

NUMÉRO SPÉCIAL
MARS 2020

REVUE FRANCOPHONE DE LA
PROPRIÉTÉ INTELLECTUELLE

LES LOIS DE LA ROBOTIQUE D'ASIMOV ET LE DROIT

Sous la direction de Franck Macrez



ASSOCIATION FRANCOPHONE DE
LA PROPRIÉTÉ INTELLECTUELLE

www.revue-rfpi.com

ISSN 2490-8347

Direction de la Revue

Yann BASIRE

Maître de conférences à l'Université de Strasbourg, Directeur Général du CEIPI

Rédactrice en chef

Clémence DE MARASSE-ÉNOUF

Comité Éditorial Central

Yann BASIRE

Maître de conférences à l'Université de Strasbourg, Directeur Général du CEIPI

Valérie-Laure BENABOU

Professeur à l'Université d'Aix-Marseille

Nicolas BRONZO

Maître de conférences à l'Université d'Aix-Marseille [en disponibilité]

Sylvain CHATRY

Maître de conférences à l'Université de Perpignan Via Domitia

José Roberto D'AFFONSECA GUSMAO

Professeur à l'Université Catholique de São Paulo, Avocat

Paulin EDOU EDOU

Docteur en droit

Amélie FAVREAU

Maître de conférences à l'Université de Grenoble-Alpes

Karlo FONSECA TINOCO

Docteur en droit, Avocat

Francis GURRY

Directeur général de l'OMPI

Caroline LE GOFFIC

Maître de conférences HDR à l'Université Paris Descartes

Jacques LARRIEU

Professeur émérite

Christian LE STANC

Professeur émérite, Avocat

Pilar MONTERO

Professeur à l'Université d'Alicante

Patrick TAFFOREAU

Professeur à l'Université de Lorraine

Edouard TREPPOZ

Professeur à l'Université Jean Moulin Lyon3, directeur du Centre Paul Roubier

Comité Régional Europe

Alexis BOISSON

Maître de conférences à l'Université Pierre-Mendès-France

Uroš CEMALOVIC

Professeur à la Faculté de droit, d'administration publique et de sécurité, Université John Naisbitt, Belgrade, Serbie

Jean-Pierre GASNIER

Avocat, Professeur associé à l'Université Aix-Marseille

Natalia KAPYRINA

Docteur en droit privé

Dusan POPOVIC

Professeur à l'Université de Belgrade, Serbie

Ciprian Raul ROMÎȚAN

Maître de conférences, Université « Roumaine-Américaine » de Bucarest

Viorel ROS

Professeur, Université « Nicolae Titulescu » de Bucarest

Eric SERGHERAERT

Professeur, Université Lille, membre du CRDP - l'ERADP

Eleni-Tatiani SYNODINOU

Maître de Conférences à l'Université de Chypre

Comité Régional Océanie

Philippe GIRARD-FOLEY

Avocat

Comité Régional Afrique

Yasser OMAR AMINE

Chercheur en droit de la propriété intellectuelle, Avocat au Barreau du Caire

Ampah JOHNSON

Enseignant-chercheur à l'Université de Lomé

François-Xavier KALINDA

Docteur en droit, ancien Doyen de la faculté de droit de l'Université du Rwanda

Daoud Salmouni ZERHOUNI

Avocat

Comité Régional Amérique du Nord

Charles DE HAAS

Avocat

Emmanuel GILLET

Docteur en droit

Comité Régional Amérique Latine

Lola KANDELAFT

Avocate

Cynthia SOLIS

Avocate

Comité Régional Asie

Shujie FENG

Professeur de droit, Université Tsinghua (Pékin)

Makoto NAGATSUKA

Professeur à l'Université Hitotsubashi

Aso TSUKASA

Maître de conférences à l'Université de Kyushu

Textes révisés par :

Yann BASIRE

Nicolas BRONZO

Nathalie KAPYRINA

Caroline LE GOFFIC





OUR TALENTS » YOUR BUSINESS



**Fidal est le plus grand cabinet d'avocats
d'affaires français indépendant.**

fidal.com

Partenaires stratégiques des entreprises, des institutions et des organisations, nous nous attachons à faire du droit un levier de leur performance et de leur croissance, en France et à l'international. Tout autant experts dans leur discipline que transverses dans leur approche, nos talents parlent le même langage que nos clients et comprennent leurs enjeux. Nous encourageons le partage de la connaissance et de l'expérience. C'est notre manière d'offrir à nos clients - quels que soient leur taille, leur activité, leur implantation géographique ou les problèmes qu'ils nous soumettent - des conseils engagés, éclairés et avisés. Des conseils opérationnels qui les protègent et contribuent activement à leur développement stratégique et commercial.

SUIVEZ-NOUS



Les 29 avocats et juristes du bureau de Strasbourg se tiennent aux côtés des entreprises alsaciennes en leur apportant une expertise juridique complète et adaptée.

* Nos talents au service de votre développement

Fidal Strasbourg
9 avenue de l'Europe
Espace Européen de l'Entreprise
67 300 Schiltigheim Cedex
03 90 22 06 30

Revue Francophone de la Propriété Intellectuelle

ISSN 2490-8347

N° spécial - Mars 2020

Les lois de la robotique d'Asimov et le Droit

SOMMAIRE

F. Macrez et L. Dreyfuss - Éditorial.....	3
F. Macrez - Les lois de la robotique d'Asimov, modèle pour le système juridique ?	7
N. Nevejans - Le robot, responsable ? À propos de la Première Loi de la Robotique d'Asimov	15
U. Bellagamba - Le robot, obéissant ? Réflexions juridiques et anthropologiques sur la deuxième loi de la robotique.....	31
J.-M. Deltron - Le robot, entité autonome ? Entre norme et technique, contextes et limites de l'autonomie cybernétique	41
A. Vallat - Le robot, gardien de l'humanité. La Loi zéro dans les cycles des Robots et de Fondation d'Asimov.	59
G. Vercken Conclusion : l'œil du praticien	71

- **Propriété littéraire et artistique**
- **Médias et communication**
- **Informatique, numérique et internet**
- **Concurrence déloyale et parasitisme**
- **Données à caractère personnel**
- **E-Commerce**

Créé en 2005, l'activité du cabinet Vercken & Gaullier est uniquement centrée sur deux axes spécialisés : la **propriété littéraire et artistique** (droit d'auteur, droits voisins, droit du producteur de bases de données) d'une part, et le **droit du numérique**, d'autre part (contrats et contentieux informatiques, responsabilité des intermédiaires, données à caractère personnel, e-commerce, e-réputation, etc.).

