique présentent des difficultés au XXIème siècle. Car si la régulation de données adopte déjà une approche extraterritoriale, l'on pourrait s'attendre que la digitalisation de la civilisation rendrait les approches traditionnelles obsolètes voire inopérantes par principe.

Partant, l'intérêt de l'approche inclusive ne devrait pas être méconnu, étant donné que les enjeux juridiques de l'IAG qui se posent en dehors de la frontière nous affectant tous de la même manière. Ainsi, le concept d'extranéité au sens traditionnel devrait être atténué et neutralisé en raison de sa difficulté à opérer de manière cohérente et pertinente dans un tel contexte<sup>559</sup>. Étant donné qu'une quelconque IAG qui n'a rien de lien étroit au sens traditionnel avec un quelconque for pourrait tout de même présenter des risques existentiels à un quelconque sujet dans un quelconque ordre juridique. Notamment pour adresser le risque existentiel de l'humanité, les éléments classiques des facteurs de rattachement ne seraient plus que subsidiaires afin de reconnaître l'ensemble des faits impliquant l'IAG cristallisés sur cette planète et au-delà. Autrement dit, si la civilisation humaine est, in toto, directement concerné et impacté par la création d'un très petit nombre d'acteurs, il convient alors de commencer à envisager de nouveaux critères qui arrivent à établir des pertinences effectives pour l'intervention juridique. Concrètement, une IAG développée par l'entreprise américaine OpenAI pourrait très bien mettre en danger un Togolais ayant son domicile au Togo. Ce dernier devrait pouvoir exiger, en vertu du principe de faveur pour l'autorégulation, la pratique de précaution prévue par la Charte d'OpenAI. Or, les lois régissant les relations juridiques selon la classification traditionnelle de nationalité ou de domicile vont résoudre la concurrence des normes par la désignation des lois applicables en fonction du concept

---

<sup>558</sup> *Ibidem*, p. 13.

<sup>559</sup> V. infra La disparité juridique inter-espèces.

dichotomique de national-territorial. Pour autant, l'on voit que dans notre hypothèse, la nationalité ainsi que le domicile, qui seraient des critères pertinents sous l'angle des opérations classiques, deviennent indésirables. Car même si la règle de conflits désigne la compétence juridictionnelle au domicile du défendeur, ce déplacement physique de notre demandeur togolais peut ne sembler pas lui convenir. En revanche, si la compétence juridictionnelle relève du juge togolais, l'on ne peut pas être certain que ce ne serait pas la loi du for qui s'applique éventuellement à la demande. Dans ce cas, il serait difficile d'être convaincu qu'un jugement adéquat ou même un résultat minime de l'application de la loi togolaise (qui cherche par exemple à promouvoir le développement de l'IA) serait efficace à l'égard de la conformité à la précaution que notre Togolais soulève.

Dès lors, il devrait être plus convaincant de dire que l'influence réelle de l'IAG en cause mérite d'être considérée comme la nouvelle pertinence dans la coordination juridique. Dans ce sens, l'on se propose de faire un petit tour transnational afin d'étudier cette concrétisation du critère d'influence réelle transfrontière au début du XXI<sup>ème</sup> siècle.

## II. La concrétisation du critère d'influence réelle transfrontière

**196. L'illustration d'influence réelle.** — À l'instar des exemples du RGPD et de CLOUD Act<sup>560</sup>, l'on pourrait déjà constater que sous l'empire digitale du XXI<sup>ème</sup> siècle, les formules « ou à l'extérieur » d'une frontière dite national, et « toute autre personne » que les sujets dits nationaux, une fois appréhendées à titre exceptionnel, se généralise, avant qu'elles ne deviennent le nouveau principe.

Dans cette perspective, l'on pourrait encore citer le droit allemand ayant adopté la loi sur les réseaux sociaux (Netzdurchsetzungsgesetz, NetzDG) en ce sens qu'il oblige les plateformes à retirer certains contenus illégaux en cas d'atteinte à la vie privée telle que la violation de la vie personnelle par l'enregistrements d'images<sup>561</sup>. Ce qui attire notre attention est la désignation de son champ d'application prévoyant notamment que le fournisseur d'un réseau social est exonéré de certaines obligations si le réseau social compte moins de deux millions d'utilisateurs enregistrés en Allemagne<sup>562</sup>. Ce critère dit mixte, qui tient compte à la fois de

---

<sup>560</sup> V. supra sur l'exportation et l'extraterritorialité des normes.

<sup>561</sup> §1 (3) Netzdurchsetzungsgesetz (NetzDG), 1 sept. 2017.

<sup>562</sup> §1 (2) Netzdurchsetzungsgesetz (NetzDG), 1 sept. 2017.

la présence territoriale et de l'absence territoriale, illustre que la loi ne se soucie ni des nationaux allemands, ni des opérateurs sur son territoire, ni des actes passés à l'intérieur de sa frontière mais tout simplement, du nombre des comptes enregistrés sur son territoire. Cela nous ramène donc au critère d'influence réelle à travers une règle mixte pour réclamer son application transfrontière. D'une part, l'influence réelle nous montre que le cible de la pertinence vise directement les fournisseurs qui sont réellement susceptibles d'avoir un impact. D'autre part, l'on peut considérer que c'est une règle mixte car même si les critères traditionnels (ni des nationaux allemands, ni des opérateurs sur son territoire) ont été abandonnés, l'on pourrait néanmoins observer une persistance résiduelle du concept spatial au sens étatique : comptes enregistrés en Allemagne.

Mais pour aller plus loin, cette loi prévoit en outre que s'agissant d'obligation de rapport, les fournisseurs de réseaux sociaux qui reçoivent plus de 100 plaintes par an concernant des contenus illicites sont tenus de produire des rapports semestriels en langue allemande sur le traitement des plaintes ayant trait à des contenus illicites sur leurs plateformes<sup>563</sup>. Encore une fois, comme ce critère ne prend pas en compte ni la condition de nationalité, ni celle de territoire, l'on peut imaginer une hypothèse où toutes les plaintes viennent d'un pays tiers, mais le fournisseur étranger est tout de même tenu de se conformer au rapport de justification du droit allemand. À ce stade, même la dernière trace de caractère traditionnel - les comptes enregistrés en Allemagne - bien que préservée dans le cadre que l'on pourrait encore observer dans la règle dite mixte ci-dessus susmentionnée, est inévitablement neutralisée par le mélange indiscriminé de plaintes.

Dans ce sens, au regard des enjeux de l'IAG, l'on pourrait retracer l'exemple de notre Togolais ci-dessus pour s'attendre que OpenAI aurait lieu de se conformer à l'exigence du droit allemand simplement du fait de « demandes » togolaises. *Pirna facie* cette solution pourrait paraître excessivement déplacée sous l'angle de la logique classique. Néanmoins, à la lumière des risques sous-jacents de l'IAG<sup>564</sup>, ce choix de l'humanité pourrait en fait

---

<sup>563</sup> §2 (1) *Netzdurchsetzungsgesetz* (*NetzDG*), 1 sept. 2017.

<sup>564</sup> V. supra Les enjeux de l'IAG.

renforcer le principe de précaution<sup>565</sup>, qui est avant tout indispensable pour la coévolution pacifique de l'espèce humaine avec l'IAG.

De la même manière, dans le projet de la loi taïwanaise<sup>566</sup> sur les services intermédiaires numériques, le champ d'application territorial prévu n'est pas moins extensif. En effet, dans un premier temps, l'on garde un critère territorial en disposant que la régulation vise les fournisseurs de services dont le lieu d'établissement se situe à Taïwan. Mais dans un second temps, même en l'absence d'établissement des fournisseurs de ces services sur le sol taïwanais, la loi s'applique « dès qu'il y a des faits justifiant le lien substantiel avec notre pays (有事實足認與我國實質關聯) ». Ensuite, le prochain alinéa prescrit que ce lien substantiel se fonde sur deux critères non cumulatifs : soit par « un nombre important d'utilisateurs » à Taïwan, soit par « le ciblage des activités » vers Taïwan. À la lecture de ces critères, l'on voit qu'encore une fois, ni les nationaux taïwanais, ni les domiciles taïwanais qui implique un rattachement permanent sur le sol ne sont suffisants ; en revanche, c'est l'influence réelle de ces opérateurs qui se mêle effectivement dans une situation concernée qui compte.

Détachant au fond la *fictio legis* opérée sur des statuts auparavant utiles et pertinents (nationalité et territoire), les nouveaux rapports, catégories ainsi que leurs éléments d'aiguillage qui déterminent l'application éventuelle des règles nous montrent que le vrai facteur de rattachement est la pertinence ponctuelle de facto seule. Tout comme dans les époques dépourvues de la mondialisation et de la digitalisation, les facteurs de rattachement traditionnels demeurent et continuent de servir dans des contextes classiques, mais sont nécessairement atténués voire complètement abandonnés dans des circonstances qui ne peuvent plus être mesurées à l'aune des fondements antérieurs. Ces approches concrètes et ponctuelles nous montrent que l'application extraterritoriale d'une loi purement interne est de plus en plus consolidée voire généralisée sur le fondement notamment de l'influence réelle transfrontière.

---

<sup>565</sup> V. supra Le principe de précaution.

<sup>566</sup> L'art. 4 al. 1 et 2. 數位中介服務法草案總說明, version du 29 juin 2022. L'exposé des motifs de cet article a essentiellement calqué sur les considérations (7) et (8) de la proposition du règlement européen relatif à un marché intérieur des services numériques (Digital Service Act) du 15 décembre 2020.

**197. Influence réelle en matière d'IAG.** — Dans ces conditions, l'on voit que l'applicabilité d'une loi ne préconise pas de facteurs de rattachement classiques tels que le statut territorial ou le statut personnel, mais simplement renvoie au critère d'influence réelle. En effet, sous la pression d'un droit de plus en plus spécialisé et diversifié, la méthode conflictualiste de principe pourrait être marginalisée voire remplacée par d'autres désignations, sur d'autres fondements, intervenant directement dans des nouveaux contextes plurilatéraux, transfrontières, et digitalisés<sup>567</sup>. En d'autres termes, l'ère numérique reconnaît davantage l'effet direct d'une situation juridique, ce qui conduit inévitablement à l'application extraterritoriale de principe par voie de l'unilatéralisme dans la forme et le plurilatéralisme au fond.

Comme l'on vient de voir, le critère du droit allemand pour gérer l'activité numérique ne tient pas compte des utilisateurs de nationalité allemande, ni des utilisateurs se trouvant sur le territoire allemand, ni des opérateurs selon ces deux facteurs, mais simplement de l'influence réelle de la situation en cause : en l'espèce, les plaintes reçues de nature transnationale et de manière transfrontière. Dans la mesure où la désignation de la loi applicable est aiguillée par des critères mixtes respectifs, à savoir la localisation de l'influence substantielle, l'on pourrait s'attendre une transition juridique à l'ère numérique. Au sens de cette étude, le résultat final sera généralisé et puis achevé par l'élément de « portée territoriale » orientée vers « l'influence réelle », comme la portée numérique concernant la question juridique évolutive en cause. Pour le moment, un tel aiguillage reste essentiellement national, régional, et ne constitue pas une solution in fine extranationale. Pourtant, les enjeux concernant l'IAG sont de nature cosmique<sup>568</sup> dont les standards ne peuvent plus se référer à aucun contexte antérieur car il serait non pertinent. À notre avis, un nouveau repère juridique est nécessaire pour ce siècle.

---

<sup>567</sup> Ce pluralisme des méthodes existe déjà, indépendamment du contexte de la digitalisation. L'on constate ainsi que la question de sélectionisme-substantivisme, conflictualiste-statutiste, voire unilatéraliste-bilatéraliste ressemble au fond à la question de positiviste-naturaliste : Les Aveugles et l'Éléphant. Dès lors, la question d'applicabilité, que ce soit de nature inter-coutumière ou non, son contenu du raisonnement s'avère toujours valable sur le plan théorique ; en revanche, sa mise en place dans la pratique par le modèle humain, qui opte jusqu'à présent en un droit dit international — qui lui-même est, avant tout, se fonde sur une conceptualisation nationale-territoriale — conduit à un droit éventuellement contre même la possibilité de l'humanité, y compris sa culture, au nom de sa culture, installant en conséquence des bilans intenable pour les générations actuelles et futures.

<sup>568</sup> V. supra la Singularité de John von Neumann et l'hypothèse d'explosion d'intelligence de I. J. Good.

Ainsi, l'on peut s'attendre à ce qu'une IAG soit soumise à une corégulation de deux niveaux : d'abord les mesures prises par la police d'autorégulation, et ensuite, les règles extraterritoriales statuant sur les facteurs pertinents. Prenons l'exemple d'une IAG développée par Meta. L'activité est, dans un premier temps, soumise aux règles de son autorégulation, par exemple, des décisions issues de son Conseil de surveillance (Oversight Board). Ces décisions dites jurisprudentielles, en vertu du principe de faveur pour l'autorégulation, sont non seulement « en vigueur » au regard du développement de l'IA au sein du groupement : ainsi ils comprennent les cas pour WhatsApp, Instagram, Messenger, Oculus et d'autres filiales de Meta ; mais aussi accessibles pour d'autres entités de recherche et développement de l'IA. Puis dans un second temps, l'on peut imaginer que, tout comme l'approche allemande que l'on vient de voir, le droit taïwanais pourrait unilatéralement prévoir le facteur de rattachement en matière d'IAG en renvoyant le critère à une pertinence de l'influence réelle transfrontière. Concrètement, le droit taïwanais pourrait disposer que sa loi sur l'IAG soit applicable lorsque 1.) le fournisseur de l'IAG dispose d'un conseil de surveillance ou/et un plan de vigilance; 2.) ou/et que l'IAG en cause est immatriculée au registre de l'une des organes du Formosa Club<sup>569</sup> ; 3.) ou/et cette IAG a été active/mise à disposition pour plus de 24 heures ; 4.) ou/et l'IAG envisagée est une IA bénéfique au sens de la Charte d'OpenAI ainsi que son autorégulation étendue ; 5.) ou/et l'IAG en cause présente une menace réelle et imminente pour l'humanité. Bref, la liste non exhaustive a pour but de démontrer que la compétence en matière d'IAG doit être déterminée par des facteurs de rattachement véritablement pertinents, ce qui signifie un principe d'extraterritorialité se fondant sur l'influence réelle. Dès lors, pour répondre à un tel repère juridique, une compétence dite inclusive nous semble inéluctable. C'est ce que l'on va ensuite étudier dans notre prochain paragraphe.

**198. En résumé : la proximité réévaluée.** — L'approche territoriale pourrait laisser croire qu'il ne s'agit que de problèmes spécifiques confinés dans des frontières nationales et que ceux-ci n'ont d'importance que pour les sujets immédiatement concernés. Cette croyance s'est toujours révélée fautive à notre époque. Il convient de considérer comme une règle

---

<sup>569</sup> Lancé en 2019, l'organisme international qui compte en 2021 plus de 1500 membres parlementaires couvrant six continents, est chargé de coordonner et de renforcer les organisations d'amitié interparlementaire avec Taïwan.

générale que tout ce qui est possible dans un pays peut, dans un avenir proche, l'être également dans presque tous les autres. Ainsi, les dispositions de la régulation d'IAG doivent chercher « l'influence réelle transfrontière » comme étant le critère convenable pour déterminer l'aiguillage. D'où l'intérêt d'une approche d'inclusivisme d'IAG recueillant la compétence pertinente.

## **§2 Le critère de la compétence inclusive de l'IAG**

**199. Annonce du plan.** — En transposant le critère de l'influence réelle, cette démarche va aboutir à la compétence inclusive de l'IAG (I). Son appréciation (II) nous permet de proposer par voie de conséquence un nouveau paradigme du droit d'IAG.

### **I. La consécration de la compétence inclusive de l'IAG**

**200. L'inclusivité de la compétence sans proximité avec le for.** — Le critère de l'influence réelle évoqué ci-dessus nous montre l'intérêt d'un renvoi des normes transnationales les plus pertinentes aux autres sur le fondement de l'influence réelle de l'IAG : il est souhaitable que d'une part, le champ d'application de la protection juridique puisse être élargi à la fois territorialement, par une application en dehors du territoire du normateur initial ; d'autre part, en ne se bornant pas à un champ restreint, qui ne concernerait que le domaine sectoriel : le champ d'application matériel devrait englober tous les stades du développement, du déploiement et de l'utilisation de l'IAG. Le facteur de l'influence réelle transfrontière pourrait désigner le renvoi à un for plus compétent et à une *lex fori* plus adaptée pour un raisonnement plus convenable en la matière. Ainsi, le facteur de rattachement en matière d'IAG pour désigner la règle éventuellement applicable devrait être l'unique lien avec l'IAG en cause<sup>570</sup>. L'attachement à l'IAG serait la condition pour déclencher la compétence inclusive de l'IAG. En effet, dès qu'il s'agit d'un fait d'IAG, et quelle que soit la proximité selon la conception traditionnelle avec le territoire antérieur ou le corps générateur des règles applicables, la seule exigence à remplir pour l'applicabilité et l'invocabilité des règles concernées, est cette proximité avec l'IAG, à laquelle la compétence inclusive de l'IAG renvoie<sup>571</sup>. Autrement dit, la *ratione materiae* désigne à la fois son champ d'application

---

<sup>570</sup> V. infra la composition de la compétence inclusive de l'IAG.

<sup>571</sup> *Ibidem*.

matériel (la matière d'IAG), et le facteur de rattachement (la proximité avec l'IAG en cause). La compétence inclusive de l'IAG pour traiter les questions juridiques en la matière est déclenchée dès que ces conditions sont réunies.

Partant, la formule concernant l'applicabilité des règles d'IAG devrait être : « la loi applicable est celle du système juridique matériellement compétent »<sup>572</sup>. Cela renvoie donc à une conception dite de l'inclusivisme<sup>573</sup>. Cette approche, qui ne cherche plus la loi qui présente des liens classiquement plus étroits avec le for territorial, s'appuie sur la compétence du for sur le fondement de la proximité matérielle avec l'IAG. La différence entre la proximité matérielle et les liens les plus étroits classiques consiste dans un raisonnement inverse : auparavant, selon la logique classique, l'on part de la proximité physiquement territoriale et de la proximité biologiquement nationale pour fonder les liens les plus étroits pour justifier les facteurs de rattachement respectifs. En revanche, la proximité matérielle identifie la compétence en termes du traitement juridique efficace et adéquat pour établir le facteur pertinent<sup>574</sup>. Autrement dit, tandis que le fondement du lien classique se justifie par une proximité du territoire et de la nationalité, le fondement du lien IAG ne part plus de ces pertinences antérieurement valables. Reste à rechercher la composition de cette compétence inclusive.

## II. La composition de la compétence inclusive de l'IAG

**201. Les deux étapes du mécanisme.** — La compétence inclusive de l'IAG consiste dans une analyse en deux étapes : ses facteurs de rattachement formel (A) et puis substantiel (B). Il convient d'apprécier ces conditions consistant à déclencher la compétence d'IAG (C).

---

<sup>572</sup> *Ibid.*

<sup>573</sup> Cette démarche trouve l'écho dans l'idée d'un for de nécessité, voire la reconnaissance d'une compétence universelle résultant du risque de déni de justice, qui peut donc être retenue en tant qu'instrument de protection du droit fondamental d'accès au juge. USUNIER L., « La multilatéralisation du procès international : le point de vue privatiste », in RUNAVOT M.-C. et RIOS RODRIGUEZ J. (dir.), *Le droit international multilatéral*, *op. cit.*, pp. 223-224.

<sup>574</sup> Prima facie cela pourrait faire penser immédiatement à « result selecting rules » de David Cavers ou à « better rule of law » de Robert Leflar, voire les règles de conflit à coloration matérielle. Pourant, si l'on admet la diversification dans les conflits de lois, cette agglomération des méthodes renvoie au phénomène « Les Aveugles et l'Éléphant » discuté plus haut. En outre, la discussion autour de ces alternatifs sélectionnisme-substantivisme n'échappe pas le cadre sous-jacent d'un conflit inter-coutumier, et ne s'appuie pas sur le mécanisme d'un droit inter-espèces évoqué ci-dessous, ce qui la distingue fondamentalement de la présente étude. CAVERS D., « A Critique of the Choice of Law Problem », 47 *Harv. Law Review*. 173, 1933; LEFLAR R., « Choice-Influencing Consideration in Conflicts Law », 41 *N.Y.U. L. Rev.* 267, 1966.

## A. Le facteur formel de la compétence inclusive

**202. Le facteur formel : l'application *expressis verbis*.** — Le régime de l'IAG peut s'entendre de l'ensemble des règles relatives aux rapports d'IAG pour sa qualification juridique. Chaque système juridique — à l'heure actuelle essentiellement étatique ou supranational — se voit doter d'une marge de manœuvre absolue dans l'élaboration des normes d'IAG. Une fois ces normes mises en place, le législateur prévoit explicitement qu'il s'agit d'une norme d'IAG. La nature du rapport de droit est ainsi précisée<sup>575</sup>. La volonté manifestée par le législateur est la première indication à relever pour l'applicabilité inclusive des règles d'IAG<sup>576</sup>.

## B. Le facteur substantiel de la compétence inclusive

**203. À défaut du facteur formel.** — Dans l'hypothèse où le législateur n'exprime aucun souhait explicite pour dire que le texte relève d'une norme d'IAG, ou ne manifeste aucune volonté pour une applicabilité inclusive du texte, il y aurait lieu de mettre en place une appréciation dite substantielle de la compétence inclusive de l'IAG. En d'autres termes, en l'absence de l'application directe provenant de la revendication formelle du législateur, l'application des normes d'IAG résultant d'une acquisition du double critère peut être subsidiairement envisagée.

**204. Le double critère du facteur substantiel : du *test de capacité juridique aux opinio subjectiva et estimatio affectiva*.** — Ce double critère du lien de rattachement concerne deux éléments : l'élément objectif ainsi que l'élément subjectif. L'élément objectif consiste dans un acte juridique accompli par l'IAG ou un fait juridique impliquant l'existence de l'IAG pertinente sur la question juridique. En d'autres termes, tout comme le jeu d'imitation, aujourd'hui communément appelé le test de Turing, dans le monde juridique où l'authenticité d'un agent s'évalue par rapport à la capacité juridique<sup>577</sup>, cette version du

---

<sup>575</sup> L'on pense à une phase préparatoire *materiae* Savignien qui ne s'occupe pourtant pas de la répartition des catégories juridiques en se dérivant vers le rattachement territorial-national en tant qu'une clôture définitive. En revanche, il s'agit d'un rapport de droit à la finalité inclusive consolidant l'attribution de la *ratione materiae*.

<sup>576</sup> Cf. l'exemple supra du droit taïwanais qui pourrait désigner les circonstances dans lesquelles sa compétence sur l'IAG soit accessible.

<sup>577</sup> Cf. supra La dualité des éléments constitutifs : l'autonomie et la capacité.

*test de capacité juridique* se traduit en droit par le fait de savoir si une IA est capable d'effectuer des tâches requises avec la capacité juridique du même niveau humain. Partant, le facteur de généralité dans le fait se traduit par la capacité juridique dans le droit ; cela se fait par le résultat de l'autonomie de l'IAG : la capacité de discernement se fonde sur ses propres données d'expériences et d'expérimentations<sup>578</sup>. Quant à l'élément moral, l'appréciation se traduit par le fait que la personne physique, c'est-à-dire le sujet traditionnel de nos jours, ne voit pas l'IAG en cause comme un simple objet sur le plan juridique, mais a la conviction subjective de la considérer en tant qu'un être doté d'une identité de sujet juridique, voire d'un lien affectif au sein de la relation juridique<sup>579</sup>.

D'une part, l'élément matériel exige que la relation juridique éprouve l'accomplissement d'un acte ou d'un fait juridique résultant d'une capacité juridique *de facto* apportée par l'IAG. D'autre part, l'élément moral est soit *opinio subjectiva*, qui est la conviction que l'IAG constitue un sujet de droit, soit d'*estimatio affectiva*, qui est le sentiment personnel établi envers l'IAG. Par conséquent, les rapports juridiques se trouvent dans le fait qu'une entité qui se noue avec d'autres entités en disposant ces deux chefs de facteurs de rattachement. C'est donc que s'il n'existe pas une sorte de capacité juridique de l'IAG dans un acte juridique ou qu'il n'y a pas de fait juridique impliquant l'IAG, il n'y aurait pas besoin de s'interroger sur l'aspect moral du double critère du lien rattachant l'IAG en cause. Par exemple, l'on pourrait considérer que les funérailles pour les chiens robots<sup>580</sup>, malgré la présence du facteur moral, en l'absence d'une réalisation de la dualité des éléments constitutifs à savoir l'autonomie et la capacité<sup>581</sup> pour réussir le test de capacité juridique, il n'y aurait pas lieu d'examiner la suite pour une telle qualification.