Avec la collaboration de **Valérie-Laure Benabou**, professeure agrégée de droit et de **Benjamin Jean**, consultant technique « Open Source / Open Data », une équipe structurée de **cinq avocates** est au service des clients du cabinet : **grands groupes internationaux, PME, ministères, collectivités publiques, institutions.**

Fort de l'expérience de ses **deux associés, Florence Gaullier et Gilles Vercken**, le cabinet s'est construit une réputation dans tous les **secteurs traditionnels des médias** (édition, musique, télévision et radio, art, etc.), mais également dans les **nouveaux secteurs des médias** (jeux vidéo, multimédia, sites internet, plateformes, market place, réseaux sociaux, logiciels, etc.). L'équipe est **passionnée** par les technologies de l'information et comprend les aspects tant techniques que commerciaux de l'environnement numérique.

© Tomas Heuer

Cabinet Principal

14 rue Séguier – 75006 Paris – France
Tel : +33 (0)1 42 21 34 75 - Fax : +33 (0)1 42 21 32 77

Bureau secondaire

C/O Fidal - 9 avenue de l'Europe – 67300 Schiltigheim
Tel : +33 (0)3 90 22 06 30 – Fax : +33 (0)3 90 22 06 31

ÉDITORIAL

Franck MACREZ

*Maitre de conférences
Centre d'Études internationales de la propriété intellectuelle (CEIPI)
Université de Strasbourg*

Laurence DREYFUSS

*Avocat au barreau de Strasbourg
Cabinet Fidal*

Les fameuses lois de la robotique édictée par Isaac Asimov dans son cycle des robots constituent bien souvent un modèle de référence, conscient ou non, pour qui veut penser la régulation de l'intelligence artificielle. La référence à cet univers qui appartient à l'imaginaire collectif est néanmoins bien souvent purement intuitive, même lorsque c'est le Parlement européen qui s'en empare¹, et aucune étude n'a été menée pour établir la validité et l'intérêt de ces trois Lois pour le système juridique et pour penser son évolution. C'est à ce questionnement que le présent numéro spécial de la RFPI ambitionne de répondre, avec quatre questions correspondant à chacune des Lois : Le robot, gardien de l'humanité ? Le robot, responsable ? Le robot, obéissant ? Le robot, entité autonome ?

Il ne s'agit plus de demain... En effet, l'avocat spécialisé en droit de la propriété intellectuelle est d'ores et déjà fréquemment confronté à la question de la création par un robot.

Quelles protections des créations, par exemple d'info-graphisme, pour lesquelles la « machine » agit beaucoup et l'humain très peu ?

Lorsqu'il s'agit d'un logo, une marque peut être déposée. Le droit revient au titulaire. Le créateur est absent du corpus juridique.

Il en est de même du dessin et modèle. Seuls les critères objectifs de nouveauté et de caractère propre sont pris en compte.

Concernant les brevets, la question demeure ouverte, l'Office Européen des Brevets ayant récemment rejeté deux demandes désignant une « machine » comme inventeur. Si une personne physique avait été désignée, la situation aurait certainement été différente, même si l'intelligence artificielle était intervenue dans le processus inventif.

Se pose la délicate question du droit d'auteur : l'humain est-il l'auteur ? la « machine » est-elle l'auteur ? sont-ils co-auteurs ? l'œuvre est-elle protégeable ? Quant à cette dernière question, la protection requiert que l'œuvre soit marquée de l'empreinte de la personnalité de l'auteur... mais quel auteur ?

Tous les intervenants ont apporté des pistes passionnantes de réflexion en réponse à cette question. Un caractère qui fait la distinctivité de ce numéro tient sans doute à l'éclectisme de ses auteurs : tous gravitent autour de la sphère juridique, mais tout en étant autre chose que juriste... Historien, écrivain de science-fiction, essayiste, sociologue, philosophe, éthicien, spécialiste en intelligence artificielle, voire même astro-physicien... C'est à cette diversité que nous convient les lois d'Asimov, sans perdre de vue que c'est de leur apport au Droit qu'il s'agit, et chacun s'est prêté au jeu, pour ne pas dire

¹ Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL)).

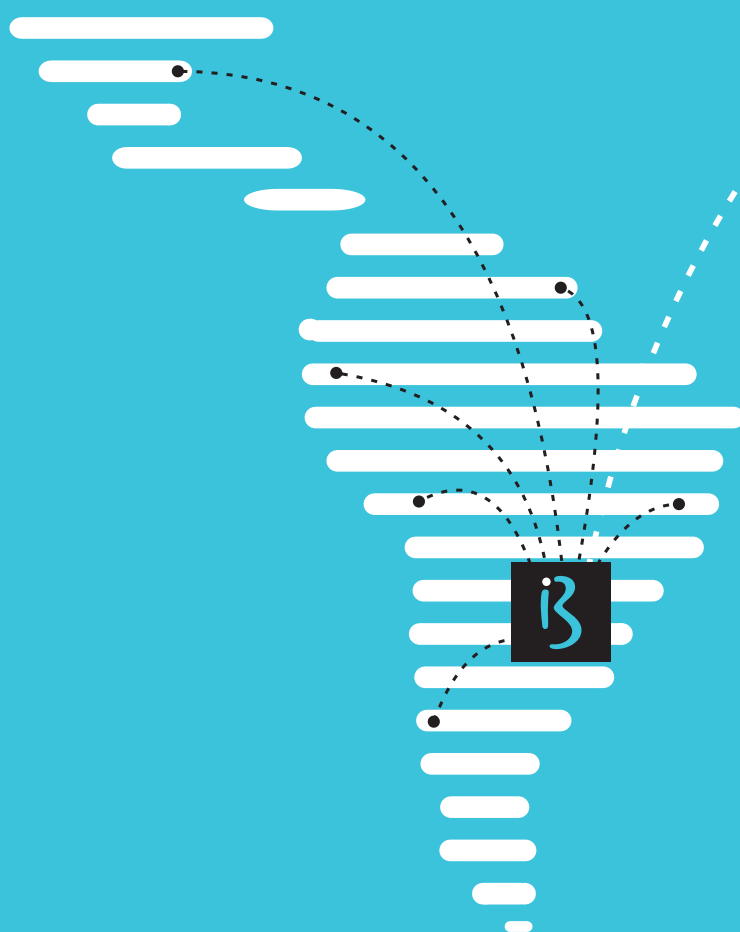
l'exercice, le résultat de notre après-midi de conférence ayant procédé d'une remarquable alchimie, étonnante pour les organisateurs eux-mêmes.

« Chacun » : pas tout-à-fait. Parmi ceux qui avaient accepté de venir contribuer à notre conférence figurait Gilbert Hottois, professeur de philosophie à l'université libre de Bruxelles. Nous lui avons adressé une invitation sans vraiment y croire, un peu comme un enfant s'adressant à une grande personne ou, plus proche de la réalité, un ignorant osant solliciter le Sage. Il avait, très gentiment et très simplement, accepté de venir parler de la loi zéro, manifestement intéressé et enthousiaste à la perspective de cette présentation.

Quelques semaines plus tard, il nous avait prévenu que, hospitalisé, il ne pourrait venir à Strasbourg. Nous avons appris sa disparition quelques jours plus tard, le 16 mars 2019. Nous croyons pouvoir dire sans risque de se tromper, bien que sans aucune légitimité pour le faire, que nous avons perdu un grand penseur de la technoscience et de son éthique. À l'heure où l'intelligence artificielle déchaîne les passions et exacerbe les manichéismes, Hottois savait se situer « entre technophobie et technophilie ». Son intérêt pour la science-fiction (il avait écrit deux romans de ce genre, ou encore dirigé une étude portant sur « philosophie et science-fiction ») allié à cette pondération face au phénomène technoscientifique, en faisait le candidat idéal pour nous parler de la loi Zéro et des rapports entre l'humanité et les robots. Anthony Vallat a accepté de relever le défi d'un tel sujet, nous transmettant sa passion pour l'œuvre d'Asimov.

Du philosophe à l'avocat en passant par l'écrivain de science-fiction, l'interdisciplinarité des champs de réponses permet, ce numéro nous semble en témoigner, d'appréhender globalement les rapports entre le Droit et l'intelligence artificielle, pour le présent et pour l'avenir.

Experts en propriété intellectuelle en Amérique latine



Dépôt de demandes
Convention de Paris / PCT

**Recherches
d'antériorités**

**Traduction de
brevets**

**Commercialisation
de la PI**

i3
BERKEN IP

Les lois de la robotique d'Asimov, modèle pour le système juridique ?

Franck MACREZ

Maitre de conférences
Centre d'Études internationales de la propriété intellectuelle (CEIPI)
Université de Strasbourg

Introduction

1.- De *Fondation* aux *Robots*. Le génie d'Asimov ne se mesure pas uniquement aux succès de librairie, aux prix littéraires, ou encore à son influence chez d'autres auteurs.

Bien entendu, on sait qu'il est l'auteur de *Fondation*, élue meilleure série de science-fiction de tous les temps (prix Hugo 1966), qui plonge le lecteur dans un empire galactique à son apogée, inexorablement amené à décliner : grâce à la psychohistoire, science statistique permettant de prédire l'avenir, Hari Seldon a établi que l'Empire allait s'effondrer et laisser place à 30 000 ans de barbarie... Ce cycle s'étend sur cinq siècles, aux moments cruciaux de l'évolution de la Fondation, tandis qu'advient la décadence de l'Empire, à la manière de celle de l'Empire romain, dont Asimov s'est inspiré.

Mais son cycle *Les Robots* n'en est pas moins une œuvre de science-fiction exceptionnelle,

qui a renouvelé le genre, en particulier en évitant soigneusement le thème du robot se retournant contre son créateur, grâce au postulat des lois de la robotique, qui gouvernent les actions de tout robot au sein de cet univers.

Les autres volumes du cycle *Fondation*, écrits bien plus tard par Asimov, feront se rejoindre les deux univers, leur donnant une

remarquable cohérence. Au final, l'extraordinaire supériorité intellectuelle du Robot lui permet de considérer qu'il doit et peut mener l'humanité vers l'Utopie d'une société parfaite : celle du Second Empire, qu'il contribue à créer dans l'ombre.

Ainsi, le juriste amateur de science-fiction peut y trouver une source d'inspiration à plusieurs égards. D'abord, la perfection rationnelle du robot lui permet de tendre vers l'Utopie¹, laquelle contient évidemment l'idéal de Justice. Ensuite, l'invention du terme de « robotique » permet de penser l'ensemble des techniques permettant la conception et la réalisation de machines « pensantes » comme un tout. Enfin, la formulation des « lois » invite à penser la régulation de cet ensemble de machines. Il semble bien qu'au final, la science-fiction soit un puissant outil pour penser la société contemporaine et son évolution².