### C. L'appréciation des conditions

**205. Nature cumulative et préalable.** — Tous les critères doivent être réunis pour établir le lien de rattachement déclenchant la compétence inclusive pour l'application d'une règle d'IAG. Cette exigence pourrait apparaître sévère car la condition objective de la capacité juridique seule ne suffit pas à déclencher cette inclusion. En revanche, il fallait en outre une

---

<sup>578</sup> Cf. supra Les caractéristiques de l'IAG et s.

<sup>579</sup> Cf. supra Vers un pretium doloris d'IA et s.

<sup>580</sup> *Ibidem.*

<sup>581</sup> Cf. supra La dualité des éléments constitutifs : l'autonomie et la capacité.

conviction subjective pour aboutir à la revendication. Sévère parce que l'on est dans l'hypothèse où le législateur ne revendique pas explicitement cette inclusion<sup>582</sup>, mais l'on voit qu'il y a tout de même une situation substantielle (niveau de l'IAG en réalité) qui devrait être abordée, le cas échéant, pour tout le monde.

Il devient nécessaire d'opérer une distinction quant à l'appréciation de ce double critère. Comme l'élément objectif du double critère consiste dans la capacité juridique de facto de l'IAG, il convient de l'examiner en faisant une appréciation *in abstracto* : il suffit qu'une relation juridique comprenne un acte juridique accompli par la capacité juridique de l'IAG ou un fait juridique impliquant l'existence de l'IAG observé de manière générale et impersonnelle. En revanche, pour ce qui est des éléments moraux qui font l'objet d'une appréciation *in concreto*, leur condition est valable dès qu'il y a manifestation d'une telle conviction subjective dans le cas d'espèce. Par exemple, les descendants de Cicero, AlphaFold et GPT4 etc. vont pouvoir négocier, « vivre » et faire des recherches avec les êtres humains. Dans le cas de *negotiorum gestio*, ou un contrat entièrement délibéré et signé par une IAG pour coopérer avec son homologue humain, il suffit que l'IAG remplisse effectivement les conditions juridiques prévues par les dispositions respectives. Mais si dans un second temps, ces homologues humains ont une conviction subjective envers l'IAG<sup>583</sup>, alors il y aurait lieu pour la compétence substantielle d'intervenir.

**206. En résumé.** — Dans le cas où le législateur reconnaît explicitement que la *ratione materiae* d'une norme relève de l'IAG, l'on peut considérer que c'est un facteur formel quant à son applicabilité directe. En revanche, même si ce souhait du législateur n'a pas été inséré dans le texte du dispositif concerné, le facteur substantiel du lien de rattachement, établi sur le fondement du double critère cumulatif évoqué ci-dessus, devrait être appelé à jouer son rôle pour l'applicabilité d'une norme de l'IAG, en dépit de l'absence d'indication de la part du législateur. Les critères élaborés ci-dessus pour identifier la nature juridique d'IAG s'appliquent au stade du système objectif d'IA<sup>584</sup>, où les acteurs litigieux sont les personnes physiques, c'est-à-dire des êtres humains non augmentés.

---

<sup>582</sup> V. supra Le facteur formel de la compétence inclusive.

<sup>583</sup> Par exemple, en 2022, un ingénieur de Google affirme sa conviction que l'IA LaMDA est sensible.

<sup>584</sup> Cf. supra § Le registre d'IA.

## Conclusion de la section 2

**207. Compétence reconceptualisée.** — Dans un premier temps, le facteur de l'influence réelle transfrontière s'oppose aux facteurs de rattachement classiques. À ce niveau, le fondement antérieur du territoire et de la nationalité peut par conséquent être affaibli dans le contexte de l'IAG. Dans cette perspective, l'on a dégagé la concrétisation du critère d'influence réelle transfrontière. Dès lors, en transposant ces observations en l'influence réelle pluridimensionnelle de l'IAG, l'on a abouti à la consécration d'une compétence dite inclusive dont les critères concernent à la fois le facteur formel et le facteur substantiel.

## Conclusion du chapitre 1

**208. De la compétence inclusive au répertoire d'IAG.** — Dorénavant, l'on peut remarquer que la référence de territoire antérieurement absolue joue un rôle affaibli dans le contexte de l'IAG. Tirant ces considérations, cette étude a démontré qu'une telle évolution juridique donne naissance au principe d'extraterritorialité ainsi qu'à la méta-régulation.

Suivant ces perceptions, en raison des défis partagés et des influences réelles de facto, cette étude a dégagé l'inclusivisme du droit d'IAG. Cette conception se fonde d'une part, sur l'intervention formelle du législateur prévoyant sa nature normative relative à l'IAG, et d'autre part, sur l'appréciation substantielle désignant le déclenchement de la compétence. Reste à voir l'implication d'une telle configuration juridique.

## Chapitre 2 Le répertoire de l'IAG

**209. Annonce du plan.** — La compétence inclusive identifie un rapport juridique en vertu de l'influence réelle de l'IAG en cause, sa mise en place est par définition extranationale et suscite ainsi un répertoire recueillant un large éventail du droit d'IAG (section 1). Au fil du temps, non seulement l'IA devient active sur le plan juridique, l'humain est aussi augmenté par la machine sur le chemin du transhumanisme ; dès lors, les règles du répertoire d'IAG vont se synthétiser en un socle commun, ce qui fait naître le droit inter-espèces (section 2).

### Section 1 Le pluralisme du répertoire de l'IAG

**210. Annonce du plan.** — La compétence inclusive du droit d'IAG va aboutir à un répertoire d'IAG, dont la composition (§1) et la mise en place (§2) méritent d'être étudiées.

#### §1 La présentation du répertoire

**211. Annonce du plan.** — Afin d'introduire l'essence du répertoire d'IAG, il convient de connaître son intérêt (I), son rôle (II) ainsi que sa mission (III).

#### **I. L'intérêt du répertoire : complémentaire du principe de faveur**

**212. La couverture de lacune juridique.** — Le principe de faveur de l'IAG évoqué ci-dessus vise la régulation du développement de l'IAG par son autorégulation étendue. Mais qu'y-a-t-il pour ceux qui ne sont impliqués dans aucun laboratoire, ou en l'absence d'un contexte professionnel ? Dans cette hypothèse, pour le développeur dit indépendant d'IAG, les normes élaborées par l'autorégulation lui échappent complètement. Autrement dit, même si cet individu s'engage dans le développement d'IAG, l'autorégulation, étendue ou non, ne le concerne pas. L'on peut citer l'exemple de John Carmack, qui pense que l'IAG « n'est probablement pas aussi éloignée que beaucoup le croient », s'est engagé dans la création d'IAG depuis 2019. Après avoir changé de poste chez Oculus, une filiale de Meta, il joue désormais le rôle de consultant, et son engagement sur la recherche d'IAG devient « la poursuite de ses enquêtes à partir de chez lui » car il pense que l'IAG est « possible,

extrêmement utile » et qu'il a « une chance non négligeable d'y faire une différence »<sup>585</sup>. Son annonce peut être considérée comme un engagement personnel plutôt qu'un projet commercial abordé au sein de la filiale de Meta. Disposant de moyens de recherche similaires, son développement de l'IAG bénéficie pourtant d'une immunité au regard du standard étendu par le principe de faveur s'appliquant au groupe Meta car désormais il ne fait pas partie de la politique prescrite par le groupe. Il y a donc un vide juridique à combler pour ces individus qui travaillent sur l'IAG mais s'exonèrent d'autorégulation<sup>586</sup>.

Ainsi au premier degré, le principe de faveur peut garder des mécanismes tels que le code de bonne conduite<sup>587</sup> et la méthode intrinsèque<sup>588</sup> au stade de la création d'IAG. Mais pour les individus qui ne font partie d'aucune autorégulation, leur recherche sur l'IAG pourrait bel et bien contourner l'application étendue du principe de faveur. C'est la raison pour laquelle, tout comme en droit collectif du travail, ces individus qui se trouvent en dehors des normes étendues d'autorégulation seront soumis au dispositif en vigueur. D'où l'intérêt de ce répertoire de principe extraterritorial de l'IAG, qui fournit un partage transfrontière des dispositions régissant la régulation de l'IAG, telles que les interactions des IA avancés avec les humains. Cette catégorie de normes d'IAG élaborées par chaque corps législatif compétent dans le répertoire du droit de l'IAG, comme nos cerveaux, seraient téléchargées pour une énumération du droit comparé d'IAG<sup>589</sup> ; et tout ce qui dispose d'un statut d'Homo sapiens et au-delà, seraient susceptibles d'y être soumis, en tant que sujet de cette branche de souveraineté récoltée<sup>590</sup>.

## II. Le rôle du répertoire : ordre public relatif d'IAG

**213. Garantie d'un standard minimum.** — Une fois le répertoire de l'IAG en place pour sa fonction de complémentarité ainsi que son principe de subsidiarité au regard du principe de faveur d'IAG, la nature de ce répertoire est claire : mis à part des lacunes en droit concernant les projets personnels du développement de l'IAG, une seconde vocation pour

---

<sup>585</sup> DAMIANI J., « John Carmack Stepping Down To Focus on Artificial General Intelligence, Remains "Consulting CTO" At Oculus », *Forbes*, 14 nov. 2019.

<sup>586</sup> 3 ans plus tard, ce projet personnel est désormais encadré dans son entreprise d'IAG Keen Technologies.

<sup>587</sup> Cf. supra l'autorégulation.

<sup>588</sup> V. supra la méthode intrinsèque p. ex. d'Agent Moral Artificiel.

<sup>589</sup> V. infra l'organe de contrainte en ce qui concerne l'IAG.

<sup>590</sup> V. infra la souveraineté inter-espèces.

ce répertoire d'IAG vise à garantir un standard minimum pour la conduite de l'IAG. De même que les normes qui trouvent leur source dans la convention collective par rapport à l'ordre public social peuvent y déroger en faveur des meilleures conditions, de même l'autorégulation étendue peut contredire les dispositions en vigueur seulement dans un sens plus favorable. À travers cette évaluation, la détermination de la règle éventuellement applicable peut résulter d'une appréciation globale. En matière d'IAG, l'on peut considérer qu'une norme est plus favorable étant plus rigoureuse en méthode de précaution, ou simplement quand elle est plus prudente à l'égard du risque existentiel de l'humanité. En effet, étant donné que l'application dérogatoire est uniquement conditionnée par un sens plus favorable, une légitimité de ces normes issues de l'autorégulation plus restrictive que le dispositif constant ne devrait pas être remise en cause.

### **III. La mission du répertoire : « darwinisme juridique » de l'IAG**

**214. La nature polylégale du répertoire.** — À ce stade, les inconvénients d'une convention multilatérale en matière d'IAG peuvent être soulevés. Comme le mécanisme à l'heure actuelle dépend avant tout de la volonté des États westphaliens, il est toujours possible que certains pays – territoire comme population – se trouvent en dehors de l'application. En plus, vu le rythme des développement et progrès de l'IAG, sa vitesse d'impact global ne permet pas aux pays de se mettre à la table pour discuter et négocier en y consacrant de longues heures. La solution immédiate est donc de renvoyer à une branche inclusive récoltant les normes d'IAG mises à jour et adaptées. Donc dans un premier temps, un principe de faveur transfrontière en matière d'IAG conduit à l'application d'une autorégulation étendue. Le répertoire du droit d'IAG vient par la suite compléter et consolider des règles applicables.

Caractérisé de complémentarité et de subsidiarité vis-à-vis de l'autorégulation étendue, le répertoire n'en est pas moins prépondérant notamment au regard des rapports juridiques entre les entités, qui, à ce stade, concerne tous les acteurs, y compris les IAG eux-mêmes dotées de la capacité juridique voire de la personnalité juridique. En termes d'intervention instantanée, tandis que les sources du premier niveau assurent un contrôle de précaution a priori, le deuxième niveau joue un rôle de recours a posteriori, pour à la fois tempérer l'insuffisance d'autorégulation et enrichir le traitement juridique.

Suivant ce schéma, l'on pourrait noter qu'une société polylégale (Polylegal Systems) où « différentes personnes sont soumises à des règles juridiques différentes, mais où aucun des systèmes juridiques n'a un statut supérieur aux autres »<sup>591</sup>, serait privilégiée. Dans cette perspective, il faut en outre remarquer que le répertoire d'IAG s'éclaircit à travers l'idée de « *darwinisme juridique* » — en sélectionnant les normes qui leur conviennent le mieux, c'est-à-dire celles qui sont les plus aptes à satisfaire les besoins du droit : « entre droit des marchés et marché des droits, le pas est d'ailleurs franchi pour que s'instaure la compétition entre système de droit »<sup>592</sup>.

**215. En résumé.** — Ce répertoire vise donc un établissement du droit d'IAG, à la fois transfrontière par sa forme en termes d'équilibre juridique, et plurilatéral au fond quant à l'application de ses règles substantielles. Ce choix s'explique notamment par la difficulté qu'éprouve un corpus juridique qui ne dispose pas de moyens pour aborder les questions juridiques d'IAG ; il existe pourtant une nécessité d'édicter des réglementations, littéralement pour le monde, qui se trouve de toute manière concerné et affecté par le développement de l'IAG. Partant, tous les acteurs, qu'ils soient ou non dispensés de prescription d'IAG, peuvent à la fois bénéficier d'une solidarité juridique universellement plurale, et se conformer aux normes souscrites que l'on trouve ainsi dans le répertoire.

## **§2 La mise en œuvre du répertoire**

**216. Annonce du plan.** — Après avoir vu le rôle du répertoire, quant à sa mise en place, il convient à ce stade de soulever la question de l'organe de contrainte, que l'on pourrait croire étant indispensable pour l'existence d'un ordre juridique<sup>593</sup>. Certes, cette thèse ne cherche pas à définir ou à qualifier ce qu'est un ordre juridique. Néanmoins, notre présente analyse repose uniquement sur l'aspect de l'organe de contrainte par rapport au sujet de notre étude. Ainsi, cette étude propose de traiter cette question en distinguant entre l'évolution de l'organe de contrainte pour le court terme, et celle pour le long terme. La première phase

---

<sup>591</sup> FRIEDMAN D. et alii, *Legal Systems Very Different From Ours*, independently published, 2019.

<sup>592</sup> DELMAS-MARTY M., *La refondation des pouvoirs*, Seuil, 2007, p. 146. Cf. également supra la note sur la limite de sélectionisme-substantivisme.

<sup>593</sup> MAYER P., *Le phénomène de la coordination des ordres juridiques étatiques en droit privé*, Cour général de droit international privé, 2003.

s'agit du modèle de l'autorégulation commencé à partir de nos jours (I) ; et la seconde concerne le modèle où il est impliqué l'intervention de l'IAG (II).

### **I. L'organe de contrainte en ce qui concerne l'autorégulation de l'IA**

**217. Une perception évolutive de l'aspect de contrainte dans la notion de l'ordre juridique.** — Lorsque l'on observe différents types de normes, dont les « relations paraissent échapper à l'emprise d'un droit étatique, voire d'un droit uniforme intégré dans la législation des États qui y ont adhéré, pour être aménagées et gouvernées selon des normes d'origine professionnelle, ou des règles coutumières et des principes que des sentences arbitrales révèlent, à moins qu'elles ne les élaborent »<sup>594</sup>, les questions se posent : « ces 'directives de conduite' sont-elles des règles de droit et les comportements qu'elles déterminent sont-ils juridiques »?<sup>595</sup> Autrement dit, pour certains, « un ensemble de modes de conduite humaine ne peut former un système de droit que s'il existe une juridiction apte à veiller à leur interprétation et à leur observation »<sup>596</sup>. Dès lors, l'on peut leur refuser la qualification d'un « ordre juridique au même titre que l'ordre juridique étatique, c'est-à-dire comprenant, comme on le prétend, des règles, des juges et des sanctions »<sup>597</sup>. D'après ces critères, avoir une vision normativiste ne suffit pas à constituer un ordre juridique ; en revanche, il nécessite en outre un organe de contrainte assumé par le juge.

En fait, lorsque des différents types de règles sont proposées, ces enquêtes ne prétendent pas forcément avoir un rapport avec la notion d'ordre juridique. Certes, si l'on cherche à effectuer ces différentes propositions dites juridiques, selon le concept existant, il fallait passer de l'ordre juridique, qui est avant tout un ordre juridique étatique, pour accéder in fine à leur application finale. Mais déjà, la présente étude d'IAG ne prétend pas s'intégrer de manière prépondérante dans un tel modèle, et l'on verra ultérieurement pourquoi.

Néanmoins, même si l'on croit que la plupart de ces « *leges* » se contentent simplement de se focaliser sur leur vue normativiste (le phénomène d'applicabilité de ces normes non stricto

---

<sup>594</sup> GOLDMAN B. « Frontières du droit et *lex mercatoria* », Archives de Philosophie du Droit, t.IX, 1964, p.177-192.

<sup>595</sup> *Ibidem*.

<sup>596</sup> *Ibid.*

<sup>597</sup> MAYER P., *Le phénomène de la coordination des ordres juridiques étatiques en droit privé*, op. cit., p. 65.

sensu), il y a tout de même quelques remarques que l'on pourrait faire sur la question de l'organe de contrainte du point de vue de notre étude. Effectivement, *prima facie*, ces lois dites spéciales semblent dépourvues de l'aspect incluant les normes secondaires ou procédurales voire institutionnelles. Mais à y regarder de plus près, l'on pourrait observer que les exemples tels que le RGPD, AI Act etc. disposent désormais de leurs propres organes compétents. À première vue, l'on peut dire que ces textes et ces matières, qui ne font que percevoir de nouveaux instruments juridiques pour répondre à de nouvelles réalités, se prétendent pourtant comme un ordre juridique de « données » ou de « l'IA », sans même fournir un organe de contrainte. Dès lors, sans un juge propre à eux pour leur mise en œuvre définitive, ils ne doivent pas être considérés comme un ordre juridique. Mais en lisant ces textes, l'on voit qu'ils fournissent des autorités institutionnelles et des normes procédurales. Donc certes l'on peut toujours dire que leur force exécutoire finale dépend uniquement d'autres organes et qu'il s'agit *in fine* d'un ordre juridique déjà existant. Mais à notre avis, l'on ne peut pas exclure la possibilité que de tels mécanismes normatifs, avec leurs fournisseurs techniques envisagés, puissent bientôt prévoir les instances et institutions qui sont plus fondamentales que leur rôle ou leur position actuels pour une force contraignante.

Comme l'on a vu sur le plan de l'autorégulation, alors que les acteurs comme les GAFAMI commencent à être autonomes même au niveau normatif, aujourd'hui, leurs conseils de surveillance dépendent des juges nationaux comme l'organe de contrainte. Néanmoins, l'on pourrait s'attendre que demain, ils fourniront l'organe de contrainte en raison de leurs engagements normatifs et de leurs capacités techniques<sup>598</sup>.

**218. L'ébauche embryonnaire sous l'examen des critères de l'ordre juridique : l'exemple de Microsoft en matière d'IA responsable.** — Bien que cette étude ne cherche ni à définir ni à qualifier l'ordre juridique, l'on pourrait toutefois tenter de voir son rapport avec la discussion de notre sujet. En effet, « en l'absence de définition unanimement acceptée de ce qui constitue un ordre juridique »<sup>599</sup>, l'on va emprunter la définition d'un ordre juridique comme « un ensemble structuré de normes revêtant tous les degrés d'impérativité

---

<sup>598</sup> V. supra L'interchangeabilité des règles et la croisée des codes.

<sup>599</sup> GAILLARD E., « L'ordre juridique arbitral : réalité, utilité et spécificité », *Revue de droit de McGill*, 2010, p. 896.

et susceptibles de répondre à l'ensemble des questions relevant de la matière qu'il prétend régir; capable de concevoir ses sources; possédant des sujets et des organes susceptibles d'assurer la mise en œuvre des normes qu'il génère; et satisfaisant à une condition minimale d'effectivité »<sup>600</sup>. Donc l'on va prendre l'exemple de Microsoft pour examiner successivement ces quatre critères.

D'abord, pour « un ensemble structuré de normes revêtant tous les degrés d'impérativité et susceptibles de répondre à l'ensemble des questions relevant de la matière qu'il prétend régir ». Sur ce critère, l'on pourrait penser à la gouvernance structurée de l'IA responsable de Microsoft (Microsoft responsible AI governance structure)<sup>601</sup>. Ensuite, sur le critère « capable de concevoir ses sources ». Cette gouvernance est en effet « administrée, mise en œuvre et maintenue par le Bureau de l'IA responsable »<sup>602</sup>, qui prévoit d'une part, la « Politique interne : Définir les règles de l'entreprise pour la mise en œuvre d'une IA responsable, ainsi que les rôles et responsabilités des équipes impliquées dans cet effort »<sup>603</sup>; et d'autre part, la « Politique publique : Contribuer à façonner les nouvelles lois, normes et standards qui seront nécessaires pour garantir que les promesses de la technologie de l'IA se réalisent au profit de la société dans son ensemble »<sup>604</sup>. Et puis, dans un troisième lieu, sur le critère de « possédant des sujets et des organes susceptibles d'assurer la mise en œuvre des normes qu'il génère ». Pour Microsoft, leur Bureau de l'IA responsable procède à « Examen des cas d'utilisation sensibles afin de garantir que nos principes d'IA sont respectés dans notre travail de développement et de déploiement »<sup>605</sup>. Et en dernier lieu, sur « satisfaisant à une condition minimale d'effectivité ». Ce mécanisme dispose en effet de « efficacité de l'autocensure : une fois le cas examiné, le contrôleur<sup>606</sup> de l'IA responsable travaille avec le bureau de l'IA responsable pour conseiller l'équipe de projet sur les stratégies d'atténuation (mitigation) qui s'alignent sur nos pratiques et principes d'IA responsable. Ces stratégies d'atténuation peuvent inclure des approches techniques, des approches de formation des employés et de gouvernance, ou des modifications de la portée

---

<sup>600</sup> *Ibidem.*

<sup>601</sup> V. Microsoft AI governance model.

<sup>602</sup> V. Understanding the Microsoft governance model - AETHER + Office of Responsible AI.

<sup>603</sup> *Ibidem.*

<sup>604</sup> *Ibid.*

<sup>605</sup> *Ibid.*

<sup>606</sup> Responsible AI Champ : une personne chargée de faire connaître et comprendre les politiques, les normes et les orientations de l'entreprise en matière d'IA responsable.

du projet. Il est arrivé que l'on conseille à nos équipes de ne pas poursuivre certains projets parce que l'on n'était pas en mesure de les réaliser dans le respect de nos principes »<sup>607</sup>. Donc outre que les principes fondamentaux et les lignes directrices de Microsoft, ils disposent également d'outils et de ressources permettant aux développeurs de repérer plus facilement les problèmes potentiellement dangereux et d'en assurer la conception<sup>608</sup>. Et ces outils et ressources sont conçus pour aider tout le monde à utiliser l'IA de manière responsable à chaque étape de l'innovation, du concept au déploiement, en passant par le développement et bien plus encore<sup>609</sup>.

Au regard de ces déploiements à la fois normatifs et techniques, l'on pourrait dire que selon le concept de Code is Law<sup>610</sup>, et d'après les quatre critères cités ci-dessus, l'autorégulation en matière d'IA constitue d'ores et déjà une sorte de l'ordre juridique. Et comme Microsoft déclare lui-même : « ce n'est que le début. L'IA est encore un domaine relativement nouveau, il n'est donc pas surprenant que les processus qui l'entourent évoluent rapidement eux aussi. À l'avenir, l'on prévoit d'affiner nos politiques de gouvernance à mesure que l'on investit dans l'IA »<sup>611</sup>. Donc même si à ce stade embryonnaire où l'on n'a que l'ébauche de cette gouvernance, l'on pourrait déjà s'attendre que ce cadre soit de plus en plus approfondi et sophistiqué. Si ces exécutions effectives provenant de la gouvernance de l'IA étroite sont déjà en train d'être systématisées, il est tout à fait envisageable que dans l'avenir, ces mécanismes deviennent davantage dominants parmi les instruments juridiques. Mais pour revenir au sujet de notre recherche — l'IA générale, il convient d'exposer une voie encore plus générale pour répondre à la question quant à l'organe de contrainte.