2.- Des « robots » aux lois de la robotique.

L'étymologie du mot « robot » montre que c'est avant tout une machine destinée à supplanter l'homme dans des tâches pénibles : le salve *Rabota* signifie corvée, le russe *rab*, esclave. Le robot relève d'abord du dispositif automate répétitif : machine à coudre, moissonneuse-batteuse, presse-purée, aspirateur, robot de la chaîne de montage industrielle... Mais aujourd'hui, le « robot » est bien plus : il serait intelligent, grâce aux traitements algorithmiques des données ;

¹ Pour une analyse dans l'œuvre d'Asimov, v. : A. Vallat, Utopie et raison dans le cycle de Fondation d'Isaac Asimov, ActuSF, 2014, *passim* ; A. Vallat, « Le robot, gardien de l'humanité ? », in *Les lois de la*

robotique d'Asimov et le Droit, conférence CEIPI, Strasbourg 26 avril 2019, <www.canalc2.fr>.

² V. « Les lois de la robotique d'Asimov et le Droit », conférence du 26 avril 2019 à l'Université de Strasbourg, disponible sur www.canalc2.tv.

parfois même, on ne désigne par ce terme rien de plus qu'un pur algorithme (les « robots » de Google qui scannent le web pour en référencer les pages). Le mythe de l'« intelligence artificielle », expression utilisée pour désigner essentiellement la technologie des algorithmes d'apprentissage, rapproche les robots contemporains des robots d'Asimov aux cerveaux positroniques. Nous sommes pourtant loin, techniquement, du robot humanoïde qu'on ne sait distinguer d'un être humain...

Une des fiertés d'Asimov est d'avoir inventé, sans le savoir, le terme de « robotique »³, ce qui lui assure en soi une certaine postérité. Les désormais fameuses lois de la robotique sont rédigées ainsi :

« 1. Un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ;

2. 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 ;

3.- Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi. »

Une quatrième loi, la loi Zéro est induite des trois autres :

« 0. Un robot ne peut pas faire de mal à l'humanité, ni, par son inaction, permettre que l'humanité soit blessée »

Si, intuitivement et en première intention, on peut se dire que le juriste amateur de l'œuvre d'Asimov peut éventuellement s'inspirer de l'œuvre de fiction pour penser son activité de juriste, il nous semble que la question peut être posée très sérieusement : ces lois de la robotique peuvent être constituer un modèle pour le Droit ?

3.- Un modèle pour le Droit ? Traiter une telle question implique avant tout de déterminer ce qu'il convient d'entendre par « modèle » dans

un contexte de science juridique. Dans un sens général, le modèle est une représentation du fonctionnement de quelque chose. Plus particulièrement, il convient d'y voir une « représentation simplifiée d'un objet ou d'un processus qui existe ou doit exister »⁴, et de distinguer entre modèle sémantique, modèle théorique, et modèle normatif. Il est possible, en première intention, de se demander s'il faudrait identifier les trois Lois comme constitutives d'un modèle théorique, c'est-à-dire comme « idéalisation » de la réalité. Mais l'œuvre de fiction fonctionnerait plutôt par analogie avec la réalité alors que le modèle théorique est descriptif. À la vérité, les trois Lois pourraient relever plutôt du modèle normatif, qui « indique ce qui doit être »⁵, que ce soit en tant que modèle politique, éthique, voire de science juridique. À l'extrémité, ce questionnement conduira à s'interroger sur le fait de savoir si les Lois d'Asimov ne pourraient pas être vues comme sources de Droit.

En tout état de cause, il conviendra, à considérer que l'œuvre de fiction puisse servir de modèle, d'identifier quel rôle elle peut jouer, de savoir à quel type de modèle on se réfère. Bien entendu, il s'agit en l'occurrence de s'interroger sur la régulation de l'« intelligence artificielle », aujourd'hui en devenir, c'est-à-dire de se demander en quoi les lois de la robotique peuvent constituer un modèle pour penser les rapports entre intelligence artificielle et Droit.

Il nous semble que ce modèle peut être triplement caractérisé : les trois Lois apparaissent avant tout comme un modèle crédible (I). Même si ce modèle est imparfait (II), il apparaît utile pour penser l'appréhension du numérique par le système juridique (III).

³ Apparu pour la première fois dans la nouvelle *Menteur*, en 1941.

⁴ V° « modèle » in A.-J. Aranaud (dir.), Dictionnaire encyclopédique de théorie et de sociologie du droit, LGDJ, Paris, 1993.

⁵ Ibid.

I.- Les lois de la robotique, modèle crédible

4.- L'ontologie des Lois, modèle idéal. Sans entrer dans le détail du contenu des lois, qui conduisent à considérer que le robot doit être obéissant, et à s'interroger sur le point de savoir s'il peut être autonome ou responsable⁶, il est inévitable de considérer que les Lois constituent un modèle crédible en raison de leur existence même et de leurs caractères. Elles sont en effet conçues en relation les unes avec les autres (A), et également impératives et immuables (B).

A. Des lois coordonnées

5.- La hiérarchie des Lois d'Asimov. Le premier caractère, sur lequel nous passerons rapidement, tient au fait que les Lois sont coordonnées, d'une manière qui fait songer à la hiérarchie des normes kelsénienne. Chacune des lois s'applique en effet sous réserve de ne pas entrer en conflit avec la précédente, traduisant ainsi une hiérarchie de valeurs et d'objectif : d'abord, protéger l'humanité toute entière, ensuite les êtres humains individuellement, puis assurer l'obéissance du robot et enfin protéger l'existence du robot lui-même. Asimov a d'ailleurs exploré les potentialités de cette forme de hiérarchie dans certaines nouvelles, notamment en trouvant des situations indécidables⁷ ou qui conduisaient à des comportements absurdes⁸. Cette nécessaire structuration rappelle le système juridique tel qui est réellement conçu, ce qui participe de la crédibilité des Lois en tant que modèle réel.

6.- Importance de la structuration du système juridique. Dès lors que l'on veut bien

concevoir le Droit comme un système, son devoir-être est la cohérence, et cette cohérence implique de prévoir la manière dont les droits s'articulent entre eux. Un des moyens logiques est l'opération de rangement, qui permet de résoudre une antinomie. Autrement dit, la nécessité de cohérence conduit à la découverte d'un principe hiérarchique destiné à mettre de l'ordre au sein du système juridique⁹ : la négation d'une des règles « est la solution naturelle, fatale et irréversible de toute antinomie ; dès qu'il y a contradiction (conceptuelle, politique, humaine, etc.), l'un des contraires tend à l'emporter sur l'autre »¹⁰. Asimov, qui pourtant faisait preuve d'une érudition assez exceptionnelle, n'avait sans doute pas lu de littérature juridique... Pourtant, cette structuration des Lois, identique dans sa conception au système juridique tel qu'il est élaboré réellement, invite à l'analogie et la comparaison.

B. Des lois immuables

7.- « Code is law ». Un caractère plus original des trois Lois est que, dans l'univers d'Asimov, les Lois sont absolument impératives et immuables. À aucun moment il n'est question d'imaginer des robots qui ne seraient pas régis par ces Lois car cela est une condition de leur acceptation par l'humanité, méfiante envers ces machines intelligentes, c'est-à-dire que c'est une condition de leur existence même. Cela signifie que la règle et son effectivité même se situent à l'intérieur de la machine, au plus profond des cerveaux positroniques imaginés par l'écrivain. C'est dire que les Lois, écrites en langage naturel pour être expliquées aux humains, sont avant tout formalisées et encapsulées au sein d'un dispositif technique (qu'Asimov ne décrit pas). En cela les Lois constituent un exemple de la technique prise comme norme, comme devoir-être¹¹. Elle est un

⁶ V. « Les lois de la robotique d'Asimov et le Droit », conférence pécitée.

⁷ Cercle vicieux, 1942.

⁸ *Menteur*, 1941.

⁹ M. Villey, *Philosophie du droit*, Dalloz, Paris, 2001 (1986), p.280, et p.323, n°244.

¹⁰ P. Malaurie, « Les antinomies des règles et de leurs fondements », *Le droit privé français à la fin du XXe siècle*, Études offertes à Pierre Catala, Dalloz 2001, p.25, pp.27-28

¹¹ Même si, de manière générale, la technique se situe vraisemblablement à la frontière de l'être et du devoir-être : P. Amselek, « Norme et loi », *Archives de*

mode de régulation de la conduite des hommes, parmi trois autres modes que sont la loi, la norme sociale et le marché¹² : le code (informatique), c'est la loi¹³.

L'idée d'implémenter une norme (juridique) au sein de dispositifs techniques n'est pas nouvelle, et a été âprement discutée à propos des mesures techniques de protection des œuvres et des dispositions juridiques interdisant de les contourner, dans le domaine du droit d'auteur¹⁴. Car les dispositions anti-contournement sont apparues nécessaires eu égard au fait qu'il n'est pas de dispositif informatique qui ne puisse être hacké. C'est, au-delà du contenu des Lois, ce caractère d'immutabilité (présumée), cette implémentation par défaut qui a séduit certains législateurs qui se sont expressément référés aux Lois d'Asimov.

8.- Des Lois consacrées. De fait, et de manière tout à fait remarquable, le Parlement européen a fait explicitement référence à la série de science-fiction dans sa résolution de février 2017, en ces termes :

« considérant 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'autoapprentissage, étant donné que lesdites lois ne peuvent être traduites en langage de programmation »¹⁵.

Il semble donc que c'est avant tout le positionnement « en amont » de l'implémentation des règles qui conduit le

législateur européen à s'inspirer des Lois d'Asimov et à les prendre comme modèle.

Le gouvernement coréen avait adopté une posture analogue dès 2007, avec un projet de charte éthique des robots, qui prévoit de réglementer avant toute chose les normes de fabrication¹⁶. La démarche en Grande-Bretagne est comparable, dans un guide officiel de l'agence de standardisation et certification¹⁷. Et le Conseil d'État va dans le même sens en préconisant la mise en place de « mécanismes directement implantés dans le programme initial »¹⁸, nommés par lui-même « bouton Asimov » et « porte Asimov ». Certaines règles, fondamentales, nécessitent d'être techniquement implémentées par défaut et en se situant en amont au stade de la conception, et le « modèle Asimov » apparaît bien présent dans les esprits quand il s'agit de concevoir la norme juridique de demain pour réglementer l'intelligence artificielle.

Mais si, en leur principe, les trois Lois servent de modèle en quelque sorte ontologique, voire même symbolique, elles sont limitées au stade de la mise en œuvre.

II.- Les lois de la robotique, modèle imparfait

9.- Incomplétude du modèle. L'imperfection du modèle peut se constater à deux niveaux : d'une part, le contenu des Lois ne correspond pas, aujourd'hui, à la volonté des législateurs

philosophie du droit 1980, XXV, p.89, spéc. p.100 ; F. Violet, *Articulation de la norme technique et de la règle de droit*, PUAM, 2003, p.33

¹² L. Lessig, « The Law of the Horse : What Cyberlaw Might Teach », *Harvard Law Review* 1999, 113, p.501, <www.lessig.org>, spéc. p.507.