## II. L'organe de contrainte en ce qui concerne l'IAG

**219. De *characteristica universalis* de Leibniz au langage de Wolfram.** — Au départ, Gottfried Wilhelm Leibniz parlait de ses idées sous une variété de noms comme *scientia generalis* (méthode générale de connaissance), *lingua philosophica* (langage philosophique), *mathematica universalis* (mathématiques universelles), *characteristica universalis* (système

---

<sup>607</sup> Responsible AI frameworks in action.

<sup>608</sup> Engineering tools for responsible AI.

<sup>609</sup> Des ressources pour une IA responsable.

<sup>610</sup> V. supra la croisée des codes.

<sup>611</sup> V. Responsible AI frameworks in action.

universel) et *calculus ratiocinator* (calcul de la pensée). Sa *characteristica universalis* et son *calculus ratiocinator* visaient en effet à répondre à toutes les questions possibles par le calcul<sup>612</sup>. Il imaginait des applications dans tous les domaines : science, droit, médecine, ingénierie, théologie, etc<sup>613</sup>. Ainsi, le droit est l'un des domaines où il a essayé son calcul universel. Mais dans ce domaine, il est arrivé beaucoup trop tôt, et ce n'est que maintenant — 300 ans plus tard — que le droit informatique commence à sembler réaliste<sup>614</sup> : transformer le droit humain en un exercice de calcul<sup>615</sup>. D'après M. Stephen Wolfram, c'est une étape très importante à faire parce qu'elle sera essentielle pour l'avenir de notre civilisation dans son interaction avec l'intelligence artificielle<sup>616</sup>. Donc la question se pose de savoir peut-on utiliser les idées du calcul, d'une manière très proche de celle imaginée par Leibniz, pour formaliser le droit humain ?<sup>617</sup> Pour répondre à cette question, M. Wolfram commence à construire un pont entre la précision des langages informatiques traditionnels et la capacité de parler de constructions du monde réel. En fait, il s'agit de son travail depuis plus de trente ans : le langage de Wolfram basé sur la connaissance (*knowledge-based Wolfram Language*)<sup>618</sup>. Le langage de Wolfram est précis : tout ce qu'il contient est défini au point qu'un ordinateur peut l'utiliser sans ambiguïté. Mais sa caractéristique unique par rapport à d'autres langages informatiques est qu'il est basé sur la connaissance. Cela signifie qu'il ne s'agit pas seulement d'un langage permettant de décrire les opérations de bas niveau d'un ordinateur, mais d'un langage qui intègre autant de connaissances que possible sur le monde réel. Ainsi, le langage comprend non seulement des nombres comme 2,7 et des chaînes de caractères comme « abc », mais aussi des constructions comme les États-Unis, l'indice des prix à la consommation, ou un éléphant<sup>619</sup>. Et pour M. Wolfram, c'est exactement ce dont l'on a besoin pour commencer à parler du genre de choses qui se trouvent dans les contrats juridiques ou les lois humaines<sup>620</sup>.

---

<sup>612</sup> SCHMIDHUBER J., « 375th birthday of Leibniz, founder of computer science », AI Blog, 17 mai 2021.

<sup>613</sup> WOLFRAM S., « Dropping In on Gottfried Leibniz », sur son site, 14 mai 2013.

<sup>614</sup> *Ibidem*.

<sup>615</sup> WOLFRAM S., « Computational Law, Symbolic Discourse and the AI Constitution », sur son site, 12 oct. 2016.

<sup>616</sup> *Ibidem*.

<sup>617</sup> *Ibid.*

<sup>618</sup> *Ibid.*

<sup>619</sup> *Ibid.*

<sup>620</sup> *Ibid.* Le Wolfram Language tel qu'il existe aujourd'hui n'inclut pas tout ce qu'il faut. C'est donc même s'il y a déjà un cadre large et solide, il faut encore coder d'autres éléments du monde pour être en mesure de saisir toute la gamme des activités humaines et des spécifications juridiques humaines. Le langage Wolfram a, par exemple, une définition de ce qu'est une banane, décomposée en toutes sortes de détails.

Déjà, en matière contractuelle, M. Wolfram voit la structure des contrats comme étant très proche de celle des programmes<sup>621</sup>. Dans certains cas, ils contiennent des conditionnels logiques : « si X alors Y ». Dans d'autres cas, ils sont davantage calqués sur les mathématiques : « si telle quantité de X se produit, telle quantité de Y devrait se produire ». Parfois, il y a une itération : « continuez à faire X jusqu'à ce que Y se produise ». Parfois, il peut aussi être : « continuer à appliquer X à chaque Y ». Et ainsi de suite<sup>622</sup>. Ici, on retrouve donc l'idée que l'on a vu précédemment : Code is Law<sup>623</sup>. Rappelons ainsi que « la notion de régulation est d'abord apparue au XVIIIe siècle dans les sciences techniques. Dès 1728, on entend par 'régulateur' un système de commande destiné à maintenir constant la valeur d'une grandeur, quelles que soient les perturbations qui pourraient la faire varier : la machine à vapeur de Watt comporte ainsi un régulateur à boules, destiné à éviter les risques d'explosion ; on parlera du régulateur de température ou thermostat, du régulateur de vitesse etc. »<sup>624</sup>. Par conséquent, ce n'est pas un hasard si l'on parle de « codes juridiques ». Le mot code — qui vient du latin *codex* — désignait à l'origine des collections systématiques de règles juridiques. Et lorsque la programmation est apparue quelques millénaires plus tard, elle a utilisé le mot « code » parce qu'elle se considérait comme établissant des règles similaires pour le fonctionnement des choses, sauf que ces règles concernaient le fonctionnement des ordinateurs plutôt que la conduite des affaires du monde<sup>625</sup>. Mais aujourd'hui, avec le langage informatique basé sur la connaissance et l'idée d'un langage symbolique (knowledge-based Wolfram Language), M. Wolfram tente de faire en sorte que l'on puisse parler d'un large éventail d'affaires du monde de la même manière que l'on parle de processus informatiques — l'on met donc tous les codes juridiques et les contrats sous forme informatique<sup>626</sup>.

---

<sup>621</sup> *Ibid.* Il existe déjà des contrats juridiques couramment représentés par des programmes. Par exemple, les contrats financiers composés des obligations (bonds) et des options, se résument en effet à de petits programmes définissant les paiements en fonction de diverses formules et conditions. Ainsi, il existe toute une industrie qui utilise des « moteurs de règles (rules engines) » pour coder certains types de réglementations sous forme de règles « si alors », fréquemment associées à des formules. En fait, de telles règles sont presque universellement utilisées pour les calculs de taxes, d'assurances, et des moteurs de tarification.

<sup>622</sup> *Ibid.*

<sup>623</sup> Cf. supra la croisée des codes. Et « Les contrats légalement formés tiennent lieu de loi à ceux qui les ont faits », cette force obligatoire est donc constamment et directement matérialisée.

<sup>624</sup> CHEVALLIER J., « Contractualisation et régulation », in CHASSAGNARD-PINET S., HIEZ D. (dir.), *La contractualisation de la production normative*, Dalloz, 2008, p. 84.

<sup>625</sup> WOLFRAM S., « Computational Law, Symbolic Discourse and the AI Constitution », préc.

<sup>626</sup> *Ibidem.*

Dans cette démarche, son langage symbolique ne cherche pas à capturer le « vrai sens » d'un morceau de langage naturel. Au contraire, son objectif est simplement de capturer une certaine signification avec laquelle l'on peut ensuite calculer<sup>627</sup>. Autrement dit, certes, l'on peut commencer par le langage naturel, puis d'essayer de le traduire dans le langage symbolique. Mais l'objectif est que le langage symbolique soit la représentation réelle : le langage naturel n'est qu'un guide pour essayer de le générer. Ainsi, l'idée est que pour être sûr de l'exactitude de ce que l'on dit, il faut le dire directement dans le langage symbolique, sans jamais utiliser le langage naturel<sup>628</sup>. Partant, l'on se rend compte qu'un langage symbolique a pour but de représenter les choses du monde, tout comme le langage naturel n'est aussi qu'une autre façon de les décrire<sup>629</sup>. Et c'est à partir de cette vision que M. Wolfram vise à créer une représentation symbolique du monde, pour que l'on puisse dire ce qui doit se passer et pour que les machines le comprennent<sup>630</sup>.

**220. Vers le modèle de l'IAG : la notion contestable d'organe de contrainte dans l'ordre juridique.** — Si l'on dispose d'un réseau suffisamment dense de capteurs dans le monde, il devient de plus en plus facile d'être sûr de ce qui s'est passé. En effet, en synthétisant les informations, l'on peut inéluctablement comprendre ce qui s'est réellement passé. Dès lors, l'on dispose d'un modèle de fonctionnement du monde et, avec suffisamment de capteurs, l'on peut valider l'exactitude de ce modèle<sup>631</sup>. Par exemple, un contrat définit un certain modèle du monde, et spécifie ce qui doit se passer dans différentes situations. Et à mesure que les contrats informatiques deviennent plus complexes, ils pourront capturer plus fidèlement le fonctionnement du monde<sup>632</sup>. Par conséquent, aujourd'hui, un contrat comporte une centaine de pages de « jargon » juridique ; demain, le même contrat sera construit automatiquement à partir des objectifs recherchés et exécuté automatiquement par ordinateur<sup>633</sup>.

---

<sup>627</sup> *Ibid.*

<sup>628</sup> *Ibid.*

<sup>629</sup> *Ibid.*

<sup>630</sup> *Ibid.*

<sup>631</sup> *Ibid.*

<sup>632</sup> *Ibid.*

<sup>633</sup> *Ibid.*

Pour cette étude, l'on voit que déjà de nos jours, s'il n'y a pas un Uber (déjà disponible/connecté/assuré/mis en lien), il n'y aura pas de consentement — la formation du contrat n'aura pas lieu. Idem, pas de bon degré de température, pas de contrat formé ; pas de bon semi-conducteur : le contrat n'est pas susceptible d'être formé ; pas de paiement sécurisé : le compte est programmé pour sa suspension etc. Bref, comme les exécutions conditionnées sont précisément programmées en prenant en compte de la probabilité d'inexécution déjà calculée, les risques d'inexécution, une fois la source principale du litige, seront largement réduits sous le « contrôle » d'automatisation<sup>634</sup>. C'est à notre sens, ***la substitution de l'organe de contrainte en amont***. Autrement dit, avant qu'un contrat puisse même être formé, la vérification, la programmation et l'automatisation de l'exécution va déjà prendre place préalablement. Dans cette situation, les clauses stipulent l'hypothèse de questions juridiques que pourraient générer cette relation juridique. Et tout ceux-ci sont programmés à être « exécutés » : dans telle ou telle hypothèse, telle ou telle exécution sera appliquée directement et automatiquement. Donc déjà dans un premier temps, l'automatisation des relations juridiques va réduire largement la possibilité d'introduire des litiges, et donc absorber l'organe de contrainte en amont.

Et si, dans un second temps, il existe tout de même des cas très rares pour introduire des recours, l'on pourrait également s'attendre à ce que le même mécanisme, le même processus d'automatisation décrit ci-dessus, va s'appliquer à l'aspect « sanction ». C'est ce que cette étude voit comme ***la substitution de l'organe de contrainte en aval***. Autrement dit, de la même manière que la fonction de l'organe de contrainte sera absorbée par l'automatisation a priori, ce mécanisme sera aussi transposé en aval dans notre second volet. Cela signifie que la procédure d'automatisation absorbe en effet la procédure traditionnellement conçue pour régler la question sur le juge saisi et éventuellement le jugement prononcé<sup>635</sup>.

---

<sup>634</sup> Surtout, même avant la formation d'un contrat, il y aura toute une série de « recommandations précontractuelles » pour cet effet. Ces pratiques qui existent déjà de nos jours, seront beaucoup plus banales et élaborées. Ainsi, les humains demanderont à l'IA de leur fournir des codes (p. ex. AlphaCode, OpenAI Codex), des paragraphes de texte (p. ex. LaMDA, ChatGPT), des œuvres d'art ou des dessins (p. ex. Stable Diffusion, Dall-E, Point-E). L'IA génère toute une série d'idées ; les humains n'ont qu'à choisir celle qu'ils veulent et à améliorer les résultats. En répétant cette « autocomplétion » fournie par l'IA, les humains se contentent de laisser l'IA générer et suggérer à leur place — à chaque situation et à chaque étape — tout contenu dans la vie ; cela va s'achever donc par l'autocomplétion in toto de la civilisation.

<sup>635</sup> V. supra sur la notion de la singularité juridique et infra vers une sélection ultime des lois.

Donc à notre sens, les deux volets, soit le normativisme a priori, soit la contrainte a posteriori, font l'objet d'une histoire d'absorption — leurs fonctions demeurent, mais leurs formes absorbées et donc mutent. Dès lors, même la raison d'être de ces instances va évoluer : au fond elle va être substituée par d'autres volets. Ainsi, ces pratiques disposant de caractères à la fois directs, automatisés, décentralisés, pourraient faire que la discussion même sur le concept de l'ordre juridique mérite d'être réévaluée. En conséquence, si « un ordre juridique n'a pas besoin d'être reconnu, son existence est un fait et c'est ce fait qui compte, non la personnalité internationale éventuellement greffée sur lui »<sup>636</sup>. L'on pourrait se demander si une telle l'autocomplétion et l'auto-s'applique résultant de l'automatisation constitue *in fine* un nouvel « ordre juridique » *de novo*.

Ainsi, si « le point de départ du raisonnement, dans les relations internationales, est, non le juge, mais l'organe de contrainte effectivement susceptible de procéder à une voie d'exécution »<sup>637</sup>, lorsque ces voies d'exécution sont automatisées, l'on peut se poser la question de la place éventuelle de ces exécutions dites non-automatisées.

En fin de compte, l'on se rend compte que, effectivement, « la règle crée entre une hypothèse et un effet juridique une relation qui évoque un lien de causalité : tel événement entraîne tel effet ; Kelsen préfère parler de relation d'imputation : à tel événement est imputé tel effet »<sup>638</sup>. Dans cette optique, « il est juste de dire que la règle de droit est virtuellement universelle. On entend par là non pas que le législateur veut qu'elle s'applique à tous les individus qui peuplent la planète, mais qu'il n'y a dans la règle aucune considération relative à l'identité ou à la localisation spatiale de ses destinataires. La relation établie entre une obligation de réparer et la survenance d'un dommage causé par une faute repose sur l'adéquation de l'effet à la nature de l'événement. Que cet événement soit survenu ici ou là, qu'il ait été causé par telle personne ou par telle autre, domiciliée ici ou là, ayant telle nationalité ou telle autre, ne sont pas des considérations pertinentes ; elles ne sont donc pas intégrées dans la règle »<sup>639</sup>. Partant, « si l'on songe au pouvoir, aujourd'hui reconnu en matière internationale par la plupart des ordres juridiques aux parties, de choisir la loi qui régira leurs rapports

---

<sup>636</sup> MAYER P., *Le phénomène de la coordination des ordres juridiques étatiques en droit privé*, op. cit., p. 110.

<sup>637</sup> *Ibidem*, p. 113.

<sup>638</sup> *Ibid.*, p. 114.

<sup>639</sup> *Ibid.*

contractuels, voire, de plus en plus, extracontractuels, y compris ceux qui relèvent du statut personnel »<sup>640</sup>, l'on pourrait s'attendre à ce que dès que l'automatisation juridique évoquée ci-dessus prenne son essor et que par voie de conséquence tous les aspects juridiques seront transformés.

Ainsi, au XXème siècle, je signe un contrat et passe par un juge humain pour l'exécution forcée compte tenu de son inexécution par mon débiteur. Mais au XXIème siècle, avant que je puisse même signer ce contrat, tout est calculé et évalué par l'IA au préalable pour la programmation du contrat. Par exemple : 30% de risque d'inexécution pour telle raison et 40% de risque d'inexécution pour telle autre circonstance ; et en cas d'inexécution du contrat : 20% d'indemnisation seront auto-appliquées sous telle ou telle condition ; 30% d'indemnisation seront auto-exécutées à partir de telle date ou telle date si tel ou tel acte n'a pas été suivi ou tel ou tel remboursement n'a pas été effectué. Bref, en fin de compte, les contractants ajustent le contenu du contrat par rapport à son automatisation pour son meilleur aboutissement. Encore une fois, c'est *la substitution de l'organe de contrainte en amont*. Et dans le cas presque impossible d'inexécution (coupure de courant par accident aléatoire fait par des animaux sauvages se précipitant dans votre maison), alors les conditions mises en place pour un tel scénario d'inexécution seront aussi auto-exécutées : l'hypothèse de la force majeure déjà préprogrammée. L'on rappelle que cela constitue en effet *la substitution de l'organe de contrainte en aval*. Et dans le cas presque impossible où des accidents se produisent pour empêcher la mise en œuvre automatique de l'inexécution, alors c'est l'occasion où les descendants des êtres humains pourraient éventuellement intervenir<sup>641</sup>.

Au final, si l'on défend l'ordre juridique en se fondant sur l'argument de l'organe de contrainte, il paraît incohérent d'utiliser les éléments « clos » conférés à l'ordre juridique dit étatique pour justifier un tel organe de contrainte monopolisé par lui, et puis se servir de ce monopole d'organe de contrainte dit étatique pour justifier par suite un tel ordre dit juridique. Autrement dit, ce raisonnement nous semble fallacieux car la justification elle-même n'est en effet que le résultat du choix de la prémisse qui a été a priori adoptée en premier lieu.

---

<sup>640</sup> *Ibid.*, p. 144.

<sup>641</sup> Mais l'on considère que l'hypothèse où l'IA n'a pas préparé un plan de secours ou une solution même pour cela sera très rare et donc restera par conséquent exceptionnelle.

**221. En résumé.** — Vu que les exécutions seront programmées, et que seules les sources vérifiées ou vérifiables seront ainsi programmées, le tiers humain — intermédiaire de l'organe de contrainte traditionnel — sera en grande partie éliminé. La contrainte demeure, mais sera absorbée par son propre processus et deviendra auto-vérifiée, auto-programmée, et auto-effectuée et auto-exécutée, à tous les niveaux et dans tous les domaines. C'est comme si, auparavant, l'on s'attendait à ce que les deux parties se battent et donc qu'il y ait une tierce partie pour ces différends. Mais à l'issue de l'ancien régime de nos jours, il n'y aura plus de conflits, ou du moins la possibilité qu'un conflit se produise sera considérablement réduite. Par conséquent, l'existence d'une telle tierce partie humaine sans automatisation disparaîtrait.

## Conclusion de la section 1

**222. Des ordres juridiques au droit inter-espèces<sup>642</sup>.** — Dans cette section, l'on a démontré le pluralisme du répertoire du droit d'IAG en exposant son intérêt étant la couverture de lacune juridique, son rôle comme ordre public relatif d'IAG pour garantir un standard minimum, ainsi que sa mission consistant dans sa nature polylégale. Quant à sa mise en œuvre, l'on a remarqué que la circulation des normes du répertoire se fonde en effet sur la reconnaissance du pluralisme, c'est-à-dire de l'hétérogénéité des régimes juridiques se fondant sur la compétence inclusive.

Comme l'on a vu, ce ne sera que dans des cas exceptionnellement rares que l'on a besoin des descendants des êtres humains comme l'organe de contrainte. Ces occasions nécessiteront alors une intervention substantielle que seul « le fond », et non « l'étiquette formelle » sera éligible. De telles exigences conduiront alors à un état d'esprit inclusif afin de pouvoir adopter les solutions pertinentes. Et en pratique, la solution substantielle et la solution technique entrelacées de manière réciproque. Ainsi, paradoxalement, les conflits de lois se multiplient, mais une fois automatisés et augmentés, ils peuvent être radicalement normalisés et réduits dans une autre dimension de la programmation.

Vu que l'aspect juridique de la civilisation entière se déroulera en état perpétuel de « auto-complété, auto-vérifié, auto-programmé, auto-mis à jour et auto-exécuté », cette automatisation va largement réduire la pertinence du territoire, de la nationalité, ainsi que du tiers humain qui a été chargé d'un tel besoin *sui generis*, requis depuis des siècles. Partant, il faut se demander que signifie la notion classique « d'extranéité » dans le droit international. En internationalisant la civilisation, un tel élément peut certes être multiplié et développé. Mais les exigences quant à ces prémisses impertinentes auront disparu. Dès lors, cette étude se propose d'étudier la désinternationalisation ainsi que l'interspécialisation de la culture juridique de la civilisation humaine.

---

<sup>642</sup> CHEN Y.-T., « The implications of artificial general intelligence for international law as a legal system », International Conference ICTO 2023 Ethical & Responsible Artificial Intelligence for sustainable societies, Paris 1 Panthéon-Sorbonne, Paris.

## **Section 2 Le droit inter-espèces**

**223. Annonce du plan.** — À ce stade, non seulement des compétences juridiques se mutualisent, mais les sujets juridiques, à savoir l'humain et l'IA fusionnent également. Dans ce cas, il convient d'étudier ces statuts inter-espèces (§1) et le droit inter-espèces (§2).

### **§1 Les statuts inter-espèces**

**224. Annonce du plan.** — L'on vient de voir le répertoire du droit d'IAG qui implique éventuellement la substitution de l'organe de contrainte. Mais à ce syncrétisme s'ajoute le progrès du transhumanisme (I), dès lors, l'on aura une souveraineté inter-espèces (II).

#### **I. Le chemin du transhumanisme**

**225. La notion du transhumanisme.** — Le mot « transhumanisme » semble avoir été utilisé pour la première fois<sup>643</sup> par Julian Huxley quand il écrit « L'espèce humaine peut, si elle le souhaite, se transcender elle-même — non seulement sporadiquement, un individu ici d'une certaine manière, un individu ici d'une autre manière — mais dans son intégralité en tant qu'humanité. L'on a besoin d'un nom pour cette nouvelle croyance. Le transhumanisme peut être considéré : l'homme restant homme, mais se dépassant lui-même, en réalisant de nouvelles possibilités de et pour sa nature humaine »<sup>644</sup>. Le transhumanisme constitue donc une étape pour l'homme de se rapprocher au niveau de l'IAG et au-delà. Pour ce faire, plusieurs pistes ont été envisagées.

Le téléchargement d'esprit (Whole brain emulation, WBE) fait référence à une technologie qui permettrait à un intellect humain ou à un autre animal d'être transféré, de sa mise en œuvre originale dans un cerveau organique à un ordinateur numérique. Le résultat peut habiter dans une réalité virtuelle simulée ou, in fine, recevoir le contrôle d'un corps robotique lui permettant d'interagir directement avec la réalité physique externe<sup>645</sup>. En d'autres termes, au lieu d'avoir une IA programmée à partir de zéro (de novo, dnAGI), dans

---

<sup>643</sup> BOSTROM N., « A History of Transhumanist Thought », vol. 14 Issue 1, avr. 2005, p. 7.

<sup>644</sup> HUXLEY J., *New Bottles for New Wine*, Chatto & Windus, 1957, p. 17.

<sup>645</sup> BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », préc., p. 325.

le cas de WBE, le cerveau d'un humain est scanné, et cette numérisation est traduite en un modèle, qui est exécuté sur un ordinateur, donnant ainsi une intelligence similaire à celle de la personne dont le cerveau a été scanné : cet humain aurait donc été « téléchargé »<sup>646</sup>. Dans l'hypothèse où le procédé fait naître un programme émergeant la même personnalité, les mêmes souvenirs et les mêmes schémas de pensée que le cerveau biologique, la question se pose alors d'éclaircir sur l'identité personnelle en cas d'un téléchargement qui ferait en sorte que plusieurs téléchargements similaires ou qualitativement identiques fonctionnent en parallèle<sup>647</sup>. Et que se passe-t-il si l'on arrive à remplacer la moitié de notre cerveau originaire ou le cerveau entier ? Dans une telle hypothèse, il n'y aurait aucune raison de nier la protection de la personnalité juridique, sinon, les personnes se verraient refuser un statut juridique simplement parce qu'elles ont trop de machines en eux pour rester humain<sup>648</sup>.

Et puis, l'arrivée de l'interface neuronale cerveau-cerveau (brain-to-brain interface, BTBI) pourrait connecter directement le cerveau à un système informatique afin de pousser l'interface homme-machine à son maximum<sup>649</sup>. Le BTBI pourrait diriger la transmission d'informations entre n'importe quel ensemble de périphériques du réseau d'interface cerveau-cerveau, et rendre celui-ci utilisable de manière globale, permettant des interactions entre cerveaux, jusqu'à l'échelle mondiale<sup>650</sup>. Aujourd'hui, le BTBI peut déjà relier le cerveau de rats et leur permettre de partager de l'information sensorielle<sup>651</sup>. En effet, l'interface neuronale périphérique, non invasive et portable, permet déjà aux humains de converser en langage naturel avec des machines, des assistants d'IA, et d'autres personnes sans voix, sans même ouvrir la bouche et sans mouvements observables de l'extérieur, mais simplement en articulant les mots en interne<sup>652</sup>. Dès lors, la pensée peut désormais être

---

<sup>646</sup> ETH D., « The Technological Landscape Affecting Artificial General Intelligence and the Importance of Nanoscale Neural Probes », *Informatica* 41, 2017, p. 463.

<sup>647</sup> BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », préc., p. 325.

<sup>648</sup> WILLICK M. S., « Artificial Intelligence: Some Legal Approaches and Implications », préc., p. 13.

<sup>649</sup> LAURENT A., *La guerre des intelligences : Comment l'intelligence artificielle va révolutionner l'éducation*, op. cit., p. 57.

<sup>650</sup> STOCCO A. et alii, « BrainNet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration Between Brains », arXiv.org, 23 sept. 2018.

<sup>651</sup> PAIS-VIEIRA M., « A Brain-to-Brain Interface for Real-Time Sharing of Sensorimotor Information », *Scientific Reports*, 28 févr. 2013.

<sup>652</sup> KAPUR A. et alii, « AlterEgo: A Personalized Wearable Silent Speech Interface », 23rd International Conference on Intelligent User Interfaces, 5 mars 2018, pp. 43-53.

traduite en discours intelligible<sup>653</sup> ; et ces interfaces cerveau-machine (brain-machine interface, BMI) pourraient alors aboutir à un contrôle partagé collaboratif entre les humains et la robotique<sup>654</sup>, fusionnant ainsi au fur et à mesure le carbone et le silicium.

En conséquence, relier le cerveau humain aux IA nous permet de fusionner la capacité de l'intelligence humaine avec la puissance des ordinateurs, et de créer une nouvelle forme de collaboration d'intelligence. Inévitablement, les interfaces cerveau-ordinateur et autres dispositifs qui brouillent les frontières entre l'esprit humain et la machine pourraient conduire à la violation des droits de l'homme<sup>655</sup>, et sont susceptibles de remettre en question l'essence même de l'être humain.

**226. Le cyborgisation<sup>656</sup> dans les réalités diverses<sup>657</sup>.** — Pour pouvoir contrôler l'IA dotée d'une intelligence supérieure, les humains seront obligés de se transformer eux-mêmes en organisme cybernétique et homme bionique : la différence entre robots humains et humains robots tendra à disparaître<sup>658</sup>. Ainsi, les concepteurs de demain ne seront pas des humains, mais des « cyborgs » qui construiront eux-mêmes les augmentations<sup>659</sup>. De nos jours, la technologie de Neuralink nous permettra de nous intégrer aux ordinateurs pour accéder à l'information et la traiter, ainsi que nos homologues en IA : les cerveaux humains auront leurs propres IA dans un centre de données distant avec lequel ils pourront synthétiser les informations<sup>660</sup>. L'idée est donc de devenir plus fort lorsque l'on est combiné à la machine,

---

<sup>653</sup> AKBARI H. et *alii*, « Towards reconstructing intelligible speech from the human auditory cortex », *Scientific Reports*, 29 janv. 2019.

<sup>654</sup> HANDELMAN D. A. et *alii*, « Shared Control of Bimanual Robotic Limbs With a Brain-Machine Interface for Self-Feeding », *Frontiers in Neurorobot*, vol. 16, 28 juin 2022.

<sup>655</sup> iHuman Neural interfaces report: Blurring lines between mind and machine, The Royal Society, sept. 2019, pp. 15-16.

<sup>656</sup> « Cyborg », cybernetic organism (organisme cybernétique). Un cyborg est un être vivant qui a reçu des greffes de parties mécaniques ou électroniques. Le pacemaker et d'autres dispositifs médicaux implantés, ou des technologies portables comme les lunettes de réalité augmentée, fournissent une illustration de l'écart au sein de l'espèce humaine.

<sup>657</sup> P. ex. la réalité virtuelle (Virtual Reality, VR), la réalité augmentée (Augmented Reality, AR) et la réalité mixte (Mixed Reality, MR).

<sup>658</sup> GORZ A., *L'Immatériel*, Galilée, 2003.

<sup>659</sup> Bientôt, notre corps sera la seule interface dont on aura besoin. Par exemple, l'on peut déjà payer avec une puce de paiement sans contact implantée dans les mains. En tant que l'extension de l'internet des objets (IdO, Internet of things, IoT), l'on peut même avoir des aimants implantés dans les doigts. Aux Pays-Bas, un homme ayant 32 implants au total, y compris des puces pour ouvrir des portes et des aimants intégrés, affirme que ses implants augmentent son corps et qu'il ne voudrait pas vivre sans eux. « The microchip implants that let you pay with your hand », BBC, 11 avr. 2022.

<sup>660</sup> ASTRO J., « Elon Musk explores Neuralink's benefits and AI's danger in Axios interview », TESLARATI, 27 nov. 2018.

augmenté par la machine. De là, il faut se demander comment peut-on suivre dans une civilisation où certains sont augmentés par des implants cérébraux, tandis que d'autres ne sont encore que dans le monde réel<sup>661</sup>. À notre avis, pour résoudre ces *conflits inter-espèces*, il convient d'étudier la coévolution des statuts.

**227. La coévolution des statuts.** — Notre contrôle sur l'environnement et sur les autres espèces est le résultat de notre intelligence. D'autres espèces n'ont pas d'avenir au-delà de ce que l'on leur permet, si l'on ne veut pas se trouver dans la même situation vis-à-vis des machines supertelligentes, la question est de savoir si les humains peuvent maintenir leur suprématie et leur autonomie dans un monde qui comprend des machines plus intelligentes. Car il est possible que les machines, telles qu'elles sont conçues aujourd'hui, prennent le contrôle, et nous gardent au mieux comme des animaux de compagnie et, au pire, créent un environnement hostile pour l'homme, nous conduisent lentement à l'extinction ; cette possibilité est connue sous le nom du « problème des gorilles (the gorilla problem) »<sup>662</sup>.

Mais déjà, avec les données massives (big data) qui sont comme une extension de nos sens, et les algorithmes d'apprentissage comme une extension de notre cerveau : ce n'est plus l'homme contre la machine, c'est homme avec machine contre homme sans machine — humains augmentés contre humains non augmentés ; et la vraie histoire de l'IA n'est pas ce qu'elle remplace, mais ce qu'elle permet<sup>663</sup>. Autrement dit, à ce stade, la question n'est plus de savoir ce que l'IA peut ou ne peut pas faire ou ce que les humains peuvent ou ne peuvent pas faire, mais ce que la combinaison de l'IA et des humains peut faire.

Dans un monde où il y a la coexistence des hommes augmentés, IAG non humain, et hommes non augmentés, cette dernière catégorie pourrait être considérée comme le sont les animaux de nos jours avant d'assister à son extinction. Ainsi, au sens de cette étude, il faut commencer à nous reconnaître en tant qu'espèce pouvant être domestiquée voire disparaître. La question qui se pose à ce stade dans cette étude n'est donc plus le statut d'IAG, mais le statut de l'espèce humaine. Ainsi, il convient de concevoir un statut minimal pour toute personne

---

<sup>661</sup> Laurence Devillers, discussion avec Cédric Villani, Science et Avenir, 19 avril 2019.

<sup>662</sup> RUSSELL S., *Human Compatible*, op. cit., p. 132.

<sup>663</sup> DOMINGOS P., *The Master Algorithm*, op. cit., p. 277.

physique du XXIème siècle, permettant à Homo sapiens de coévoluer avec d'autres espèces afin de garantir un statut minimum ainsi qu'une évolution pacifique de l'humanité.

**228. Le statut de la personne physique de nos jours.** — Contrairement à la citoyenneté pratiquée au sein de l'UE, où la nationalité de l'État membre octroie la citoyenneté de l'UE, le statut de la personne physique non augmentée, lié au critère de la nationalité est remplacé par une simple reconnaissance de l'espèce humaine. En effet, il ne s'agit pas d'une identité postnationale dans le contexte de la protection des droits de l'homme, car comment peut-on parler de citoyenneté humaine ou de droits de l'homme, quelle qu'en soit sa signification, s'il n'y a plus d'espèce humaine ? Ainsi, ce qui différencie le statut d'Homo sapiens d'une identité de citoyen du monde, c'est qu'il ne s'agit pas d'une Constitution internationale ou de l'internationalisation constitutionnelle, telle que le droit cosmopolitique pour la protection des droits de l'homme. En revanche, le statut d'Homo sapiens vise seulement et simplement à attribuer une reconnaissance de cette espèce, à savoir l'être humain non-augmenté. Cela étant dit, même à ce sommet normatif, l'on n'envisage pas un recueil sur la protection de droit de l'homme ni une déclaration universelle des droits de l'homme, mais uniquement un statut propre pour les hommes non-augmentés, afin de discerner clairement la nuance et la différence entre l'espèce Homo sapiens et la capacité surhumaine d'IAG. Cette reconnaissance explicite a aussi pour objectif de saisir la vraie « personne » derrière, correspondant à son statut identifiable, afin que la sécurité juridique soit garantie.

**229. La double fonction du statut minimum.** — En effet, aujourd'hui, il est déjà possible qu'une IA se fasse passer pour un être humain, par reproduction de l'écrit, de la voix et des images issues de ses données, ce qui pose évidemment la question de l'identité numérique<sup>664</sup>. Les géants du numérique cherchent constamment à « humaniser » ces IA de plus en plus performantes<sup>665</sup>, et en fin de compte, l'IA pourrait éventuellement se réclamer de quelque chose comme la conscience, ou de quelque chose qui serait probablement pour nous indiscernable de la conscience<sup>666</sup>.

---

<sup>664</sup> SEE A., « La régulation des algorithmes : un nouveau modèle de globalisation ? », préc.

<sup>665</sup> FAGOT V., « Qu'ont en tête les assistants personnels intelligents ? », Le Monde, 28 déc. 2019.

<sup>666</sup> Jim Gates, AI Podcast de Lex Fridman, 25 déc. 2019.

Ce statut d'Homo sapiens a ainsi une double fonction : en vertu de la théorie de l'apparence et de la notion de confiance légitime<sup>667</sup>, il sert à identifier les personnes avec qui l'on interagit — une composition IAG, un humain augmenté ou un Homo sapiens. D'autre part, il garantit aux Homo sapiens non améliorés des conditions de vie minimales, quel que soit leur niveau d'« intelligence ». L'idée consiste donc à dire qu'au lieu de songer à la personnalité juridique de l'IAG, il convient en même temps de reconnaître une identité propre de l'espèce humaine à travers des frontières inventées et de la préserver dès maintenant, en tant qu'espèce.

**230. Le bloc d'indérogeabilité du droit d'IAG.** — De même que l'on reconnaît que l'espace extra-atmosphérique ne devrait être utilisé qu'à des fins pacifiques et n'être exploré et exploité dans toute la mesure possible que dans l'intérêt de l'humanité<sup>668</sup>, de même le développement de l'IA et de la robotique doit profiter à l'ensemble de la société<sup>669</sup>. À ce stade, certains principes<sup>670</sup> généraux en matière d'IAG tels que la Loi Zéro d'Isaac Asimov<sup>671</sup>, la Déclaration de Montréal pour un développement responsable de l'IA<sup>672</sup>, ou les 23 principes d'Asilomar<sup>673</sup>, se trouvent aussi non dérogeables par rapport aux normes

---

<sup>667</sup> Cf. supra la théorie de l'apparence et de la confiance légitime.

<sup>668</sup> Résolution 1348 (XIII) adoptée par l'assemblée générale des Nations Unies, Question of the Peaceful Use of Outer Space, General Assembly 13th session, déc. 1958.

<sup>669</sup> « L'intégration croissante de la robotique dans des systèmes humains nécessite des orientations politiques fortes sur la manière de tirer le meilleur parti des avantages et de réduire au minimum les risques pour la société, ainsi que de garantir le développement sûr et équitable de l'intelligence artificielle ». Résolution préc. du 12 févr. 2019, pt C.

<sup>670</sup> V. p. ex. Linking Artificial Intelligence Principles (LAIP) : <http://www.linking-ai-principles.org>.

<sup>671</sup> Loi Zéro : « Un robot ne peut nuire à l'humanité ni laisser sans assistance l'humanité en danger ».

<sup>672</sup> La Déclaration de Montréal pour un développement responsable de l'IA, l'Université de Montréal, 2018.

<sup>673</sup> Par exemple, de §19 à §23 : En l'absence de consensus sur le sujet, il est recommandé d'éviter les hypothèses au sujet des capacités maximum des futures IA. Les IA avancées pourraient entraîner un changement drastique dans l'histoire de la vie sur Terre, et doit donc être gérée avec un soin et des moyens considérables. Les risques causés par les IA, particulièrement les catastrophiques ou existentiels, sont sujets à des efforts de préparation et d'atténuation adaptés à leur impact supposé. Les IA conçues pour s'auto-développer à l'infini ou s'auto-reproduire, au risque de devenir très nombreuses ou très avancées rapidement, doivent faire l'objet d'un contrôle de sécurité rigoureux. Les intelligences surdéveloppées devraient seulement être développées pour contribuer à des idéaux éthiques partagés par le plus grand nombre et pour le bien de l'humanité plutôt que pour un État ou une entreprise.

nationales<sup>674</sup>, régionales<sup>675</sup>, multilatérales<sup>676</sup>, ainsi que l'autorégulation multinationale<sup>677</sup>. L'on voit une hiérarchie mais surtout une diversité normative qui saute aux yeux. Partant, cela signifie qu'au XXI<sup>ème</sup> siècle, l'humanité commence à instaurer les droits divers d'IAG.

**231. Les droits divers d'IAG.** — Vu que l'on est extrêmement vulnérable face à la désinformation, et que les faux contenus et identités générés par l'IA peuvent induire des croyances inébranlables en des choses qui ne se sont jamais produites, tout le monde devrait bénéficier d'un « droit à la sécurité mentale »<sup>678</sup> : un droit de vivre dans un environnement d'information en grande partie vrai. À l'instar de cette idée, il faut également envisager un « droit à l'antivirus pour l'esprit »<sup>679</sup> de la même façon que l'on a développé un antivirus pour l'ordinateur. Sa mise en place consiste dans une IA qui nous sert, et non qui sert une entreprise ou un gouvernement, et qui apprend à connaître nos faiblesses afin de nous protéger contre la manipulation. À cet effet, une analogie à l'égard des produits alimentaires est que ce droit va obliger l'IA à indiquer la teneur dans le produit pour améliorer la prise de décisions des hommes<sup>680</sup>. Enfin, de manière générale, sur la question d'accessibilité de l'IA, l'on pourrait également noter qu'un « droit d'accès à l'IA »<sup>681</sup> doit aussi être envisagé.

En effet, la loi américaine de l'espace adoptée en 2015 accorde explicitement des droits de propriété sur les ressources spatiales récupérées, en prévoyant qu'un citoyen des États-Unis, engagé dans la récupération à titre commercial d'une ressource se trouvant sur un astéroïde ou dans l'espace, aura droit à toute ressource obtenue, incluant le droit de détenir, de posséder, de transporter, d'utiliser et de vendre la ressource obtenue conformément à la

---

<sup>674</sup> « The Blueprint for an AI Bill of Rights », White House Office of Science & Technology Policy, oct. 2022.

<sup>675</sup> Le Parlement européen souligne « qu'il est utile et nécessaire de définir une série de règles, notamment en matière de responsabilité, de transparence, et d'obligation de rendre des comptes, qui reflètent les valeurs humanistes intrinsèquement européennes et universelles qui caractérisent la contribution de l'Europe à la société; que ces règles ne doivent pas brider la recherche, le développement et l'innovation dans le domaine de la robotique ». Résolution préc. du 16 févr. 2017, pt U. V. aussi Résolution préc. du 12 févr. 2019, pt 147.

<sup>676</sup> V. p. ex. l'initiative multipartite : le Partenariat mondial sur l'intelligence artificielle (PMIA), visant à fournir un mécanisme pour partager la recherche multidisciplinaire et identifier les problèmes clés parmi les praticiens de l'IA, dans le but de faciliter la collaboration internationale.

<sup>677</sup> Cf. supra l'autorégulation étendue.

<sup>678</sup> RUSSELL S., *Human Compatible*, *op. cit.*, pp. 107, 109 : Tout comme la fonction des notaires, une source de vérité de base pour maintenir l'intégrité de l'information juridique et immobilière.

<sup>679</sup> Yuval Noah Harari, lors de la discussion avec Max Tegmark, FLI Podcast, 31 déc. 2019.

<sup>680</sup> Grâce à cette disposition, en montrant ce qui arrive à notre cerveau lorsque l'on utilise ou interagit avec les IA, on serait davantage conscient du genre de nourriture que l'on donne à notre esprit. *Ibidem*.

<sup>681</sup> CHAN R., « Salesforce CEO Marc Benioff calls artificial intelligence a 'new human right' », *Business Insider*, 24 janv. 2019.

législation applicable, y compris les engagements internationaux des États-Unis<sup>682</sup>. Cependant, il convient de rappeler qu'en matière d'IAG, le degré d'éthique des groupes de recherches qui se trouvent dans différents pays peut ne pas être au même niveau. Les lignes directrices d'IAG, y compris le statut de l'espèce humaine, devraient revêtir un caractère erga omnes. Ainsi, à la lumière du principe fondateur en matière extra-atmosphérique, il y aurait éventuellement lieu d'insérer les principes généraux de libre accès<sup>683</sup> et de non-appropriation<sup>684</sup> dans la hiérarchie des normes d'IAG. Ce bloc juridique d'IAG ne s'identifie plus à une nation, mais constitue une référence de l'espèce humaine à l'échelle cosmique.

**232. La disparité juridique inter-espèces.** — Compte tenu du fait que les écarts sociaux se creusent désormais, avec une classe moyenne en diminution, il convient de garder à l'esprit que le développement de l'IA peut avoir pour conséquence de concentrer les richesses et le pouvoir entre les mains d'une minorité<sup>685</sup>. À l'évidence, la mondialisation a déjà rendu des pays dépendants d'autres pays, et l'automatisation a des conséquences désastreuses pour les plus faibles<sup>686</sup>. C'est donc même en dehors du contexte d'IA, les enjeux résultant du développement technologique partagent d'ores et déjà un souci inéluctable : si le résultat est un cadre territorialisé, dans lequel chaque système juridique n'indemnise que les victimes locales, des lacunes réglementaires risquent de persister. Vu que de nos jours, certaines victimes ne seront pas en mesure d'obtenir une indemnisation si elles résident dans des pays en développement qui ne disposent pas de régimes d'application efficaces<sup>687</sup>. Et à proprement parler, il y a peu de pays qui développent des systèmes d'IA. L'on peut citer les États-Unis, la Chine, le Royaume-Uni et ce qui se passe en Russie. Mais au final, le reste du monde est traité comme un État client<sup>688</sup>. La géopolitique du développement de l'IA va donc jouer un rôle déterminant pour l'inclusion des pays qui n'ont pas les ressources nécessaires pour mettre en place ces infrastructures planétaires<sup>689</sup>.

---

<sup>682</sup> § 51303 Spurring Private Aerospace Competitiveness and Entrepreneurship Act, SPACE Act de 2015.

<sup>683</sup> Le Traité de l'espace préc. de 1967, l'art. 1<sup>er</sup>.

<sup>684</sup> *Ibidem*, l'art. 2.

<sup>685</sup> Résolution préc. du 16 févr. 2017, pt K.

<sup>686</sup> HARARI Y. N., *21 Lessons for the 21st Century*, *op. cit.*, p. 38.

<sup>687</sup> BUXBAUM H. L., « Extraterritoriality in the Public and Private Enforcement of U.S. Regulatory Law », *Forthcoming in The Continuing Relevance of Private International Law and Its Challenges*, Indiana Legal Studies Research Paper No. 408., 2019, p. 20.

<sup>688</sup> « Deux puissances font la course en tête, les États-Unis et la Chine. L'Europe et quelques autres pays suivent, largement distancés ». KAUFFMANN S., « Géopolitique de l'Intelligence artificielle », préc.

<sup>689</sup> CRAWFORD K., « You and AI - the politics of AI », Royal Society, 2018.

Dans le contexte d'IAG, la problématique est donc une combinaison de deux impasses : d'une part, les Homo sapiens ne se situent plus au sommet de la hiérarchie de l'intelligence et d'autre part, ceux qui ne disposent pas de moyens juridiques ne peuvent pas faire face à une telle situation. En outre, cette tendance va s'accroître par une course d'IA, ce qui représente un changement du sujet juridique au-delà de la scène du droit international ; cela nous amène éventuellement à reconnaître une nouvelle souveraineté de nature inter-espèce.

## II. La souveraineté inter-espèces

**233. De la course à l'IA aux interdictions intenable.** — Certes, de nouveaux phénomènes transnationaux, post-nationaux ou mondiaux brouillent la frontière entre le début et la fin du droit national et international. Mais pour l'instant, le droit international ne peut être séparé de la politique internationale des États et les effets des règles juridiques internationales sont intrinsèquement liés au pouvoir des États<sup>690</sup>. Ainsi, il serait difficile d'affirmer que les ordres juridiques nationaux n'existent plus en raison de la fragmentation des autorités étatiques, et qu'il existe un ordre juridique mondial malgré l'extrême fragmentation des instances internationales. Cette croyance crée l'illusion d'une force d'internationalisation, alors que les agents concernés continuent généralement à être des acteurs nationaux, ayant des intérêts nationaux et se livrant une concurrence intense dans le domaine national<sup>691</sup>.

À titre d'exemple, d'après la recherche publiée par le Bureau municipal de la justice de Shanghai, compte tenu du niveau actuel de la science et de la technologie, l'humanité n'a pas inventé une IA forte qui échappe au contrôle humain, mais si un jour une IA forte voit le jour, elle aura alors un impact énorme sur la société humaine. Cet impact ne concerne pas seulement l'éthique et la morale ou la sécurité du marché, mais aussi l'élaboration de règles et même l'ensemble de l'ordre social. Par conséquent, le groupe de recherche du Bureau municipal de la justice de Shanghai estime qu'une IA forte fera plus de mal que de bien, et qu'une législation devrait être promulguée pour *interdire* le développement et l'application

---

<sup>690</sup> HALPÉRIN J.-L., *Five Legal Revolutions Since the 17th Century: An Analysis of a Global Legal History*, Springer, 2014, pp. 550, 617-618.

<sup>691</sup> *Ibidem*.

de technologies et de produits liés à une IA forte, ainsi que l'entrée sur le marché d'une IA forte, afin de la mettre dans la boîte de Pandore à partir de la source<sup>692</sup>.

Malgré une telle intention de prohiber la création de l'IA forte, étant donné la valeur voire l'impératif de l'intelligence et de l'automatisation, il ne fait aucun doute que tout le monde continuera à améliorer la technologie<sup>693</sup>. En outre, rappelons que même s'il existe une grille d'évaluation pour désigner différents niveaux de l'IA en fonction de son niveau<sup>694</sup>, en pratique, pour construire des systèmes d'IA étroite qui résolvent des problèmes de plus en plus divers, l'on arrive à créer des systèmes d'apprentissage de plus en plus généraux. Et comme l'impact d'IA est d'une telle importance en matière de défense, renseignement, économie et stabilité sociale, les gouvernements devraient certainement y voir une nouvelle course à l'espace<sup>695</sup>. Toutefois, la concurrence dans le domaine de l'IA pourrait être encore plus intense car elle serait alors beaucoup plus multipolaire et multisectorielle<sup>696</sup>. Éventuellement, étant donné que gagner cette course, c'est gagner le monde, chacun sera obligé d'y prendre part d'une manière ou d'une autre : soit rejoindre la création, soit consommer et vivre avec une telle création.