¹³ L. Lessig, *Code and other laws of cyberspace*, Basic Books, 1999.

¹⁴ Article 11 du Traité de l'OMPI sur le droit d'auteur, 1996 ; Directive n° 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information, JOCE, 22 juin 2001, considérants 47 à 50 et article 6.

¹⁵ Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL)).

¹⁶ Corée du Sud, Charte éthique des robots, 2007, *Le Monde*, 7 mars 2007.

¹⁷ The British Standards Institution, *Robots and robotic devices. Guide to the ethical design and application of robots and robotic system*, BS 8611:2016, 30 avr. 2016.

¹⁸ Conseil d'État, *Étude annuelle 2017, Puissance publique et plateformes numériques : accompagner l'« ubérisation »*, p.117.

(A) ; d'autre part, le caractère immuable est problématique au stade de la mise en œuvre des règles (B).

A. Vers un autre contenu des « trois Lois » ?

10.- Conseil d'État : un « bouton Asimov » et une « porte Asimov ». Si la référence à l'écrivain américain de science-fiction est remarquable, pour ne pas dire étonnante, dans le cadre qui est celui d'un rapport du Conseil d'État, force est de constater que le modèle ne va pas jusqu'au contenu des règles à implémenter. En effet, alors qu'il s'agirait de consacrer le principe de loyauté et de responsabilité dans l'utilisation des algorithmes, la haute juridiction suggère de mettre en place ce qu'elle nomme elle-même un « bouton Asimov » ou une « porte Asimov »¹⁹. Il s'agirait de pouvoir geler l'exécution d'une tâche (le bouton) ou de permettre aux services de sécurité et de renseignement de pénétrer le système (la porte) lorsque certaines conditions sont réunies, en particulier la méconnaissance de droits fondamentaux ou l'atteinte à des êtres humains. La référence aux trois Lois s'arrête donc à l'idée d'immuabilité, mais le contenu de la règle est différent.

11.- Des trois Lois adaptées aux vingt-trois règles d'Asilomar. Les lois d'Asimov sont donc source d'inspiration, mais leur contenu apparaît inadapté aux problématiques contemporaines. Révélateur de cet état de fait, un spécialiste en intelligence artificielle a pertinemment suggéré, tout en conservant ces grands principes au nombre de trois, de réglementer l'intelligence artificielle en se fondant sur ces trois règles²⁰ :

1.- Le propriétaire est responsable des actes de l'intelligence artificielle ;

2.- Une intelligence artificielle / un robot doit s'annoncer comme tel ;

3.- Une intelligence artificielle ne peut partager des informations qu'avec l'accord de son propriétaire

Plus en phase avec les problématiques contemporaines de responsabilité et de respect des données personnelles, ces trois principes se déclinent rapidement en une multiplicité lorsqu'il d'agit de penser un développement éthique de l'intelligence artificielle. Ainsi, les vingt-trois principes d'Asilomar déclinent un guide éthique et pratique pour les concepteurs d'intelligence artificielle²¹. De telles initiatives, qui rappellent à l'évidence les lois d'Asimov, demeurent symboliques, sans réelle portée juridique.

Si l'obligatorité leur fait défaut, faute d'être édictée par une autorité compétente, la question de leur transcription en un texte juridique effectif demeure problématique.

B. Des Lois impossibles à implémenter ?

12.- Champ d'application. La première difficulté tient au champ d'application de telles Lois de la robotique. Dans *Les Robots*, les Lois s'appliquent à des robots intelligents, qui semblent avoir conscience d'eux-mêmes. Nous sommes loin d'avoir développé une « intelligence artificielle forte » et, à la vérité, les algorithmes autoapprenants tels qu'ils existent à ce jour sont assez loin de l'idée qu'on doit se faire de l'intelligence... En réalité, ces algorithmes sont particulièrement stupides : « L'intelligence artificielle a moins de sens commun qu'un rat », reconnaît Yann Le Cun, responsable de la recherche sur l'intelligence artificielle de Facebook²². Or les trois Lois concerneraient, au mieux, une intelligence artificielle forte, qui aurait conscience d'elle-même. Une autre difficulté, qui relève du même ordre d'idée, tient au fait que les trois

¹⁹ Conseil d'État, Étude annuelle 2017, *eod. loc.*

²⁰ Oren Etzioni, "How to regulate Artificial intelligence", The New York Times, 1er sept. 2017.

²¹ FLI, « Asilomar AI Principles », 2017, <https://futureoflife.org/ai-principles/>.

²² Yann Le Cun, conférence à Station F, Paris, 23 janvier 2018.

Lois s'adressent non aux hommes, mais... aux robots. C'est une des raisons pour laquelle une étude commandée par la Commission européenne et conclu que la référence aux lois d'Asimov « est inopportune pour des raisons scientifiques et culturelles »²³.

13.- Des lois immuables et inaltérables ? Dans *Les Robots*, il n'est pas question de pouvoir supprimer les Lois, tout au plus voit-on les concepteurs de l'*US Robots* pouvoir diminuer l'importance de l'une par rapport aux autres. Cela est fondamental, car la méfiance des humains envers les robots est telle que l'hypothèse d'un robot pouvant ne pas respecter les Lois mettrait en péril l'existence de l'ensemble des robots, qui seraient rejetés par l'humanité. Ainsi, l'univers d'Asimov ne connaît pas le *hacker*... Cette possibilité est exploitée dans la série suédoise « Real Humans : 100 % humain » (2012-2014) : des robots androïdes, les *hubots*, ayant accédé à la conscience essaient de débloquent le « code Asimov » pour accéder à la liberté. Dans l'œuvre d'Asimov, l'écrivain joue souvent avec l'interprétation des lois pour voir s'il est possible de les contourner. Mais leur implémentation demeure infalsifiable. En cela, les Lois constituent un idéal que nous ne saurions atteindre aujourd'hui en l'état de la technique : dans la réalité, tout est code... et tout code peut être renversé par un autre code. C'est la raison pour laquelle, dans la législation sur les mesures techniques de protection, ont été créées des dispositions anticircumvention, c'est-à-dire des obligations juridiques venant renforcer le « verrou » technique en interdisant de le « crocheter » pour passer outre l'interdiction technique²⁴.