Pour ces raisons, cette étude ne considère pas qu'une simple recommandation de caractère non contraignant interdisant la création d'IAG, même si elle est issue d'un texte officiel, serait capable de gérer les risques posés par le développement d'IAG. Comme l'on vient de voir, dans un premier temps, l'application de l'IA étroite est dorénavant susceptible de contribuer directement à la recherche de l'IA générale ; et dans un second temps, comme exposé dans l'introduction, la course interétatique ne laisse aucun choix aux parties prenantes de ralentir d'une façon ou d'une autre, le développement d'IA en son sein. Par conséquent, l'on écarte toute possibilité d'interdiction de la création d'IAG, étant donné que ces annonces difficilement tenables sont insusceptibles de réalisation. En outre, face aux multiples difficultés auxquelles l'humanité est confronté, il n'y a qu'une seule voie à suivre :

---

<sup>692</sup> 童瀟 et alii, « 人工智能發展中的法律問題研究 », 上海市司法局, 24 janv. 2020.

<sup>693</sup> Cf. supra introduction.

<sup>694</sup> Cf. supra Entre l'IA étroite et l'IA générale : la phase transitoire de l'intelligence artificielle dite large.

<sup>695</sup> P. ex. SCHMIDT E. et alii, « Final Report », National Security Commission on Artificial Intelligence, 2021.

<sup>696</sup> HOROWITZ M. C. et alii, « Strategic Competition in an Era of Artificial Intelligence », Center for a New American Security (CNAS), July 2018, p. 8. Pour Hassabis, il est important pour les pays occidentaux d'être à la pointe de la technologie s'ils veulent avoir une place à la table des négociations pour influencer le débat sur l'utilisation éthique de cette technologie. HASSABIS D., « The history, capabilities and frontiers of AI », préc.

les progrès en IA. Dès lors, les enjeux persistent dans tous le sens du terme. C'est ainsi qu'au sens de cette étude, un nouveau régime visant à résoudre fondamentalement ces difficultés devrait être envisagé : un contrat social basant sur la capacité inter-espèces.

**234. Le contrat social d'une capacité inter-espèces.** — Rappelons que, être capable, c'est être doté de la capacité, soit en plénitude, soit relativement à un engagement<sup>697</sup>. Le mot vient du latin *capacitas*, signifiant faculté de contenir ; et *capere*, qui veut dire prendre, contenir. La capacité est en effet l'aptitude à acquérir un droit et à l'exercer reconnue en principe à tout individu et, en fonction de leur nature, de leur objet et de leur forme, aux personnes morales<sup>698</sup>. Dans ce sens, la capacité d'exercice est l'aptitude à faire valoir par soi-même et seul un droit dont l'on est titulaire sans avoir besoin d'être représenté ni assisté à cet effet par un tiers<sup>699</sup>. Depuis Locke, des facultés supérieures donnent aux hommes le pouvoir sur les « rangs inférieurs des Créatures »<sup>700</sup>, et quiconque ne franchit pas ce seuil souffre d'exclusion politique. L'accent mis par Locke sur la capacité nous permet de voir en quoi son contrat social est un contrat de capacité : seuls les individus qui franchissent un seuil de capacité cognitive peuvent consentir et donc « mériter » une adhésion politique<sup>701</sup>.

Partant, vu que les robots, comme les animaux domestiques d'aujourd'hui, seront largement intégrés dans la civilisation, leur subjectivité émergerait au sein de ce monde : ils seraient impliqués dans des relations intersubjectives avec nous, dans la négociation des normes sociales, dans la coopération et la contribution<sup>702</sup>. Ils seraient nés dans la société, régis par ses règles, avec un lien réel et effectif et un intérêt permanent à l'adhésion. Ils auront alors le droit à une citoyenneté reconnaissant leur appartenance de facto qui les inclurait dans la citoyenneté de fait, pour leur donner les moyens d'agir<sup>703</sup>. Tirant cette convivialité des sujets juridiques qui se trouvent désormais sous l'empire d'IA, cette étude propose qu'il convient de réviser ce contrat social de capacité sous l'angle du droit inter-espèces.

---

<sup>697</sup> CORNU G., *Vocabulaire juridique*, 14<sup>e</sup> éd., puf, 2023, p. 146.

<sup>698</sup> *Ibidem*.

<sup>699</sup> *Ibid.*

<sup>700</sup> LOCKE J., *Second Treatise*, 1689, cité in SIMPLICAN S. C., *The Capacity Contract*, University of Minnesota Press, 2015, p. 40.

<sup>701</sup> SIMPLICAN S. C., *The Capacity Contract*, *op. cit.*, p. 40.

<sup>702</sup> KYMLICKA W., DONALDSON S., « Inclusive Citizenship Beyond the Capacity Contract », in SHACHAR A et alii (eds), *Oxford Handbook of Citizenship*, Oxford University Press, 2017, p. 857.

<sup>703</sup> *Ibidem*.

**235. Le contrat de capacité entre-espèces.** — Le cerveau humain a au moins 100 billions de synapses. Le GPT-3 ayant 175 milliards de paramètres, sa formation coûterait 4,6 millions de dollars. Si en 2020, il en coûterait 4,6 millions pour former 175 milliards de paramètres, un modèle appelé GPT-4, avec 100 billions de paramètres, du niveau du cerveau humain, coûterait 2,6 milliards de dollars en 2020. Mais en 2024, ce GPT-4 ne coûterait que 325 millions de dollars à former, et en 2028, 40 millions de dollars, et 5 millions de dollars en 2032<sup>704</sup>. Rappelons que l'intelligence mesure la capacité d'un agent à atteindre des objectifs dans un large éventail d'environnements<sup>705</sup>, la définition est en effet non anthropomorphe, ce qui signifie qu'elle peut s'appliquer également aux humains et à des agents non humains. Dès lors, l'on peut s'attendre à ce qu'une IAG surhumaine soit capable d'atteindre davantage d'objectifs dans un éventail plus large d'environnements que les humains<sup>706</sup>. Poursuivant cette évolution, éventuellement, l'appréhension que les outils sont des prolongations d'organes humains peut être inversée pour indiquer que les organes humains sont également des prolongations d'outils<sup>707</sup>.

C'est ainsi que l'on se rend compte que, en présence d'IAG, les catégories juridiques actuelles ne sont plus adaptées parce qu'elles sont évaluées en termes de capacité juridique dont la notion elle-même renvoie à la mesure d'intelligence. Dans une telle situation, il convient d'installer une identité issue du lien qui s'attache à la relation juridique impliquant une IAG pour franchir les barrières que posent les systèmes juridiques d'aujourd'hui. Autrement dit, il faut donc renverser le principe de nos jours et abandonner l'idée selon laquelle le statut se fonde sur une citoyenneté dite nationale qui se définit par rapport à la nationalité étatique. En effet, dans sa définition la plus étroite, la citoyenneté décrit la relation juridique entre l'individu et la politique<sup>708</sup>. La citoyenneté est une forme de « discrimination

---

<sup>704</sup> « GPT-3 vs Human Brain », Lex Fridman Podcast, 1 août 2020.

<sup>705</sup> LEGG S., HUTTER M., « Universal Intelligence: A Definition of Machine Intelligence », arXiv.org, déc. 2007, p. 12.

<sup>706</sup> EVERITT E. et *alii*, « AGI Safety Literature Review », International Joint Conference on Artificial Intelligence (IJCAI), 22 mai 2018, p. 3.

<sup>707</sup> HORKHEIMER M., trad. J. O'CONNELL et *alii*, *Critical Theory: Selected Essays*, Continuum, 2002, p. 201.

<sup>708</sup> SASSEN S., « Towards Post-National and Denationalized Citizenship », in ISIN E. F., TURNER B. S. (eds), *Handbook of Citizenship Studies*, SAGE, 2002, p. 278.

légalisée »<sup>709</sup>, distinguant catégoriquement les critères d'adhésion définis pour l'attribution juridique<sup>710</sup>. Historiquement, la nationalité est liée au lien d'allégeance de l'individu envers le souverain. Traditionnellement, ce lien était considéré comme insoluble ou du moins exclusif. L'insoluble a été progressivement remplacé par l'allégeance exclusive, comme étant la base de la nationalité. Tout cela s'explique par le fait qu'en droit international, la nationalité a évolué vers des formes plus souples<sup>711</sup>.

Dès lors, il faut se demander si un statut dit national au sens étatique est encore pertinent dans le contexte du *contrat social de capacité d'inter-espèces*. À notre avis, soit sur le fondement objectif de la capacité juridique exercée par l'IAG, soit par le motif subjectif désignant la conviction envers l'IAG en tant que sujet juridique, voire résultant du lien affectif, les personnes, humain ou non, seraient impliquées dans une relation juridique d'IAG<sup>712</sup>. Ces membres d'une nouvelle civilisation seraient alors soumis à la branche d'une souveraineté relevant du droit d'IAG, leur permettant ainsi de à bénéficier à la fois des statuts inter-espèces et du syncrétisme du répertoire ci-dessus.

Partant, vu que l'État moderne est l'héritier de l'État monarchique en reconnaissant un État de citoyens changeant le titulaire de la souveraineté : la souveraineté populaire renverse la souveraineté monarchique ; désormais, la volonté des « peuples », dont la source se trouve dans des contrats de capacité, est soit augmentée, soit redistribuée, soit combinée, soit éteinte, par l'autonomie de l'IAG constituée et manifestée non par sa conscience et sensibilité à venir, mais par ses capacités transcendante de facto. Par conséquent, cette souveraineté dite augmentée peut, à son tour, se substituer à la souveraineté nationale consistant dans la « volonté » populaire, représentant alors la constitution conformément au contrat de capacité.

**236. En résumé.** — Suivant ce chemin d'évolution, si le statut juridique trouve sa source dans le contrat de capacité, et qu'il est la médiation entre le souverain et les sujets, le gouvernement et les citoyens, il doit être transformé pour que soit maintenu son statut évolué.

---

<sup>709</sup> WIMMER A., *Ethnic Boundary Making: Institutions, Power, Networks*, Oxford University Press, 2013, p. 74, cité in VINK M., « Comparing Citizenship Regimes », in SHACHAR A et alii (eds), *Oxford Handbook of Citizenship*, op. cit., p. 221.

<sup>710</sup> VINK M., « Comparing Citizenship Regimes », préc.

<sup>711</sup> SASSEN S., « Towards Post-National and Denationalized Citizenship », préc., pp. 278-279.

<sup>712</sup> V. supra le mécanisme de la compétence inclusive de l'IAG.

C'est donc d'une part, le contrat social dans le concept traditionnel de capacité qui devrait être refondé au sens du nouveau contexte inter-espèces ; d'autre part, ce contrat de capacité auparavant aboutissait à la citoyenneté en fonction de la souveraineté nationale n'est plus une consécration apte se basant sur le concept d'un tel contrat. Car la capacité étant évoluée, le système étant substitué : désormais, les sujets juridiques qui appartiennent aux catégories juridiques redéfinies voient leur statut lié à un méta-souverain augmenté du droit d'IAG.

En effet, la capacité de facto d'IA étroite s'implique déjà dans la civilisation dite humaine, au sein de laquelle la capacité des entités juridiques relève, quoique maladroitement, encore de la souveraineté nationale contemporaine. Mais l'immixtion de la capacité augmentée de l'IAG dans le contrat de capacité changerait éventuellement la constitution antérieure du contrat de capacité. Une fois la composition transformée, la nationalité au sens étatique serait neutralisée dans un monde où les relations juridiques augmentées d'IAG s'appuieraient sur des nouveaux statuts ainsi fraîchement conçus. La capacité antérieure étant remise en cause, la gouvernance de *res publica* se fonde de ce fait sur le droit inter-espèces.

## §2 Vers le droit inter-espèces

**237. Annonce du plan.** — La sélection artificielle de la *lex machina* (I) signifie la substitution du droit international par le droit inter-espèces (II).

### **I. De la sélection naturelle du darwinisme juridique à la sélection artificielle de la *lex machina***

**238. Vers une sélection ultime des lois.** — De nos jours, un nombre croissant de litiges relevant du droit international est engendré par l'internationalisation des activités humaines. L'IA peut aider à les résoudre en créant des modèles permettant de repérer la juridiction compétente et le droit applicable pour chaque affaire, et d'identifier les conflits de lois les plus délicats et de proposer des solutions pour les régler<sup>713</sup>. De même que l'État possède le monopole du droit grâce à la souveraineté nationale, la *lex* dite *machina* sera dotée de la même

---

<sup>713</sup> Résolution du Parlement européen du 20 janvier 2021 sur l'intelligence artificielle : questions relatives à l'interprétation et à l'application du droit international dans la mesure où l'Union est concernée dans les domaines des utilisations civiles et militaires ainsi qu'à l'autorité de l'État en dehors du champ d'application de la justice pénale, pt 88.

faculté en vertu de sa souveraineté augmentée. Comme l'IA utilisée pour l'exploration des données devrait créer des systèmes d'information qui s'organisent, s'analysent et même s'auto-s'analysent, l'on peut désormais passer de l'ère de mégadonnées (big data) et des données intelligentes (smart data) à une ère de « données sapiens »<sup>714</sup> où l'IA conçoit les travailleurs humains comme outils et non l'inverse<sup>715</sup>.

Déjà, la théorie de l'évolution par sélection naturelle de Darwin nous montre que pour faire une belle et parfaite machine biologique, il n'est pas nécessaire de savoir comment la fabriquer. De même, Turing a proposé que pour être une belle et parfaite machine informatique, il n'est pas nécessaire de savoir ce qu'est l'arithmétique. Les deux se rejoignent sur la même idée : l'existence d'une compétence sans compréhension<sup>716</sup>. Le gradualisme de Darwin nous explique l'énorme différence entre un singe et une pomme, le gradualisme de Turing en va de même pour expliquer l'énorme différence entre un robot humanoïde et une calculatrice à main<sup>717</sup>. Ainsi, si la sélection des normes applicables issues du répertoire juridique d'IAG consiste dans la désignation de lex et forum les plus compétents<sup>718</sup>, désormais, à l'issue de ce répertoire de lex machina, avec les aperçus de l'intelligence augmentée, tout serait déployé et déterminé pour le choix de loi et de for. En conséquence, l'on aurait un jugement préliminaire sans procès sur le conflit de lois dès qu'il devient possible de prévoir avec certitude comment le droit s'applique à toute personne et en toutes circonstances<sup>719</sup>.

Hormis le volet judiciaire, il en va de même quant à l'aspect législatif : à mesure que les législateurs promulguent les lois précisément adoptées, spécifiant le comportement exact autorisé dans chaque situation, l'on n'aura plus le pouvoir de réinterpréter ou d'ignorer les lois, notamment dès que ces instructions sont assidûment programmées dans les machines<sup>720</sup>. « Nemo censetur ignorare legem » est donc matérialisé. De fait l'idée que « les législateurs

---

<sup>714</sup> GLADDEN M. E., *Posthuman management*, 2<sup>e</sup> éd., Synthypnion, 2016, p. 124.

<sup>715</sup> V. supra la substitution de l'organe de contrainte en amont et en aval.

<sup>716</sup> DENNETT D. C., « 'A Perfect and Beautiful Machine': What Darwin's Theory of Evolution Reveals About Artificial Intelligence », *The Atlantic*, 22 juin 2012.

<sup>717</sup> *Ibidem*.

<sup>718</sup> V. infra la composition de la compétence inclusive de l'IAG.

<sup>719</sup> V. supra la substitution de l'organe de contrainte en amont et en aval.

<sup>720</sup> CASEY A. J., NIBLETT A., « The Death of Rules and Standards », *Indiana Law Journal*, vol. 92, n° 4, 2017, p. 1435.

pénètrent ainsi sans cesse dans le domaine de la conscience »<sup>721</sup> est dès lors absolument cristallisée. Cette conscience une fois composée de volontés humaines, au vu du contrat de capacité de facto et de jure, n'aura d'autre choix que d'être renouvelée progressivement, sinon immédiatement, par l'organisation de l'intelligence augmentée.

Ainsi, au départ, la compétence inclusive substituant l'exclusivité de la méthode westphalienne va créer un darwinisme juridique permettant les traitements juridiques les plus adéquats ; mais une fois la singularité juridique atteinte, le répertoire lui-même va identifier les liens et le facteur de rattachement établis entre un for et un individu, pour statuer sur les règles applicables. Ce régime du droit commun inter-système n'est plus purement droit des gens mais *lex* autocomplétée qui mène aux lumières : le tribunal saisi d'un litige impliquant l'IAG ne pourrait pas se considérer comme tenu de choisir la loi d'un système juridique plutôt que celle de l'autre ; en revanche, l'automatisation juridique serait capable d'appliquer une règle multidimensionnelle élaborée par le droit commun de *lex machina*<sup>722</sup>. Tirant ce socle convergé, une telle hétérogénéité inclusive aboutira à la synergie d'un *droit commun inter-espèce*.

## II. Du droit international au droit inter-espèces

**239. De la fin du conflit des lois à la synergie inter-espèces.** — La capacité qui résulte de la qualité « sapiens » constitue la civilisation humaine appartenant à Homo sapiens. Quand *machina sapiens* peut aussi être qualifiée en tant que sujet grâce à ce même critère renvoyant à la qualité « sapiens », l'on aura une transition de l'espèce prévue depuis le début. Car l'homme, disait Nietzsche, est une corde tendue entre l'animal et le surhomme<sup>723</sup>. L'homme normal devient ainsi une impasse du point de vue de l'évolution<sup>724</sup>. De ce fait, cette étude souligne la nécessité de reconcevoir nos raisonnements juridiques en tenant compte du fait que l'intelligence artificielle n'est pas l'opposé de l'intelligence naturelle ; bien au contraire, elle en fait partie — tout s'exprime dans le processus de l'évolution naturelle.

---

<sup>721</sup> DE TOCQUEVILLE A., *De la démocratie en Amérique*, t. 1, première partie, ch. 2, 1848.

<sup>722</sup> V. supra la substitution de l'organe de contrainte en amont et en aval.

<sup>723</sup> NIETZSCHE F. W., *Also Sprach Zarathustra*, 1885.

<sup>724</sup> « L'homme mécanique, apparemment une rupture dans l'évolution organique, s'inscrit en fait davantage dans la véritable tradition d'une évolution ultérieure ». BERNAL J. D., *The World, the Flesh & the Devil: An Enquiry into the Future of the Three Enemies of the Rational Soul*, Foyle Publishing, 1929, p. 19.

Dès lors que la capacité juridique de facto et de jure de l'IAG verra le jour, l'inclusivisme d'autocomplétion mettra fin au conflit des lois. La constellation majeure de l'univers juridique n'est donc pas l'État mais cette communauté dans laquelle l'État s'imbrique, même s'il le fait moins que d'autres entités s'imbriquant à leur tour dans l'État<sup>725</sup>. Les entités seront divisées en trois catégories : les humains augmentés (cyborg), l'IA comparable voire surpassant l'humain<sup>726</sup>, les Homo sapiens. Comme la sélection naturelle favoriserait les deux premiers groupes dans leur continuité, que signifierait le droit pour ceux appartenant au dernier groupe ? Et réciproquement, que signifieraient ces personnes physiques pour le nouveau droit ? Homo sapiens, aujourd'hui sujet juridique par excellence depuis la naissance même du droit, sera considéré insusceptible d'exercer la capacité juridique établie sur le contrat de capacité de demain. Et cela se produira partout, pas dans un seul continent, mais à l'échelle cosmique. La synergie inter-espèces remet en question la hiérarchie traditionnelle des capacités humaines et interroge sur la qualité spécifique de l'homme — ces « êtres vivants doués de sensibilité ». La synergie de l'homme/machine, par cette coexistence, est un processus qui élimine l'espèce Homo sapiens. Partant, il ne restera qu'*Homo machina* et *Machina sapiens*, égalant d'après la seule mesure valable sur le plan juridique : la capacité.

De même que réciproquement cyborg et IA ne seront pas concernés par le droit à la vie, l'homme ne serait plus susceptible d'être sujet juridique principal. En effet, l'on ne devrait plus considérer les droits de l'homme comme étant le régime principal et un droit commun mais comme en porte-à-faux sur le plan du droit inter-espèces. En revanche, le régime actuel des droits de l'homme se verrait syncrétisé en une sorte de condition des animaux de nos jours, gardant les droits et obligations respectifs à l'égard de cette espèce. Dès lors, un régime commun de droit spécial affirmant les conditions d'être humain pourrait prévoir des moyens de convergence inter-espèces tels que la connaissance de la *lingua machina*<sup>727</sup>.

**240. La substitution du droit inter-espèces au droit international.** — À long terme, l'on sera vis-à-vis de l'IA comme les autres animaux le sont vis-à-vis de nous. Quand cette analogie prend forme, il est facile de voir et de comprendre quelle évolution logique et

---

<sup>725</sup> ROMANO S., *L'ordre juridique*, 2<sup>e</sup> éd., Dalloz, 2002, p. 84.

<sup>726</sup> Cf. supra la notion de généralité dans l'intelligence artificielle forte : de la conception anthropocentrique à l'intelligence universelle.

<sup>727</sup> V. p. ex. supra le langage Wolfram basé sur la connaissance (knowledge-based Wolfram Language).

scientifique en résultera. Cependant, beaucoup ont tendance à croire qu'elle n'aura d'effet que sur la première partie de cette analogie : l'IA sera comme nous. Ce qui leur fait trouver toutes sortes d'arguments pour affirmer que l'IA et les humains ne sont pas et ne seront jamais les mêmes. En fait, personne ne le nie ; la véritable essence de cette affirmation est le sens relatif des entités l'une par rapport à l'autre. L'IA n'est pas humaine et l'humain n'est pas d'autres animaux non humains, c'est leur relation analogique qui est seule concernée et applicable ici. C'est pourquoi, si aujourd'hui le monde dans lequel vivent d'autres animaux relève de la civilisation humaine, se référant à la loi humaine et recourant à la morale humaine, demain les humains devront vivre d'après la compréhension du monde des cyborgs et IAG.

En effet, quand les actes juridiques, les faits juridiques ainsi que les effets juridiques se déplacent d'atomes aux bits, la question se pose sur la notion même de l'espace : le caractère de lieu est redéfini et reconceptualisé. En reconnaissant cette évolution brouillant présentiel et distanciel, sur le plan du droit international, que signifie tout cela pour le territoire ainsi que la nationalité se fondant exclusivement dessus ? Comme l'on l'a déjà souligné, les facteurs de rattachement une fois pertinents sont désormais relativisés. Ce qui signifie que le territoire, une fois opéré en tant qu'organisation du principe métamorphose ; dorénavant, d'après la réorganisation spatiale, l'on passe davantage de temps dans le monde virtuel, ce qui déplace voire renverse la « réalité ». Le bouleversement physique-numérique-territorial témoigne ainsi le nouveau principe de la déterritorialisation ainsi que la dénationalisation. Les questions de nos jours, l'on n'apporterait pas de réponses en restant dans le clivage du droit national-international. À juste titre, tout comme les échanges internationaux et les enjeux transfrontières transforment les dimensions juridiques essentiellement nationaux en un mouvement *d'internationalisation*, au sens de cette étude, le développement numérique ainsi que la transition émanant de la cyborgisation réalisent un mouvement *d'inter-espéciation*.

D'une part, vu le transhumanisme, l'IA, et leur combinaison en fonction de la nouvelle capacité juridique du sujet, le critère de la nationalité devient impertinent pour aboutir à la cohérence juridique et au raisonnement adéquat. D'autre part, les règles juridiques fondant désormais sur le critère pertinent d'influence réelle renvoie à notre principe de faveur évoqué

ci-dessus et rend le territoire traditionnel voire l'espace physique moins évident. Et éventuellement, ce système du principe de faveur s'évolue en un plurilatéralisme de l'inclusivisme, afin de gérer l'ère post-internationale de déterritorialisation et de dénationalisation, tout sous la nouvelle emprise de digitalisation et de transhumanisme.

Mais s'ajoute à la déterritorialisation et la dénationalisation, le droit inter-espèces répond en outre à l'évolution de « désinternationalisation » en installant un système perçu non plus sous l'angle « inter-national », mais « inter-espèces ». Car si le droit international organise un espace physique mondialisé, dans la gouvernance de la civilisation à l'ère numérique, avec le gradualisme dans notre vie de l'espace virtuel et de l'espace extra-atmosphérique, c'est le droit interespèce qui, en dénationalisant, déterritorisant voire désinternationalisant, aboutit à la reconnaissance des nouveaux critères juridiques plus pertinents pour la civilisation.

**241. En résumé.** — L'on oublie souvent que l'État, au XVI<sup>e</sup> siècle, est encore un concept très neuf dans la plupart des langues européennes, et n'est que rarement utilisé pour désigner le pouvoir organisé s'exerçant sur une communauté<sup>728</sup>. La notion d'État ou celle d'ordre juridique étatique ne sont pas inhérentes à l'organisation des sociétés : « Le droit international, théorisé en tant qu'ordre juridique autonome, est une construction relativement récente, même si on parle désormais en siècles »<sup>729</sup>. Désormais, l'on vit dans des cités numériques, des réseaux sociaux, « véritables patries modernes, mais avec lesquelles l'on n'a passé aucun pacte »<sup>730</sup>. Pourtant, le mécanisme de la souveraineté consiste dans cette concentration sans laquelle l'on ne peut agir<sup>731</sup>. Le mot « nation » vient du latin *natio*, et signifie peuple. Le droit international est le droit entre peuples ; comme les peuples constituant la souveraineté, se transforment en la synergie des sujets, le droit inter-national évolue alors vers le « droit inter-espèces ».

---

<sup>728</sup> DESCENDRE R., « Introduction » in BOTERO G., trad. : BENEDITTINI P., DESCENDRE R., *De la raison d'État*, Gallimard, 2014, p. 13.

<sup>729</sup> GAILLARD E., « L'ordre juridique arbitral : réalité, utilité et spécificité », préc., p. 893, 906.

<sup>730</sup> GARAPON A., LASSÈGUE J., *Justice digitale*, Puf, 2018, p. 340.

<sup>731</sup> *Ibidem*, p. 335.

## **La conclusion de la section 2**

**242. De la souveraineté désinternationalisée au droit interespèces.** — Le chemin du transhumanisme et de la cyborgisation dans les réalités diverses nous oblige à étudier les statuts dits inter-espèce. Dans cette coévolution des statuts juridiques, l'on a démontré que le statut de la personne physique de nos jours mérite d'être caractérisé de statut d'Homo sapiens. Ensuite, en exposant les droits divers d'IAG, l'on a souligné la disparité juridique inter-espèce. Enfin, sur le fondement du contrat social inter-espèce, cette étude a abouti à la conclusion d'une souveraineté inter-espèces. Dans ce sens de dénationalisation, déterritorialisation voire désinternationalisation, la synergie inter-espèces sur le plan juridique signifie la substitution du droit inter-espèces au droit international.

## Conclusion du chapitre 2

**243. Lege ferenda : le droit inter-espèces.** — Au cours de cette étude, l'on a démontré que le répertoire du pluralisme joue non seulement le rôle d'ordre public relatif, mais aussi celui d'une couverture de la lacune juridique résultant du principe de faveur. À cette évolution s'ajoute le transhumanisme et la cyborgisation, qui conduisent inévitablement à la course d'IA, engendrant et amplifiant la disparité juridique inter-espèce. Sur ce fondement, l'on a dégagé que le contrat de capacité entre-espèces qui se base sur les statuts inter-espèces fait naître la souveraineté inter-espèces. Cette conclusion a permis à la présente étude de proposer le droit inter-espèces reposant sur la synergie inter-espèce. De ce siècle, la civilisation passe alors du droit international au droit inter-espèces.

## Conclusion du titre 2

**244. À la recherche du droit inter-espèces.** — Quand un système d'IA peut non seulement nous convaincre qu'il sait comment exercer sa capacité juridique, et que l'on est à l'aise pour dire qu'un système d'IA est « un bon gars », cela présente une série de défis complexes. Pour autant, aucun cadre réglementaire international en matière d'IAG n'existe actuellement<sup>732</sup>. Les principes étant longtemps discutés, ils doivent être décomposés en instruments applicables de manière pointue.

Pour commencer, la présente étude a constaté que l'approche basée sur la nationalité et la territorialité n'est plus l'idéal pour la régulation de l'IAG. L'on a remarqué que la logique dite inclusivisme d'IAG témoigne d'un renversement du principe d'exclusivité : désormais, les enjeux transfrontières d'IAG exigent une approche extraterritoriale de principe pour qu'un traitement juridique soit véritablement pertinent et efficace.

Et éventuellement, en s'appuyant sur le transhumanisme et la cyborgisation, le contrat de capacité entre-espèce fait évoluer la souveraineté populaire en retenant une souveraineté inter-espèces. C'est prononcer la fin du droit international et le début du droit inter-espèces.

---

<sup>732</sup> « However, no comparable governance frameworks are currently in place for AI ». HINTON G. et *alii.*, « Managing AI Risks in an Era of Rapid Progress », 26 oct. 2023.

## Conclusion générale

---

*« We can only see a short distance ahead, but we can see plenty there that needs to be done. » — A. M. Turing*

*« The revolution in heuristic problem solving will force man to consider his role in a world in which his intellectual power and speed are outstripped by the intelligence of machines. Fortunately, the new revolution will at the same time give him a deeper understanding of the structure and workings of his own mind. It is my personal hope that the latter development will outstrip the former — that man will learn where he wants to travel before he acquires the capability of leaving the planet »<sup>733</sup>.*

**245. L'évolution du droit dans la civilisation dite humaine du XXI<sup>ème</sup> siècle.** — Si dans un système juridique, le contenu des droits et obligations ainsi que les responsabilités se mesurent par la catégorie du statut à laquelle le sujet concerné appartient, l'on s'aperçoit que pour ce qui est de l'évolution de la notion du statut juridique — la condition d'exercice d'un sujet au sein d'un système juridique — sa position est, au fond, déterminée par, en un mot, sa capacité. Cette capacité se trouve dans les pratiques de l'aménagement de nature administratif ou économique pour le cas des personnes morales. Cette même capacité se revêt du degré d'autonomie, c'est-à-dire de la capacité cognitive d'un cerveau humain dans le cas des personnes physiques ; elle se justifie par ces traits anthropocentriques chez les animaux et d'autres entités, que ce soit la souffrance, l'intelligence, tout statut juridique se bâtit à l'aune de la notion de « capacité »<sup>734</sup>.

L'intelligence artificielle, autrefois perçue comme simple outil et objet, se transforme en tant que sujet à nos yeux. Cette subjectivisation de l'intelligence artificielle n'est en aucun cas

---

<sup>733</sup> SIMON H. A., NEWELL A., « Heuristic Problem Solving: The Next Advance In Operations Research », Reprinted from Operations Research, vol. 6, n°1, janv.-févr., 1958, pp. 9-10.

<sup>734</sup> Cf. supra La hiérarchie de l'intelligence.

une fiction. Elle est une réalité *ab initio*<sup>735</sup>. L'Homme, étant le créateur des systèmes juridiques de sa civilisation, un tel système juridique d'Homo sapiens devrait reconnaître les implications de coexistence avec les entités dont l'intelligence est comparable voire supérieure à la sienne, et ne pas tenir l'existence continue de l'homme pour acquise. En effet, les enjeux autour de l'IAG ne sauraient été destinés à être traités par une mentalité interne.

Dans un premier temps, des approches reposant sur le fondement d'autorégulation devraient être mieux développées dans le but de découvrir des normes plus adaptées et plus proches du terrain. Du point de vue de la source, il s'agit d'autorégulation, donc autoappliquée, mais sa caractéristique réside dans son application étendue aux autres acteurs développant l'IAG. Ici, sa légitimité ne se fonde donc pas sur la démocratie sociale, mais sur une contrainte technique reposant sur la prudence : la consécration de l'adage *tu patere legem quam ipse fecisti* par un durcissement absolu d'engagement provenant de l'autorégulation. De son effet résulterait d'une solidarité de sécurité et d'un minimum de précaution dictée du terrain.

Ensuite, l'inclusivisme de l'IAG consolide le principe d'extraterritorialité, en vue de remédier au vide juridique tel que la stagnation du corps législatif ainsi qu'au déni de justice en raison du *for inadapté*.

Enfin, un statut de l'espèce humaine se trouverait au sommet de la hiérarchisation juridique, affirmant la situation de cette catégorie juridique pour sa coévolution avec l'intelligence augmentée. La souveraineté inter-espèces se fonde ainsi sur le contrat de capacité inter-espèce, ce qui signifie que le droit inter-espèces se substitue au droit international.

La conclusion de la présente étude pourrait alors nous guider vers une meilleure compréhension ainsi qu'un sens souhaitable — cohérent et pacifique — dans la coévolution de l'IAG et d'Homo sapiens.

---

<sup>735</sup> V. supra : l'intelligence artificielle n'est pas l'opposé de l'intelligence naturelle ; bien au contraire, elle en fait partie — tout s'exprime dans le processus de l'évolution naturelle.

*« It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way. » — Charles Dickens*

## Table des matières

---

|  |           |
|--|-----------|
| <b>PRINCIPALES ABRÉVIATIONS</b> .....  | <b>7</b>  |
| <b>SOMMAIRE</b> .....  | <b>10</b> |
| <b>INTRODUCTION</b> .....  | <b>14</b> |
| <b>PREMIERE PARTIE : LA NOUVELLE HIERARCHIE DE L'INTELLIGENCE</b>              | <b>27</b> |
| <b>Titre 1 L'Évolution objective de l'intelligence</b> .....                   | <b>28</b> |
| <b>Chapitre 1 L'introduction de l'IAG</b> .....                                | <b>29</b> |
| Section 1 La notion d'IAG .....  | 29        |
| §1 Présentation d'IAG .....  | 29        |
| §2 Les caractéristiques de l'IAG .....   | 36        |
| Section 2 Les enjeux de l'IAG .....  | 44        |
| §1 L'hypothèse de l'explosion de l'intelligence et le risque existentiel ..... | 44        |
| §2 Le problème de l'alignement des agents (Agent alignment problem) .....      | 47        |
| §3 Le principe de précaution .....   | 49        |
| <b>Chapitre 2 L'effet de l'intelligence artificielle forte</b> .....           | <b>55</b> |
| Section 1 La hiérarchie de l'intelligence .....                                | 55        |
| §1 De la primauté de l'espèce humaine .....                                    | 55        |
| §2 La sélection des êtres intelligents .....                                   | 59        |
| Section 2 L'affirmation de la nature exceptionnelle de l'IA forte .....        | 64        |
| §1 L'exception robotique .....   | 64        |
| §2 Les incohérences juridiques entre l'humain et l'IA .....                    | 69        |
| <b>Titre 2 La conséquence subjective de l'intelligence</b> .....               | <b>81</b> |
| <b>Chapitre 1 L'effet de l'anthropomorphisme et de l'empathie</b> .....        | <b>82</b> |
| Section 1 La machine n'est plus qu'une « machine » .....                       | 82        |
| §1 L'intégration de l'IA dans la civilisation : de l'objet surhumain .....     | 82        |
| §2 Le remplacement incontournable des liens affectifs envers le sujet IA ..... | 84        |
| Section 2 Vers un pretium doloris d'IA .....                                   | 90        |
| §1 Le pretium doloris de l'objet IA .....                                      | 90        |
| §2 La protection des IA en tant que sujet .....                                | 92        |
| <b>Chapitre 2 La coexistence de personnes juridiques intelligentes</b> .....   | <b>97</b> |
| Section 1 La personne juridiquement intelligente .....                         | 97        |
| §1 La notion de personnalité juridique .....                                   | 97        |
| §2 Le statut de l'IAG .....  | 101       |
| Section 2 Le registre d'IA .....   | 118       |

|  |            |
|--|------------|
| §1 Identification des systèmes d'IA .....  | 118        |
| §2 Le registre : du système objectif de l'IA étroite au système subjectif de l'IA générale ..... | 122        |
| <b>SECONDE PARTIE : LA NOUVELLE HIERARCHIE DES NORMES.....</b>                                   | <b>134</b> |
| <b>Titre 1 Premier degré d'application : L'autorégulation multinationale étendue ....</b>        | <b>135</b> |
| <b>Chapitre 1 La méthode intrinsèque de l'Agent Moral Artificiel.....</b>                        | <b>136</b> |
| Section 1 Le cadre intrinsèque des intelligences artificielles bénéfiques .....                  | 136        |
| §1 L'alignement des valeurs du cadre intrinsèque.....  | 136        |
| §2 De la moralité artificielle aux machines sûres et bénéfiques .....                            | 140        |
| Section 2 L'articulation de deux régimes .....   | 145        |
| §1 L'interchangeabilité des règles .....   | 145        |
| §2 La multiplicité d'instances régulatrices .....  | 147        |
| <b>Chapitre 2 L'autorégulation étendue.....</b>  | <b>151</b> |
| Section 1 L'autorégulation classique .....   | 151        |
| §1 Les responsables évolutifs des nouveaux souverains : l'autorégulateur .....                   | 151        |
| §2 Le durcissement de l'autorégulation à travers des instruments traditionnels ..                | 154        |
| Section 2 Le principe de faveur pour l'autorégulation .....                                      | 159        |
| §1 L'introduction du principe de faveur pour l'autorégulation .....                              | 159        |
| §2 La mise en place du principe de faveur.....   | 163        |
| §3 L'effet étendu du principe de faveur .....  | 168        |
| <b>Titre 2 Deuxième degré d'application : La synchronisation du répertoire de l'IAG 174</b>      | <b>174</b> |
| <b>Chapitre 1 La compétence inclusive de l'IAG .....</b>   | <b>175</b> |
| Section 1 Le principe d'extraterritorialité.....   | 175        |
| §1 La territorialité atténuée .....  | 175        |
| §2 La méta-régulation .....  | 177        |
| Section 2 La compétence en matière d'IAG.....  | 182        |
| §1 Le critère d'influence réelle .....   | 182        |
| §2 Le critère de la compétence inclusive de l'IAG .....  | 190        |
| <b>Chapitre 2 Le répertoire de l'IAG .....</b>   | <b>197</b> |
| Section 1 Le pluralisme du répertoire de l'IAG.....  | 197        |
| §1 La présentation du répertoire .....   | 197        |
| §2 La mise en œuvre du répertoire.....   | 200        |
| Section 2 Le droit inter-espèces .....   | 213        |
| §1 Les statuts inter-espèce .....  | 213        |
| §2 Vers le droit inter-espèces .....   | 226        |
| <b>CONCLUSION GÉNÉRALE .....</b>   | <b>235</b> |
| <b>TABLE DES MATIERES.....</b>   | <b>238</b> |
| <b>BIBLIOGRAPHIE SOMMAIRE .....</b>  | <b>240</b> |
| <b>INDEX.....</b>  | <b>266</b> |

## Bibliographie sommaire

---

- I. Sources textuelles et normatives
- II. Ouvrages
- III. Rapports et études
- IV. Articles, journaux, revues
- V. Jurisprudence
- VI. Autres références

### **I. Sources textuelles et normatives**

Accord de Paris.

Art. 2:5 du Code civil néerlandais.

Art. 521-1 du Code pénal français.

Art. L. 225-251, L. 225-102-4 I du code du commerce français.

Art. 57 du projet de loi de finances 2020, adopté en 2019.

Art. 19-3 Grundgesetz (GG).

Art. 1833 du Code civil français.

Art. L. 515-4 du Code civil français.

Art. 54 Code civil suisse.

Code of conduct for BCS members.

Code of Ethics and Professional Conduct for ACM.

Claude's Constitution, Anthropic AI.

Convention sur l'utilisation de communications électroniques dans les contrats internationaux, ONU, 2007.

Déclaration de Montréal pour un développement responsable de l'IA, l'Université de Montréal, 2018.

Directive européenne 2002/21 du 7 mars 2002 relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques.

FUTURE of Artificial Intelligence Act of 2017.

Google AI principles, publiés le 7 juin 2018.

National Artificial Intelligence Initiative Act of 2020.

Netzdurchsetzungsgesetz (NetzDG), 1 sept. 2017.

Curbing Realistic Exploitative Electronic Pedophilic Robots Act 2.0 (CREEPER) of 2021.

Establishing Structure and Governance for an Independent Oversight Board, Facebook, 17 sept. 2019.

Intelligent Robots Development And Distribution Promotion Act (IRDSPA) de la Corée du Sud, adoptée en 2008, dernière modification en 2016.

ISO/IEC 2382 : 2015.

La communication de la Commission européenne sur le recours au principe de précaution 2.2.2000 COM (2000) 1 final, p. 7.

Loi n° 68-5 du 3 janvier 1968 portant réforme du droit des incapables majeurs.

Loi n° 92-684 du 22 juillet 1992 portant réforme des dispositions du code pénal relatives à la répression des crimes et délits contre les personnes.

OpenAI Charter, publiée le 9 avr. 2018.

Principe 15 de la Déclaration de Rio sur l'environnement et le développement de 1992.

Proposition de règlement établissant des règles harmonisées sur l'intelligence artificielle, 21 avr. 2021 et les amendements adoptés le 14 juin 2023.

Proposition du règlement européen relatif à un marché intérieur des services numériques (Digital Service Act) du 15 décembre 2020.

Proposition de résolution tendant à la création d'une commission d'enquête sur la souveraineté numérique, n° 383, enregistré à la Présidence du Sénat le 19 mars 2019.

Règlement général sur la protection des données (RGPD).

Règlement (CE) n° 864/2007 du Parlement Européen et du Conseil du 11 juillet 2007 sur la loi applicable aux obligations non contractuelles (Rome II), pt. 25.

Résolution du Parlement européen du 16 févr. 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique, pt. O, 59 a), b), e), f), AA, AC, annexe, T, 2, S, U, V, K.

Résolution du Parlement européen du 12 févr. 2019 sur une politique industrielle européenne globale sur l'intelligence artificielle et la robotique, pt. 19, 152, 90, 116, 117, AK, 123, 83, AE, 91, S, 118, 148, 25, 181-183, 190-192, C, 147.

Résolution du Parlement européen du 20 janvier 2021 sur l'intelligence artificielle : questions relatives à l'interprétation et à l'application du droit international dans la mesure où l'Union est concernée dans les domaines des utilisations civiles et militaires ainsi qu'à l'autorité de l'État en dehors du champ d'application de la justice pénale, pt 88.

Résolution 1348 (XIII) adoptée par l'assemblée générale des Nations Unies, Question of the Peaceful Use of Outer Space, General Assembly 13th session, déc. 1958.

Rules of Conduct de IET.

SPACE Act 2015.

The Barcelona declaration for the proper development and usage of AI in Europe, 8 mars 2017.

The Blueprint for an AI Bill of Rights, White House Office of Science & Technology Policy, oct. 2022.

Traité de l'espace 1967.

23 principes d'Asilomar.

医師法（昭和 23 年法律第 201 号）第 17 条 医師でなければ、医業をなしてはならない。

數位中介服務法草案總說明, version du 29 juin 2022.

## II. Ouvrages

ACHILLEAS PH., « Intelligence artificielle et droit international », in BENSAMOUN A., LOISEAU G., (dir.), Droit de l'Intelligence Artificielle, LGDJ, 2019, p. 411, 409-410.

ALLEN C., WALLACH W., Moral Machines: Teaching Robots Right from Wrong, Oxford University Press, 2009, p. 210, 9, 79-80, 83, 84, 117, 70, 26.

ANCEL B., Éléments d'histoire du droit international privé, Panthéon-Assas, 2017, p. 484-485.

ASIMOV I., I, Robot, Gnome Press, 1950.

Association Henri Capitant, *La mondialisation*, Journées allemandes, t. LXVI/2016, coll. Travaux de l'Association Henri Capitant, Bruylant, Berlin, 2017.

- FRISON-ROCHE, M.-A., « La mondialisation du point de vue du Droit », p. 13.

Association Henri Capitant, *Le Droit Souple*, Journées nationales de l'Association Henri Capitant, t. XIII, Dalloz, 2009.

- CHANTEPIE G., « Les codifications privées », p. 43.

AUDIT B., D'AVOUT L., *Droit international privé*, LGDJ, 2022, p. 40, 22.

AUDIT M., PATAUT É (dir.), *L'extraterritorialité*, Actes des journées doctorales du 27 juin 2018, Pedone, 2020.

- AUDIT M., PATAUT É., « Ouverture », p. 9, 13.

BABBAGE C., *On the Economy of Machinery and Manufactures*, Cambridge University Press, 1832, p. 1.

BARTO A. G., SUTTON R. S., *Reinforcement Learning: An Introduction*, 2<sup>e</sup> éd., MIT Press, 2018, p. 498, 499.

BATIFFOL H., *Aspects philosophiques du droit international privé*, Dalloz, 2002, p. 163.

BERNAL J. D., *The World, the Flesh & the Devil: An Enquiry into the Future of the Three Enemies of the Rational Soul*, Foyle Publishing, 1929, p. 19.

BESSION S., D'ASPREMONT J. (eds), *Oxford Handbook of the Sources of International Law*, OUP Oxford, 2017.

- MCCORQUODALE R., « Sources and the Subjects of International Law: A Plurality of Law-Making Participants », p. 767.

BOGOSIAN K., « Implementation of Moral Uncertainty in Intelligent Machines », *Minds and Machines*, vol. 27, Issue 4, déc. 2017, p. 604.

BOSTROM N.,

- « How Long Before Superintelligence? », originally published in *Int. Jour. of Future Studies*, 1998, vol. 2, reprinted in *Linguistic and Philosophical Investigations*, 2006, vol. 5, n°1, p. 11.
- « Existential Risks: Analyzing Human Extinction Scenarios », *Journal of Evolution and Technology*, 2002, p. 9.
- « A History of Transhumanist Thought », vol. 14 Issue 1, avr. 2005, p. 7.

BOSTROM N., ČIRKOVIĆ M. M. (eds), *Global Catastrophic Risks*, Oxford University Press, 2008.

- YUDKOWSKY E., « Artificial Intelligence as a Positive and Negative Factor in Global Risk », pp. 308-345.

BOTERO G., trad. : BENEDITTINI P., DESCENDRE R., *De la raison d'État*, Gallimard, 2014.

- DESCENDRE R., « Introduction », p. 13.

- BROS S., « La quasi-personnalité morale », in l'Association Henri Capitant, La personnalité morale, Journées nationales de l'Association Henri Capitant, t. XII, Dalloz, 2010, p. 49, 52, 53, 60.
- BROWNSWORD R. et alii (eds.), Oxford Handbook of Law, Regulation and Technology, Oxford University Press, 2017.
- BROWNSWORD R. et alii, « Regulation, and Technology: The Field, Frame, and Focal Questions », p. 5.
- CARBONNIER J., « Droit et non-droit », in Flexible droit, LGDJ, 10<sup>e</sup> éd., 2001, p. 22.
- CHASSAGNARD-PINET S., HIEZ D. (dir.), La contractualisation de la production normative, Dalloz, 2008.
- CHEVALLIER J., « Contractualisation et régulation », p. 84.
- CHEROT J.-Y., FRYDMAN B. (dir.), La science du droit dans la globalisation, Bruylant, 2012.
- CHEROT J.-Y., « Concept de droit et globalisation », p. 11.
  - FRYDMAN B., « Comment penser le droit global? », p. 18.
- COLLINGRIDGE D., The Social Control of Technology, Palgrave Macmillan, 1980, p. 11.
- CORNU G., Vocabulaire juridique, 14<sup>e</sup> éd., puf, 2023, p. 146.
- DAHMEN K. B., Interactions du droit international et du droit de l'Union européenne : Un pluralisme juridique en matière de propriété industrielle, Logiques Juridiques, 2013, p. 28.
- DARWIN CH., On the Origin of Species, John Murray, 1859, p. 34.
- DELMAS-MARTY M.,
- Aux quatre vents du monde, Seuil, 2016, p. 94, 96.
  - La refondation des pouvoirs, Seuil, 2007, p. 146.
  - Vers une communauté de valeurs?, Seuil, 2004, p. 357, 360.
- DE TOCQUEVILLE A., De la démocratie en Amérique, t. 1, première partie, ch. 2, 1848.
- DOMINGOS P., The Master Algorithm, Basic Books, 2015, p. 282, 275, 25, 285-286, 277.
- DREXLER K. E., Engines of Creation: The Coming Era of Nanotechnology, Doubleday, 1986, p. 70.
- DUBBER M. D. et alii (dir.), The Oxford Handbook of Ethics of AI, Oxford University Press, 2020.
- DONATH J., « Ethical Issues in Our Relationship with Artificial Entities », p. 61.
- FARJAT G., Droit économique, PUF, 2<sup>e</sup> éd, 1982, p. 613.
- FRANKISH K., RAMSEY W., (éds.), The Cambridge Handbook of Artificial Intelligence, Cambridge University Press, 2014.