Ainsi, le modèle des trois Lois apparaît bien hypothétique, y compris dans leur caractère le plus séduisant, l'immuabilité dans leur mise en œuvre. Ce modèle n'en reste pas moins d'une utilité particulière pour penser le droit et ses potentialités d'évolution.

III.- Les lois de la robotique, modèle utile

14.- L'utilité du modèle. Les Lois d'Asimov sont instructives en ce que le modèle demeure une référence qui dispose d'un important pouvoir d'évocation (A) dont il est nécessaire de prendre la mesure (B).

A. Un modèle évocateur

15.- Pouvoir cognitif du modèle. La puissance du modèle est sans doute là : il constitue un outil pédagogique formidable. Les histoires de fictions en général, et celles de robots en particulier, donnent à voir : que ces robots soient aimés ou craints, il y a un élément affectif indéniable qui permet de passer un message aisément. Ainsi, la question de la régulation de l'intelligence artificielle est plus facilement posée auprès d'un public qui connaît les histoires d'Asimov, de Philip K. Dick, A. Huxley ou G. Orwell. En réalité, ces œuvres nous semblent appartenir à l'inconscient collectif, si bien qu'on peut par exemple constater que la référence aux Lois d'Asimov est récurrente quasiment dans tout débat sur la question de la construction du droit de l'intelligence artificielle. Comme toute métaphore, la référence à l'univers de science-fiction permet de transmettre, une idée, un message ou encore les termes d'une problématique, avec une efficacité remarquable.

16.- Un outil d'anticipation du réel. La science-fiction, ou plus précisément le genre de l'anticipation permettent précisément de se figurer un futur plus ou moins lointain. Sans être des œuvres divinatoires, l'univers imaginé par l'écrivain invite à imaginer le devenir de la société, ce qui à l'évidence constitue une fonction importante pour celui qui veut penser le Droit, tourné vers l'avenir. Ainsi, la lecture

²³ N. Nevejans, Règles européennes de droit civil en robotique, étude réalisée pour la Commission Juri, PE 571.379, octobre 2016.

²⁴ V. *supra* n°7.

de *Conflit évitable* d'Asimov²⁵, sans constituer une description exacte d'un réel, permet de réaliser le caractère inévitable de la prise d'importance des algorithmes dans le fonctionnement de notre société. Dans un autre registre, les univers de 1984 de Orwell et du *Meilleur des Mondes* de Huxley invitent à prendre conscience avec une acuité particulière de problématiques essentielles à notre société. Peut-être le rôle du juriste, et du législateur, est d'éviter que la réalité ne rejoigne la dystopie... Ainsi que de s'interroger sur le degré de réalisation de cette dystopie ! Le rôle de ces œuvres d'anticipation est néanmoins difficile à évaluer : Pierre Boucher²⁶ ou les parlementaires ayant voté la loi informatique et libertés de 1978 avaient-ils lu Orwell ?...

En tout état de cause, l'œuvre de fiction dispose d'un pouvoir cognitif important particulièrement précieux dans le cadre d'une réflexion *de lege ferenda*. Mais leur utilité n'en est pas moins circonscrite.

B. Une utilité circonscrite

17.- Un modèle pour des règles éthiques, mais non juridiques. Pour en revenir aux Lois d'Asimov, nous avons constaté l'impossibilité concrète de les mettre en œuvre dans un texte légal ou un code informatique²⁷. Cela a pour conséquence que la juridicité des Lois fait défaut : si elles constituent un modèle, ce ne peut être un modèle pour un texte juridique. Tout au plus ces Lois pourraient-elles relever d'un code éthique, se situant donc en amont du texte de loi. C'est d'ailleurs le phénomène que l'on peut constater aujourd'hui : les chercheurs en intelligence artificielle expriment un besoin de se doter de règles éthiques²⁸. Cette limite ne doit néanmoins pas conduire à

négliger leur importance : les trois Lois servent de modèle à un niveau fondamental de création de la règle de droit.

18.- Une « source programmatique du droit » ? Finalement, à raisonner en termes de source de droit, le modèle est peut-être utile quelque peu au-delà du simple niveau éthique. C'est ce qui fait dire à un auteur à un auteur que ces trois lois de la robotique seraient une « source programmatique du droit »²⁹, qui en fixe les grandes lignes. Cela implique de comprendre le terme de « source » de la manière la plus large, ce qui se conçoit aisément : les sources de droit sont l'ensemble des « forces d'où surgit le Droit »³⁰. Mais il nous semble qu'elle est une source indirecte : plutôt une source d'inspiration qu'une source de droit. Finalement, les trois Lois pourraient être une source d'inspiration d'une source réelle³¹, pour un texte juridique qui pourrait aller au-delà du code éthique, une norme fondamentale qui viendrait réglementer l'ensemble des comportements des acteurs de l'intelligence artificielle.

Conclusion

19.- Les trois Lois : immuables, mais « inimplémentables ». En fin de compte, la question de savoir sur les Lois de la robotique d'Asimov peuvent constituer un modèle pour le Droit appelle une réponse nuancée. Les trois Lois paraissent bien constituer un modèle normatif, qui « indique ce qui doit être », mais seulement en ce qu'elles sont immuables. Cela étant, elles ne servent ainsi qu'à affirmer un vœu pieu qui revient à souhaiter d'une bonne règle demeure règle pour l'éternité... Mais les trois Lois ne sauraient constituer un modèle normatif quant à leur contenu, car elles ne

²⁵ The Evitable Conflict, 1950.