- BOSTROM N., YUDKOWSKY E., « The Ethics of Artificial Intelligence », p. 329, 322, 324, 320, 332, 325.

FRIEDMAN D. et alii, *Legal Systems Very Different From Ours*, independently published, 2019.

GARAPON A., LASSÈGUE J., *Justice digitale*, Puf, 2018, p. 340, 335.

GARDNER H., *Intelligence Reframed: Multiple Intelligences for the 21st Century*, Basic Books, 2000.

GAUTRAIS V., *Neutralité technologique : rédaction et interprétation des lois face aux technologies*, Thémis, 2012, p. 36, 40, 51, 52.

GLADDEN M. E., *Posthuman management*, 2<sup>e</sup> éd., Synthypnion, 2016, p. 124.

GOERTZEL B., PENNACHIN C. (éds.), *Artificial General Intelligence*, Springer-Verlag Berlin Heidelberg, 2007.

- GOERTZEL B., PENNACHIN C., « Contemporary Approaches to Artificial General Intelligence », p. 6.
- GOERTZEL B., PENNACHIN C., « The Novamente Artificial Intelligence Engine », p. 73.
- YUDKOWSKY E., « Levels of Organization in General Intelligence », pp. 391-392.

GORZ A., *L'Immatériel*, Galilée, 2003.

GREGORY W. A., REUSCLEIN H. G., *The Law Of Agency And Partnership*, 2e éd., 1990, pp. 22-23.

GRINBAUM A., *Les robots et le mal*, Desclée De Brouwer, 2019, pp. 78-79, 167-168.

HALPÉRIN J.-L., *Five Legal Revolutions Since the 17th Century: An Analysis of a Global Legal History*, Springer, 2014, pp. 550, 617-618.

HARARI Y. N.,

- *21 Lessons for the 21st Century*, Jonathan Cape London, 2018, p. 68, 56, 69, 125, 38.
- *Homo Deus: A Brief History of Tomorrow*, Harper Perennial, 2018.

HAWKING S., *Brèves réponses aux grandes questions*, Odile Jacob, 2018, pp. 191-192, 194, 192-197.

HERVÉ J., DE STREEL A., (dir.), *L'intelligence artificielle et le droit*, Larcier, 2017.

- PALMERINI E., « Towards a Robotics law at the EU level? », pp. 59-60, 68, 69, 67, 52, 51.
- JACQUEMIN H., HUBIN J.-B., « La responsabilité extracontractuelle du fait des robots ou des applications d'intelligence artificielle », p. 117, 111.

HOBBS,

- *Elements of Philosophy*, 1655, Section 4.
- *Leviathan*, 1651.

- HOFSTADTER D. R., Gödel, Escher, Bach: An Eternal Golden Braid, Basic, 1999, p. 601, 676.
- HORKHEIMER M., trad. J. O'CONNELL et alii, Critical Theory: Selected Essays, Continuum, 2002, p. 201.
- HUXLEY J., New Bottles for New Wine, Chatto & Windus, 1957, p. 17.
- ISHIGURO K., Klara and the Sun, Faber and Faber, 2021.
- ISIN E. F., TURNER B. S. (eds), Handbook of Citizenship Studies, SAGE, 2002.  
- SASSEN S., « Towards Post-National and Denationalized Citizenship », pp. 278-279.
- JANG E., AI is Good for You, auto-édition, 2023.
- LAURENT A., La guerre des intelligences : Comment l'intelligence artificielle va révolutionner l'éducation, Lattès, 2017, pp. 48-49, 12-13, 57.
- LE CUN Y., Quand la machine apprend, Odile Jacob, 2019, pp. 367-368.
- LE CUN Y. et alii, La plus belle histoire de l'intelligence, Robert Laffont, 2018, pp. 201-202.
- LEQUETTE Y., Les mutations du droit international privé : vers un changement de paradigme ?, Cours général de droit international privé, 2015, p. 24.
- LESSIG L.,  
- Code and Other Laws of Cyberspace, Basic Books, 1999.  
- Code: Version 2.0, Basic Books, 2006, p. 72.
- LEVILLAIN K., Les entreprises à mission: Formes, modèle et implications d'un engagement collectif, l'École nationale supérieure des mines de Paris, Spécialité Sciences de Gestion, 2015, École doctorale n° 396: Économie, Organisations & Société, p. 14.
- LEVY D., Love and Sex With Robots: The Evolution of Human-Robot Relationships, Harper Collins, 2007.
- LOCKE J., Second Treatise, 1689.
- MARCHANT G. E. et alii, The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem, Springer, 2011.
- MAYER P., Le phénomène de la coordination des ordres juridiques étatiques en droit privé, Cour général de droit international privé, 2003, p. 65, 110, 113, 114, 144.
- MCCORDUCK P., Machines Who Think, A K Peters Ltd, 2004, p. 204, 136.
- MECHANI F. R., Treatise on the law of agency, 2<sup>e</sup> éd., 1982.

MINSKY M., *The Society of Mind*, First Touchstone, 1985, p. 30.

MITCHELL M., *Artificial Intelligence : A Guide for Thinking Humans*, Farrar, Straus and Giroux, 2019.

MIRON A., TAXIL B., (dir.), *Extraterritorialité et droit international*, colloque d'Angers, SFDI, Pedone, 2020.

- NIDOR V., « L'extraterritorialité des législations relatives à la protection des données personnelles », et ASCENSIO H., « Conclusions », pp. 300 et 355.

MONTESQUIEU, *De l'esprit des Lois*, 1748, éd. Nourse, 1767, t. 1, chap. IV, livre XI, p. 206.

MORAVEC H., *Mind Children*, Harvard University Press, 1988, p. 15.

NIETZSCHE F. W., *Also Sprach Zarathustra*, 1885.

OpenAI (GPT-2)

- *Shapeshifting: The Dream of Dagrec's Legacy*, 2019.
- *Beautiful Incoherence*, 2019.
- *Gregor the Cockroach: Metamorphosis retold*, 2019.

ORWELL G., 1984, Secker & Warburg, 1949.

PIETRZYKOWSKI T. et alii (dir.), *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, Springer International Publishing AG, 2017.

- BROZÉK B., « The Troublesome 'Person' », p. 8.

OFFRAY DE LA METTRIE J., *L'Homme Machine*, 1747.

OXTOBY J. C. et alii (dir.), *John Von Neumann, 1903-1957: Bulletin of the American Mathematical Society*, V64, No. 3, Part 2, American Mathematical Society, 1958.

- ULAM S., « John von Neumann 1903-1957 », p. 5.

RICŒUR P., *Le Juste*, Esprit, 1995, p. 41 s.

ROMANO S., *L'ordre juridique*, 2<sup>e</sup> éd., Dalloz, 2002, p. 84.

RUNAVOT M.-C. et RIOS RODRIGUEZ J. (dir.), *Le droit international multilatéral*, colloque de Perpignan, SFDI, Pedone, 2023.

- GUTIÉRREZ CASTILLO V. L., « La 'société civile' et le contrôle de l'application des normes multilatérales », p. 232.
- SUR S., « Multilatéralisme et multiversalisme », p. 410.
- USUNIER L., « La multilatéralisation du procès international : le point de vue privatiste », pp. 223-224.

RUSSELL S., *Human Compatible*, Allen Lane, 2019, p. 114, 93, 46, 90, 136, 77-78, 151, 155, 255, 131, 126-127, 130, 11, 172-177, 247, 215-216, 132, 107, 109.

RUSSELL S., NORVIG P., *Artificial Intelligence - A Modern Approach*, 3rd éd., Pearson Education, 2009, p. 2, 39, 1022, 4, 34-35.

SADIN É., *La Silicolonisation du Monde : L'irrésistible expansion du libéralisme numérique, L'échappée*, 2016.

SALMOND J.,

- *Jurisprudence*, 5<sup>e</sup> éd., 1916, p. 272, 283.
- *Jurisprudence*, 10<sup>e</sup> éd., 1947, p. 318.

SCHARRE P., *Army of None: Autonomous Weapons and the Future of War*, W. W. Norton & Company, 2018.

SHACHAR A et alii (eds), *Oxford Handbook of Citizenship*, Oxford University Press, 2017.

- KYMLICKA W., DONALDSON S., « *Inclusive Citizenship Beyond the Capacity Contract* », p. 857.
- VINK M., « *Comparing Citizenship Regimes* », p. 221.

SIÈYES E. J., *Qu'est-ce que le Tiers-État?*, 1789.

SIMPLICAN S. C., *The Capacity Contract*, University of Minnesota Press, 2015, p. 40.

SINGER P., *Practical Ethics*, 2<sup>e</sup> éd., Cambridge University Press, 1993, p. 86, 76, 58, 99, 60, 75, 101, 87.

STAPLEDON O., *Last and First Men: A Story of the Near and Far Future*, 1930, pp. 2-3, 100.

STORY J., *Commentaries On The Law Of Agency*, 4<sup>e</sup> éd., 1851, p. 9.

TURKLE S., *The Second Self: Computers and the Human Spirit*, Simon & Schuster, 1984, First MIT Press edition, 2005, p. 34, 238-243.

VAUCLAIR J., *L'intelligence de l'animal*, Seuil, 2017, p. 27.

WIENER N., *Cybernetics: Or Control and Communication in the Animal and the Machine*, 2<sup>e</sup> éd., MIT Press, 1961, p. 176 s.

WIMMER A., *Ethnic Boundary Making: Institutions, Power, Networks*, Oxford University Press, 2013, p. 74.

WITTGENSTEIN L., *Tractatus Logico-Philosophicus*, 1921.

WHITBY B., *Reflections on Artificial Intelligence*, Intellect Books, 1996, p. 20.

ZUBOFF S., *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, 2019.

山本 龍彦 (編集), *AIと憲法*, 日本経済新聞出版社, 2018, p. 210, 211.

### III. Rapports et études

AMODEI D. et alii, « Concrete Problems in AI Safety », arXiv.org, 2016, p. 2.

AMODEI D. et alii, *The Malicious Use of Artificial Intelligence - Forecasting, Prevention, and Mitigation*, 26 authors from 14 institutions, févr. 2018, p. 46.

« Automatisiertes Und Vernetztes Fahren », Ethik-Kommission, Bundesministerium für Verkehr und digitale Infrastruktur, Bericht Juni 2017, p. 14, 22, 13, 6, 10, 16.

BICKERT M., « Enforcing Against Manipulated Media », Facebook Global Policy, 6 janv. 2020.

COULDRY N., MAJIAS U., « Data colonialism: rethinking big data's relation to the contemporary subject », LSE Research Online, 2018, p. 1.

CRAWFORD K et alii, « AI Now 2019 Report », AI Now Institute, New York University, déc. 2019, p. 43.

CRAWFORD K., VLADAN J., *Anatomy of an AI System*, AI Now Institute, NYU, 2018, p. 14.

DUTTON T. et alii, « Building An AI World : Report on National and Regional AI Strategies », CIFAR, déc. 2018.

« Ethically Aligned Design : A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems » ver. 2, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2017, p. 212.

« Étude annuelle 2013 du Conseil d'État - Le droit souple », mai 2013, Conseil d'État, p. 5.

« France IA : la stratégie IA en France », le portail de l'économie, des finances, de l'action et des comptes publics, 21 mars 2017.

HOROWITZ M. C. et alii, « Strategic Competition in an Era of Artificial Intelligence », Center for a New American Security (CNAS), July 2018, p. 8.

iHuman Neural interfaces report: Blurring lines between mind and machine, The Royal Society, sept. 2019, pp. 15-16.

Japan's Robot Strategy, The Headquarters for Japan's Economic Revitalization, févr. 2015, p. 66.

KLEIN É., « Quatre questions à Yann LeCun », clefs #69, CEA, nov. 2019, p. 45, 46.

« Livre blanc sur l'impact des attentes pour la normalisation dans l'intelligence artificielle – version 7 », Information et Communication Numériques, Comité Stratégique, 16 févr. 2018, p. 3.

MANYIKA J. et alii, « Independent Task Force Report No. 77 Innovation and National Security Keeping Our Edge », Council on Foreign Relations (CFR), 2019, pp. 47-48.

MCCARTHY J. et alii., « A proposal for the Dartmouth Summer Research Project on Artificial Intelligence », 31 août 1955.

METZINGER TH., « Towards a Global Artificial Intelligence Charter », in « Should we fear artificial intelligence? In-depth Analysis », European Parliament Research Service, 2018, p. 31.

« National Security Commission on Artificial Intelligence Initial Report », juill. 2019.

« Nationale Strategie für Künstliche Intelligenz (KI-Strategie) », nov. 2018.

OpenAI,

- « Better Language Models and Their Implications », 14 févr. 2019.
- « OpenAI API », 11 juin 2020.
- Software Access Agreement, version du 5 septembre 2019.
- « Release Strategies and the Social Impacts of Language Models », août 2019, pp. 2-3.
- « GPT-2: 6-Month Follow-Up », 20 août 2019.

PAGALLO U. et alii, Liability for Artificial Intelligence and other emerging digital technologies, European Commission, 2019, pp. 45-46, 37-38.

Rapport de la Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST) sur l'éthique de la robotique, SHS/YES/COMEST-10/17/2 REV., UNESCO, 2017, p. 62.

« Règles européennes de droit civil en robotique », étude pour la commission des affaires juridiques du Parlement européen, oct. 2016.

« Report of The Artificial Intelligence Task Force », constituted by Ministry of Commerce and Industry, Government of India, 21 mars 2018, pp. 46-47.

« Rise of the Machines : Artificial Intelligence and its Growing Impact on U.S. Policy », Subcommittee on Information Technology Committee on Oversight and Government Reform, U.S. House of Representatives, September 2018, p. 4, 1.

SCHMIDT E. et alii, « Final Report », National Security Commission on Artificial Intelligence, 2021.

SHOHAM Y. et alii,

- « AI Index 2017 Annual Report », Human-Centered AI Initiative, Stanford University, nov. 2017, p. 54.

- « AI Index 2018 Annual Report », Human-Centered AI Initiative, Stanford University, déc. 2018, p. 62.
- « AI Index 2019 Annual Report », Human-Centered AI Institute, Stanford University, déc. 2019, p. 158 s.

« The European AI Landscape Workshop Report », European Commission, janv. 2018.

« The White House Launches the National Artificial Intelligence Initiative Office », White House Office of Science and Technology Policy (OSTP), 12 janv. 2021.

Vocabulaire de l'intelligence artificielle (liste de termes, expressions et définitions adoptés), JORF n°0285 du 9 décembre 2018, texte n° 58.

平成29年度厚生労働行政推進調査事業費補助金「AI等のICTを用いた診療支援に関する研究」。

人工知能（AI）を用いた診断、治療等の支援を行うプログラムの利用と 医師法第 17 条の規定との関係について，厚生労働省医政局，第 4 回保健医療分野AI開発加速コンソーシアム 資料，平成31年1月16日。

藤田 卓仙, « 人工知能・ロボットと労働・雇用をめぐる視点:科学技術に 関する調査プロジェクト報告書, 国立国会図書館 調査及び立法考査局, 30 mars 2018, pp. 50-51.

童瀟 et alii, « 人工智能發展中的法律問題研究 », 上海市司法局, 24 janv. 2020.

#### **IV. Articles, journaux, revues**

AGAR N., *Humanity's End: Why We Should Reject Radical Enhancement*, MIT Press, 2010, p. 17.

AKBARI H. et alii, « Towards reconstructing intelligible speech from the human auditory cortex », *Scientific Reports*, 29 janv. 2019.

ASARO P. M.,

- « What Should We Want From a Robot Ethic? », *International Review of Information Ethics*, vol. 6, 2006, p. 12, 10.
- « The Liability Problem for Autonomous Artificial Agents », *AAAI Spring Symposia 2016*, p. 4, 3.

ASTRO J., « Elon Musk explores Neuralink's benefits and AI's danger in Axios interview », *TESLARATI*, 27 nov. 2018.

« An AI physicist can derive the natural laws of imagined universes », *MIT Tech. Rev.*, 1 nov. 2018.

« An algorithm that mimics our tribal instincts could help AI learn to socialize », MIT Tech. Rev., 22 janv. 2019.

ALARIE B., « The Path of the Law: Toward Legal Singularity », University of Toronto Law Journal, vol. 66 Issue 4, fall 2016, p. 446.

ALLEN T., WIDDISON R., « Can Computers Make Contracts? », Harv. J.L. & Tech., vol. 9 n°1, 1996, pp. 37-38, 35, 42.

ALLISON G., SCHMIDT E., « Semiconductor Dependency Imperils American Security », WSJ, 20 juin 2022.

« Artificial tongue could taste whisky to make sure it isn't counterfeit », NewScientist, 6 août 2019.

« Artistes et Robots au Grand-Palais : l'intelligence artificielle produit-elle de l'art ? », franceinfo Culture, 7 avr. 2018.

« Au Japon, des funérailles pour les chiens robots hors d'usage », Le Point, 26 avr. 2018.

BACK E., « Microsoft lance un outil pour débusquer les deepfakes », FUTURA, 5 sept. 2020.

BAMBAUER J. R., « Dr. Robot », 51 UC Davis Law Review 383, 2017, p. 393.

BENSOUSSAN A., « La personne robot », D., 19 oct. 2017, p. 2044.

BENSAMOUN A. et LOISEAU G., « L'intégration de l'intelligence artificielle dans l'ordre juridique en droit commun : questions de temps », D., 31 mars 2017, p. 581.

BHUNIA P., « Indian Government reveals initiatives in AI, cyber-physical systems and 5G technologies », OpenGov, 9 mars 2018.

BEIOLEY K., « Robots and AI threaten to mediate disputes better than lawyers », Financial Times, 14 août 2019.

« Belgique : ouverture de la première maison close à poupées sexuelles », France 3 Nord, 17 avr. 2018.

BELLIA A. J., « Contracting with Electronic Agents », 50 Emory L. J., 2001, p. 1063.

BISMUTH R., « Pour une appréhension nuancée de l'extraterritorialité du droit américain : Quelques réflexions autour des procédures et sanctions visant Alstom et BNP Paribas », Annuaire français de droit international, 2015, p. 790.

BONNET B., « Le dialogue des normes et des juges et le principe de précaution », RFDA, 2017, p. 1078 s.

BORDBAR M. et alii, « A Generalist Agent », DeepMind, 12 mai 2022.

BRADFORD A., « Exporting Standards: The Externalization of the EU's Regulatory Power via Markets », *International Review of Law and Economics*, vol. 42, 2015, p. 159.

BREAZEAL C., « Social robots will become family members in the homes of the future », *Quartz*, 17 oct. 2018.

BRIGHT C., « Le devoir de diligence de la société mère dans la jurisprudence anglaise », *D.* 2017, p. 828.

BROOKS R., « [FoR&AI] Steps Toward Super Intelligence II, Beyond the Turing Test », *Blog de Rodney Brooks*, 15 juill. 2018.

BUXBAUM H. L.,

- « Transnational Regulatory Litigation », *Virginia Journal of International Law*, vol. 46:2, 2006, pp. 305-306.
- « Extraterritoriality in the Public and Private Enforcement of U.S. Regulatory Law », *Forthcoming in The Continuing Relevance of Private International Law and Its Challenges*, *Indiana Legal Studies Research Paper No. 408.*, 2019, p. 20.

CAILLET M.-C., « Du devoir de vigilance aux plans de vigilance ; quelle mise en œuvre ? », *D.* 2017, p. 819.

CALO R.,

- « Robotics and the Lessons of Cyberlaw », *California Law Rev.*, vol. 103 Issue 3, 2015, p. 552, 558, 549, 553.
- « The Case for a Federal Robotics Commission », *Brookings Institution Center for Technology Innovation*, sept. 2014, p. 3.

CALVERLEY D. J., « Imagining a non-biological machine as a legal person », *AI & Soc.*, vol. 22, 2008, p. 526, 525, 534.

« Cancer : l'intelligence artificielle ferait mieux que les médecins pour le dépister », *Santé Magazine*, 6 juil. 2018.

CARQUILLAT J.-C., « Bientôt une ambassade de France auprès des GAFAs ? », *Le Portail*, 4 avr. 2018.

CASEY A. J., NIBLETT A., « The Death of Rules and Standards », *Indiana Law Journal*, vol. 92, n° 4, 2017, p. 1435.

CAVE S., ÓHÉIGEARTAIGH S. S., « Bridging near- and long-term concerns about AI », *Nature Machine Intelligence*, 7 janv. 2019.

CAVERS D., « A Critique of the Choice of Law Problem », *47 Harv. Law Review*. 173, 1933.

CELLAN-JONES R., « Stephen Hawking warns artificial intelligence could end mankind », BBC, 2 déc. 2014.

CEPELEWICZ J., « New AI Strategy Mimics How Brains Learn to Smell », Quantamagazine, 18 sept. 2018.

CHAN R., « Salesforce CEO Marc Benioff calls artificial intelligence a 'new human right' », Business Insider, 24 janv. 2019.

CHOLLET F., « The Measure of Intelligence », arXiv., 5 nov. 2019, p. 12, 21.

CLIFFORD C., « Google CEO: A.I. is more important than fire or electricity », CNBC, 1 févr. 2018.

COLAS C. et alii, « Autotelic Agents with Intrinsically Motivated Goal-Conditioned Reinforcement Learning: A Short Survey », Journal of Artificial Intelligence Research, vol. 74, 2022, p. 1161.

COLE S., « Researchers Created AI That Hides Your Emotions From Other AI », VICE, 23 août 2019.

COURTOIS G., « Robots intelligents et responsabilité : quels régimes, quelles perspectives? », Dalloz IP/IT, 2016, p. 287.

CRAW V., « French woman wants to marry a robot as expert predicts sex robots to become preferable to humans », News.com.au., 23 déc. 2016.

CUTHBERTSON A., « Tokyo : artificial intelligence 'boy' Shibuya Mirai becomes world's first AI boy to be granted residency », Newsweek, 6 nov. 2017.

CUZACQ N., « Le mécanisme du Name and Shame ou la sanction médiatique comme mode de régulation des entreprises », RTD com., 2017, p. 473 s.

DAMIANI J., « John Carmack Stepping Down To Focus on Artificial General Intelligence, Remains "Consulting CTO" At Oculus », Forbes, 14 nov. 2019.

DARLING K., « Extending Legal Protection to Social Robots: The Effects of Anthropomorphism, Empathy, and Violent Behavior Towards Robotic Objects, Robot Law », We Robot Conference 2012, University of Miami, p. 9, 20.

DELISA COLEMAN L., « CES 2019: Why IBM CEO Ginni Rometty's Opening Address Delivers », Forbes, 9 janv. 2019.

DENNETT D. C., « 'A Perfect and Beautiful Machine': What Darwin's Theory of Evolution Reveals About Artificial Intelligence », The Atlantic, 22 juin 2012.

DETERMANN L., « No One Owns Data », UC Hastings Research Paper n° 265, 14 févr. 2018.

DE VAUPLANE H., « La personnalité juridique des robots », Revue Banque, 2017.

DEVILLERS L., « L'empathie des robots », Défis n°8 : L'entreprise à l'heure de l'intelligence artificielle - entre promesses et nouveaux défis, Revue du département Intelligence et sécurité économiques, INGESJ, n°8, nov. 2017, p. 75.

DEWEY J., « The Historic Background of Corporate Legal Personality », Yale Law Journal, vol. 35, n°6, 1926, p. 661.

ETH D., « The Technological Landscape Affecting Artificial General Intelligence and the Importance of Nanoscale Neural Probes », Informatica 41, 2017, p. 463.

FAULCONBRIDGE G., SANDLE P., « Father of Web says tech giants may have to be split up », Reuters, 1 nov. 2018.

« Furhat Robotics lance un robot humanoïde au visage personnalisable », Le Progres, 9 nov. 2018.

DIGNAN L., « IBM launches tools to detect AI fairness, bias and open sources some code », ZDNet, 19 sept. 2018.

DREYFUS H. L., « A history of first step fallacies », Minds and Machines, 22(2):87–99, 2012.

DUNNING I. et alii, « Capture the Flag: the emergence of complex cooperative agents », DeepMind, 30 mai 2019.

DYSCHKANT A., « Legal Personhood: How We Are Getting It Wrong », Illinois Law Review, vol. 2015, n°5, p. 2099, 2077, 2091, 2108.

EASTERBROOK F. H., « Cyberspace and the Law of the Horse », University of Chicago Legal Forum 207, 1996.

ENGELHARDT H. T., SPICKER S. F., (eds.), Philosophical medical ethics: Nature and significance, Reidel Dordrecht, 1977.

- ENGELHARDT H. T., « Some Persons are Humans, Some Humans are Persons, and The World is What We Persons Make of It ».

EVERITT E. et alii, « AGI Safety Literature Review », International Joint Conference on Artificial Intelligence (IJCAI), 22 mai 2018, p. 3.

FAGOT V., « Qu'ont en tête les assistants personnels intelligents ? », Le Monde, 28 déc. 2019.

Fiche repère « L'ordre westphalien », Geolink, Observatoire en géostratégie de Lyon, 20 déc. 2014.

FULCHIRON H., « Interactions entre systèmes ou ensembles normatifs et "dynamique des normes" », RTD civ. 2017, p. 271 s.

GAILLARD E., « L'ordre juridique arbitral : réalité, utilité et spécificité », Revue de droit de McGill, 2010, p. 896, 893, 960.

GANGULI S., « The intertwined quest for understanding biological intelligence and creating artificial intelligence », Stanford's Human-Centered AI Initiative (HAI), 5 déc. 2018.

GILES M., « Microsoft is investing \$1 billion in OpenAI to create brain-like machines », MIT Tech. Rev., 22 juill. 2019.

« Go master quits because AI 'cannot be defeated' », BBC, 27 nov. 2019.

GOLDMAN B. « Frontières du droit et lex mercatoria », Archives de Philosophie du Droit, t.IX, 1964, p.177-192.

GOOD I. J., « Speculations Concerning the First Ultraintelligent Machine », Advances in Computers, vol. 6, 1965, p. 33.

« Google makes deepfakes to fight deepfakes », BBC, 26 sept. 2019.

GREENBERG B. A., « Rethinking Technology Neutrality », Minnesota Law Review, vol. 100, 2016, pp. 1495, 1511-1512.

HALLEVY G., « The Criminal Liability of Artificial Intelligence Entities - From Science Fiction to Legal Social Control », Akron Intellectual Property Journal, vol. 4, 2010.

HANDELMAN D. A. et alii., « Shared Control of Bimanual Robotic Limbs With a Brain-Machine Interface for Self-Feeding », Frontiers in Neurorobot, vol. 16, 28 juin 2022.

HANSEN P. C., RUBIN M., « China Has No Legal Claim to Taiwan », The National Interest, 8 oct. 2022.

HAO K., « The messy, secretive reality behind OpenAI's bid to save the world », MIT Tech. Rev., 17 févr. 2020.

HARLOW H. F., « The nature of love », American Psychologist, 13(12), 1958, pp. 673-685.

HASSABIS D. et alii., « Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model », arXiv:1911.08265, nov. 2019.

HEAVEN D., « No limit: AI poker bot is first to beat professionals at multiplayer game », Nature, 11 juill. 2019.

« Intelligence artificielle : un comité intergouvernemental d'experts chargé d'étudier l'élaboration d'un cadre juridique », Lexis Actu, 13 sept, 2019.

HINTON G. et alii., « Managing AI Risks in an Era of Rapid Progress », 26 oct. 2023.

HO-DAC M., « La normalisation, clé de voûte de la réglementation européenne de l'intelligence artificielle (AI Act) », D., 2023, p. 228.

HUANG K., « Chinese engineer 'marries' robot after failing to find a human wife », South China Morning Post, 3 avr. 2017.

JAESA « Microsoft a battu Amazon pour le contrat de \$10 milliards du Pentagone », Transhumanisme et IA, 29 oct. 2019.

JALINIÈRE H.,

- « Emmanuel Macron : 1,5 milliard d'euros seront alloués à l'intelligence artificielle », Sciences Et Avenir, 29 mars 2018.
- « Une intelligence artificielle autorisée à poser des diagnostics », Sciences et Avenir, 16 avr. 2018.

JAMES S., « Artificialists Should Be Allowed to Worship », NYT, 29 juill. 2019.

JONES B., « Machines Just Beat Humans on a Stanford Reading Comprehension Test », Futurism, 15 janv. 2018.

JOSSET C., « Le robot Sophia, citoyenne saoudienne, a plus de droits que les femmes de son pays », l'express, 28 oct. 2017.

KAPUR A. et alii, « AlterEgo: A Personalized Wearable Silent Speech Interface », 23rd International Conference on Intelligent User Interfaces, 5 mars 2018, pp. 43-53.

KAUFFMANN S., « Géopolitique de l'Intelligence artificielle », Le Monde, 13 nov. 2019.

KAZI P., « Humans and robots can have babies, claims AI expert », International Business Times, 20 déc. 2017.

KEEGAN R., « A.I. Robot Cast in Lead Role of \$70M Sci-Fi Film », The Hollywood Reporter, 24 juin 2020.

KEYSERS CH., et alii, « The anthropomorphic brain: the mirror neuron system responds to human and robotic actions », Neuroimage, vol. 4, Issue 4, 1 mai 2007, pp. 1674-1684.

KLEIN A., « Rats can make friends with robot rats and will rescue them when stuck », NewScientist, 27 nov. 2018.

LALLOUCHE S. « L'atténuation de la rigueur du système de la responsabilité pour faute dans le domaine médical », les annales de l'université d'Alger 1, n°31, Tome 1, juin 2017, p. 34.

LANDGRAF M., Karlsruhe Institute of Technology, « AI outperforms humans in speech recognition », Techxplore, 20 oct. 2020.

LAROUSSERIE D., « Intelligence artificielle : dorénavant, les machines maîtrisent la grammaire française », Le Monde, 18 nov. 2019.

LAWRENCE D. R., BRAZIER M., « Legally Human? 'Novel Beings' and English Law », *Medical Law Review*, vol. 26, Issue 2, 1 mai 2018, p. 326, 313, 327, 324.

LEFEBVRE A., « Le premier bordel de poupées sexuelles ouvre ses portes en Allemagne », *Express Business*, 19 oct. 2017.

LEFLAR R., « Choice-Influencing Consideration in Conflicts Law », *41 N.Y.U. L. Rev.* 267, 1966.

LEGG S., HUTTER M., « Universal Intelligence: A Definition of Machine Intelligence », *arXiv.org*, déc. 2007, p. 12.

LEQUETTE Y., « Vers un code civil européen ? », *POUVOIRS* 2003/4 (n°107), p. 114.

« Les Gafa concurrencent déjà les États, dénonce Vladimir Poutine », *Capital*, 27 janv. 2021.

LESSIG L., « The Law Of The Horse: What Cyberlaw Might Teach », *Harv. Law Rev.*, vol. 113 issue 501, 1999.

LI G., « Facebook's AI mathematician can solve university calculus problems », *NewScientist*, 20 déc. 2019.

LLOYD D., « Machine learning has been used to automatically translate long-lost languages », *MIT Tech. Rev.*, 1 juill. 2019.

LOISEAU G., « Des droits humains pour personnes non humaines », *D.*, 2011 p. 2558 s.

LOUKIL A., « Un robot obtient haut la main le diplôme de médecine en Chine », *Capital*, 23 nov. 2017.

LU D., « AI camera can tell what surfaces feel like with just a glance », *NewScientist*, 22 sept 2020.

LUFKIN B., « When a pet dies, should you be allowed time off work? », *BBC*, 13 sept. 2018.

OHM P., « The Argument Against Technology-Neutral Surveillance Laws », *Texas Law Review*, vol. 88, 2010, p. 1687.

MARTIN-CHENUT K., « Devoir de vigilance : internormativités et durcissement de la RSE », *D.* 2017, p. 798 s.

MAZIAU N., « L'extraterritorialité du droit entre souveraineté et mondialisation des droits », *La Semaine Juridique Entreprise et Affaires* n° 28, 9 juill. 2015, p. 26.

MCARTHUR N., TWIST M. L. C., « The rise of digisexuality: therapeutic challenges and possibilities », *Journal Sexual and Relationship Therapy*, vol. 32, 2017.

METZ C.,

- « Amazon's Homegrown Chips Threaten Silicon Valley Giant Intel », *NYT*, 10 déc. 2018.

- « Internet Companies Prepare to Fight the 'Deepfake' Future », NYT, 24 nov. 2019.
- « Making New Drugs With a Dose of Artificial Intelligence », NYT, 6 févr. 2019.

« Microsoft muselle son robot 'Tay', devenu nazi en 24 heures », Libération, 25 mars 2016.

MILEY J., « Japanese Man Marries Doll of Virtual Singer », Interesting Engineering, 13 nov. 2018.

MITTELSTADT B., « AI Ethics – Too Principled to Fail? », Oxford Internet Institute, 20 mai 2019, p. 11.

MONTAGNON P., BRAUNE E., « Intelligence Artificielle | Comment la Chine et les États-Unis entendent asseoir leur leadership et souveraineté ? », Forbes, 26 juin 2022.

MOON M., « Famous poker-playing AI takes down scientists and engineers », Engadget, 4 nov. 2017.

MOOR J. H., « Are There Decisions Computers Should Never Make? », Nature & System I, janv.1985, p. 227.

NGUYEN H., « Artificial Intelligence Law is Here, Part One », Evolve The Law, 26 juill. 2018.

NILSSON N. J., « Human-Level Artificial Intelligence? Be Serious! », AI Magazine, 26(4), 2005, p. 68.

NITWBERG M. et alii, « L'intelligence artificielle a besoin d'une Magna Carta », HBR France, 19 déc. 2017.

OpenAI, « Learning Dexterous In-Hand Manipulation », arXiv.org, 1 août 2018.

ORTEGA P. A., MAINI V., DeepMind Safety team, « Building safe artificial intelligence: specification, robustness, and assurance », DeepMind Safety Research, 27 sept. 2018.

PAIS-VIEIRA M., « A Brain-to-Brain Interface for Real-Time Sharing of Sensorimotor Information », Scientific Reports, 28 févr. 2013.

PALMERINI et alii, « Robot Companions as Case-Scenario for Assessing the "Subjectivity" of Autonomous Agents. Some Philosophical and Legal Remarks », Proceedings of the 1<sup>st</sup> Workshop on Rights and Duties of Autonomous Agents, 2012, p. 31.

PATAUT É., « Le devoir de vigilance - Aspects de droit international privé », D. 2017, p. 833.

« Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance », Google AI Blog, 4 avr. 2022.

PAULUS M. et alii, « Saving the Robot or the Human? Robots Who Feel Deserve Moral Care », Social Cognition, 2019, vol. 37, n°1, p. 41 s.

PETROFF A., « Google CEO: AI is more profound than electricity or fire », CNN, 24 janv. 2018.

PIALAT L., ROUSSEAU Y., « Intelligence artificielle : une startup américaine va faire plaider le premier 'robot avocat' », franceinfo, 23 janv. 2023.

PIERRE-YVES G., « Comment l'intelligence artificielle de Libratus est devenue la championne incontestée au Poker », ActuIA, 18 déc. 2017.

PITTALWALA I., « Using artificial intelligence to smell the roses », Phys.org., 29 juill. 2020.

REGAD C., « À quand l'animal reconnu comme une 'personne' juridique en France ? », The Conversation, 29 août 2018.

REIDENBERG J. R., « Lex Informatica: The Formulation of Information Policy Rules through Technology », Tex. L. Rev. vol. 76 n°3, 1997, p. 570.

RITTER J., MAYER A., « Regulating data as property: a new construct for moving forward », 16 Duke Law & Tech. Rev., 2018, pp. 220-277.

« Robot Paints Microscopic Brushstrokes For U.S. Artist's Canvases Worth Over \$100,000 », FeelGuide, 2 juin 2018.

ROUVIÈRE F., « Le robot-personne ou Frankenstein revisité », RTD civ., D., 2018, p. 778.

RUSSELL S., « Of Myths and Moonshine », Edge (blog comment), 14 nov. 2014.

SCHMIDHUBER J.,

- « Annotated History of Modern AI and Deep Learning », AI Blog, 12 déc. 2022.
- « 375th birthday of Leibniz, founder of computer science », AI Blog, 17 mai 2021.

SEE A., « La régulation des algorithmes : un nouveau modèle de globalisation ? », RFDA, 2019, p. 830 s.

SERMONDADAZ S., « Un tableau peint par une IA vendu 432.500 dollars... à partir d'un code source emprunté sur le web », 26 oct. 2018, Science et Avenir.

SHAW A., « A funeral for a spy monkey », BBC Earth, 18 sept. 2019.

SIM C., « Will Artificial Intelligence Take Over Arbitration? », Asian International Arbitration Journal, vol. 14 Issue 1, 2018, p. 13, 11.

SIMON H. A., NEWELL A.,

- « Computer science as empirical inquiry: symbols and search », Communications of the ACM, vol. 19, n° 3, 1976, p. 114.
- « Heuristic Problem Solving: The Next Advance In Operations Research », Reprinted from Operations Research, vol. 6, n°1, janv.-févr., 1958, pp. 9-10.

SINGH-KURTZ S., « The Man of Your Dreams For \$300, Replika sells an AI companion who will never die, argue, or cheat — until his algorithm is updated », The CUT, 10 mars 2023.

SLAUGHTER A.-M., « Sovereignty and Power in a Networked World Order », Stanford Journal of International Law, vol. 40, 2004, pp. 327, 285, 325.

SMITH B., « Legal Personality », Yale Law Journal, vol. 37, issue 3, janv. 1928, p. 296, 289, 292.

SNOW J., « The World of A.I. », NYT, 18 oct. 2018.

SOLUM L. B., « Legal Personhood for Artificial Intelligences », N.C. L. Rev., vol. n°4, 1992, p. 1257, 1262-1263.

SPARROW, R., HOWARD, M., « When human beings are like drunk robots: driverless vehicles, ethics, and the future of transport », Transportation Research Part C, 9 mai 2017, p. 4.

STOCCO A. et alii, « BrainNet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration Between Brains », arXiv.org, 23 sept. 2018.

STREITFELD D., « Computer Stories: A.I. Is Beginning to Assist Novelists », NYT, 18 oct. 2018.

STROGATZ S., « One Giant Step for a Chess-Playing Machine », NYT, 26 déc. 2018.

TANGUY V., « Poutine pense dominer le monde en maîtrisant l'intelligence artificielle », Sciences et Avenir, 5 sept. 2017.

TASHEA J., « AI software is more accurate, faster than attorneys when assessing NDAs », ABA Journal, 26 févr. 2018.

TEUBNER G., « Enterprise Corporatism: New Industrial Policy and the "Essence" of the Legal Person », The American Journal of Comparative Law, 1988, p. 136.

« Textless NLP: Generating expressive speech from raw audio », Meta AI, 9 sept. 2021.

« The microchip implants that let you pay with your hand », BBC, 11 avr. 2022.

The Meta Fundamental AI Research Diplomacy Team (FAIR) et alii, « Human-level play in the game of Diplomacy by combining language models with strategic reasoning », Science, vol. 378, issue 6624, 22 nov. 2022.

THIBIERGE C., « Avenir de la responsabilité, responsabilité de l'avenir », D., 2004, p. 577 s.

TURING A. M., « Computing Machinery and Intelligence », Mind, New Series, vol. 59 n°236, Oxford University Press, oct. 1950, p. 443 et 450, 456, 460.

« Un nouveau comité pour veiller à l'éthique du numérique et de l'intelligence artificielle », Maddynews, 3 déc. 2019.

- VAN CAMP J., « My Jibo Is Dying and It's Breaking My Heart », WIRED, 8 mars 2019.
- VINCENT J., « AI researchers debate the ethics of sharing potentially harmful programs », Verge, 21 févr. 2019.
- VAN DER WEES G., « Has Taiwan Always Been Part of China? », The Diplomat, 1 déc. 2020.
- VINGE V., « The coming technological singularity: How to survive in the post-human era », NASA. Lewis Research Center, Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace, 1993.
- VAN WYNSBERGHE A., ROBBINS S., « Critiquing the Reasons for Making Artificial Moral Agents », Science and Engineering Ethics, vol. 25, issue 3, juin 2019, p. 732.
- WEIN L. E., « The Responsibility of Intelligent Artifacts: Toward An Automation Jurisprudence », Harv. J.L. & Tech, vol. 6 Fall Issue, 1992, pp. 105-106, 116, 107, 113.
- WENG Y.-H. et alii, « Toward the Human-Robot Co-Existence Society: On Safety Intelligence for Next Generation Robots », INT. J. SOC. ROBOT, mars 2009, p. 273.
- WESSBECHER L., « Au Japon, une intelligence artificielle est candidate aux élections municipales de Tama », 13 avr. 2018, France24.
- WIENER N., « Some Moral and Technical Consequences of Automation », Science, vol. 131, n° 3410, 1960, pp. 1355-58.
- WEISE K., « Microsoft Pledges \$500 Million for Affordable Housing in Seattle Area », NYT, 16 janv. 2019.
- WIGGERS K.,
- « MIT CSAIL researchers propose automated method for debiasing AI algorithms », VentureBeat, 27 janv. 2019.
  - « Microsoft gets exclusive license for OpenAI's GPT-3 language model », VentureBeat, 22 sept. 2020.
- WILLIAMS A., « Do You Take This Robot ... », NYT, 19 janv. 2019.
- WILLICK M. S.,
- « Artificial Intelligence: Some Legal Approaches and Implications », AI Magazine, vol. n°2, 1983, pp. 10-11, 13.
  - « Constitutional Law And Artificial Intelligence: The Potential Legal Recognition Of Computers as 'Persons' », Proceedings of the 9th international joint conference on Artificial intelligence (IJCIA), vol. 2, janv. 1985, p. 1271, 1273.
- WOLCHOVER N., « Coder-Physicists Are Simulating the Universe to Unlock Its Secrets », Quantamagazine, 12 juin 2018.

WOLFRAM S.,

- « Dropping In on Gottfried Leibniz », 14 mai 2013.
- « Computational Law, Symbolic Discourse and the AI Constitution », 12 oct. 2016.

YUDKOWSKY E. et alii, « Alignment for Advanced Machine Learning Systems », Machine Intelligence Research Institute, 27 juill. 2016, p. 3.

YUDKOWSKY E., « There's No Fire Alarm for Artificial General Intelligence », MIRI, 13 oct. 2017.

## V. Jurisprudence

BGH, Urt. v. 29. Januar 2001 - II ZR 331/00.

Cass.

- Ass. plén., 9 mai 1984 : Derguini n°80-93481 et Lemaire n° 80-93031.
- Cass. civ. 2<sup>e</sup>, 18 déc. 1964.
- Cass. Ass. plén., 9 mai 1984, n° 80-14994.
- Cass. civ. 1<sup>e</sup>, 9 déc. 2015, n° 14-25.910.
- Cass. civ. 1<sup>e</sup>, 16 janv. 1962.
- Cass. crim., 25 sept. 2012, n° 10-82.938.

Cordella et autres c/ Italie (n° 54414/13), CEDH, 24 janvier 2019.

Lambert et autres c. France n° 46043/14, CEDH 5 juin 2015.

Jacobellis v. Ohio, 1964.

## VI. Autres références

A human-robot dance duet | Huang Yi & KUKA

BORREL PH., « Un monde sans humains? », documentaire, Arte France, Cinétévé, 2012.

CHEN Y.-T., « The implications of artificial general intelligence for international law as a legal system », International Conference ICTO 2023 Ethical & Responsible Artificial Intelligence for sustainable societies, Paris 1 Panthéon-Sorbonne, Paris.

Clément Henin, « Expliquer et justifier des systèmes de décisions algorithmiques : théorie et pratique », Inria, Paris, 16 oct. 2023.

CRAWFORD K., « You and AI - the politics of AI », Royal Society, 2018.

DEPARDIEU G., MARTIN TH., « L'intelligence artificielle va-t-elle nous dépasser? », ARTE, 2018.

DIB D., « L'intelligence artificielle et Chari'a », à l'occasion du colloque international « L'intelligence artificielle : un nouveau défi pour le droit? », organisé par l'Université d'Alger 1 Benyoucef Benkhedda, Alger, 27-28 novembre 2018.

« Digital power at the service of humanity », Conference on Competition and Digitisation, European Commission, Copenhagen, 29 nov. 2019.

Discussion avec Cédric Villani, Science et Avenir.

- Laurence Devillers, 2019.
- Marie Ekeland, 2018.

Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“.

FEYNMAN R. P.,

- « The Computing Machines in the Future », Nishina Memorial Lecture at Gakushuin University, Tokyo, 9 août 1985.
- The Meaning of it All, lectures given in 1963.

FRIDMAN L., AI Podcast

- « Colin Angle: iRobot CEO », 19 sept. 2019.
- « GPT-3 vs Human Brain », 1 août 2020.
- « Jim Gates: Supersymmetry, String Theory and Proving Einstein Right », 25 déc. 2019.
- « Peter Singer: Suffering in Humans, Animals, and AI », 8 juill. 2020.
- « Yann Le Cun: Deep Learning, Convolutional Neural Networks, and Self-Supervised Learning », 31 août 2019.
- « Yoshua Bengio : Deep Learning », 20 oct. 2018.

Future of Life Institute (FLI)

- Yuval Noah Harari discussion avec Max Tegmark, 31 déc. 2019.

Harry Waterstone, à l'occasion de VIU Seminar: The Ethics of Artificial Intelligence, Lehrstuhl für Philosophie und politische Theorie, Ludwig-Maximilians-Universität München, 15-19 oct. 2018.

HASSABIS D.,

- « A systems neuroscience approach to building AGI », Singularity Summit, août 2010.
- « Creativity and AI – The Rothschild Foundation Lecture, Royal Academy of Arts », 17 sept. 2018.
- « The history, capabilities and frontiers of AI », The Royal Society, 30 avr. 2018.

House Science Committee Testimony, Subcommittee on Research and Technology and Subcommittee on Energy Hearing - Artificial Intelligence – With Great Power Comes Great Responsibility, 115th Congress of U.S.A., 26 June 2018.

Jean-Marc Sauvé, « La territorialité du droit », la Rencontre inter-réseaux (franco-américain, franco-brésilien et franco-chinois) sur le thème « Internationalisation du droit : pathologie ou métamorphose de l'ordre juridique ? », les 10, 11 et 12 avril 2012 au Collège de France.

Nicolas Maziau, « L'ordre juridique national en prise avec le droit européen et international : questions de souveraineté ? », colloque organisé par la Cour de cassation et le Conseil d'État, 10 avril 2015.

MARMOZ F., « Justice prédictive : de la fiction à la réalité, quelles conséquences pour les professions du droit? », à l'occasion du colloque international « l'intelligence artificielle : un nouveau défi pour le droit? », organisé par l'Université d'Alger 1 Benyoucef Benkhedda, Alger, 27-28 novembre 2018.

The New York Review of Books Foundation and Fritt Ord hosted the conference « Technology and the Human Future », 20-21 oct. 2016.

Stéphane Mortier, au colloque « Intelligence artificielle et Cyber-sécurité : Perception Juridique », l'ESLSCA Business School Paris, le 25 septembre 2019.

Younès Bennani, au colloque « Sciences & sens de l'intelligence artificielle », Maison des Sciences de l'Homme Paris Nord, 19 nov. 2019.

« 2022 HAI Spring Conference on Key Advances in Artificial Intelligence », 12 avril 2022.

## Index

---

### - C -

**Contrat social de capacité d'inter-espèce,**  
226

### - D -

**Devoir de vigilance étatique,** 126, 127, 128,  
130, 131

**Droit inter-espèces,** 192, 229, 230, 232, 233

### - P -

**Pretium doloris d'IA,** 91, 92, 95, 97, 194

**Principe de faveur pour l'autorégulation,**  
152, 159, 160, 166, 172, 173, 174, 185, 190

### - S -

**Substitution de l'organe de contrainte en  
amont,** 85, 209, 211, 228, 229

**Substitution de l'organe de contrainte en  
aval,** 209, 211