²⁶ « Safari ou la chasse aux français », *Le Monde* 21 mars 1974.

²⁷ *Supra* n°12.

²⁸ P. ex. les vingt-trois règles d'Asilomar, *supra* n°11.

²⁹ F. Defferrard, « La science-fiction, source littéraire du droit. L'exemple des Trois Lois de la Robotique », in *Les fictions en droit* (sous la dir. de F.-X. Roux-Demare et

M.-Ch. Dizès), coll. « Colloques & Essais », Institut Universitaire Varenne, 2018, p. 9.

³⁰ Association Henri Capitant, Vocabulaire juridique, par G. Cornu (dir.), PUF, 2003, p.846.

³¹ Sur la distinction « sources réelles » / « sources formelles », v. P. Roubier, « L'ordre juridique et la théorie des sources du droit », in *Le droit privé français au milieu du xx^e siècle, Études offertes à G. Ripert*, LGDJ, tome I, 1950, p.9 et s.

correspondent pas aux problématiques contemporaines.

Par ailleurs, la difficulté technique - pour ne pas dire l'impossibilité - d'implémenter de telles règles de manière inaltérable relègue ces Lois au rang de modèle symbolique. En somme, si le juriste voulait y trouver un modèle proposant des solutions, il risque

d'être déçu. Ce sont plutôt les questionnements que les trois Lois induisent qui font la richesse du modèle : le symbole est fort et la puissance cognitive de l'œuvre de fiction particulièrement précieuse pour guider des réflexions quant à la possible réglementation de l'intelligence artificielle.

F. M.

Le robot, responsable ? À propos de la Première Loi de la Robotique d'Asimov

Nathalie NEVEJANS

*Maître de conférences en droit privé, HDR, Faculté de droit de Douai
Directrice de la Chaire IA Responsable, Université d'Artois
Centre Droit, Ethique et Procédures (EA n° 2471)
Membre du Comité d'éthique du CNRS (COMETS)*

Introduction

La notion de robot. Les robots du *Cycle des robots* de l'écrivain américain d'origine russe Isaac Asimov (1920-1992) ne sont pas très éloignés de la définition qui en est généralement retenue par les roboticiens, à savoir une machine matérielle dotée d'une architecture dédiée lui offrant la possibilité de se mouvoir dans le monde réel, et de programmes d'intelligence artificielle lui permettant de décider, parfois en toute autonomie, voire d'apprendre par elle-même et d'interagir avec l'homme¹. Même si le robot des nouvelles d'Asimov est le plus souvent de forme humanoïde, ce modèle correspond également à la vision pragmatique des roboticiens. En effet, la forme du robot étant fonction de son usage, s'il est utilisé aux côtés d'un humain ou dans un environnement humain, comme dans l'œuvre d'Asimov, la forme humanoïde peut être plus avantageuse. On peut toutefois mentionner la nouvelle *Conflit évitable*², dans laquelle l'auteur met en scène une « Machine » dont on ne sait rien, sauf qu'elle dirige le monde. Il ne s'agit pas du robot classique du répertoire asimovien, puisqu'il s'apparente plutôt à une sorte de programme d'intelligence artificielle.

Du Golem au robot. Le Golem de la Bible était une créature en argile sans intelligence et libre arbitre, façonnée par l'homme à l'image de l'homme, comme la tradition talmudique

l'autorisait. Elle pouvait travailler à la condition de placer dans sa bouche un message portant le nom secret de Dieu et de l'y retirer le vendredi soir pour respecter le shabbat. Les récits rapportent une créature fabriquée par le rabbin Laew de Prague au XVI^e siècle, qui aurait échappé à sa vigilance et provoqué de formidables destructions dans la ville. Cette dévotion de la créature au travail n'échappa pas à l'écrivain tchèque Karel Čapek qui inventa pour sa pièce de théâtre *Rossum's Universal Robots* (1920) le terme « robot » à partir du mot tchèque *robota*, qui signifie « corvée » ou « travail forcé ». Dans cette pièce, les robots humanoïdes, intelligents mais sans conscience, finissent par se révolter contre les humains devenus décadents. Le thème de la créature dangereuse a également nourri le roman de Mary Shelley *Frankenstein ou le Prométhée moderne* (*Frankenstein or The Modern Prometheus*) (1818). L'un des récits peuplant cette œuvre porte sur le jeune Victor Frankenstein qui, après avoir fabriqué une créature composée de morceaux d'êtres humains, était parvenu à l'animer au moyen de l'électricité, pour finir par trouver insupportable de contempler son monstre. Un autre récit évoque ensuite le sort de cette créature qui, rejetée par l'humanité, finit par concevoir une haine féroce contre son créateur. Le Golem et cette créature composite sont les ancêtres romanesques des robots. D'ailleurs, Asimov reconnaît ce lien en tentant de mettre

¹ V. N. Nevejans, « Les robots : tentative de définition », in *Les robots*, A. Bensamoun, dir., éd. Mare et Martin,

coll. « Presses Universitaires de Sceaux », 2015, p. 79-117.

² *The Evidable Conflict*, 1950.

fin à ce qu'il appelle le complexe de Frankenstein, dont souffrent les personnages de son œuvre qui considèrent les robots comme des créatures dangereuses, en cherchant un moyen littéraire pour empêcher la révolte de ses machines. La solution apparaît avec ses Lois de la Robotique, qui voient le jour dans sa nouvelle *Runaround*³. Les robots, fabriqués par l'homme pour travailler pour l'homme, ne sont que des machines dotées d'un dispositif de sécurité composé de règles de logique contraignantes implantées dans le cerveau positronique au moment de leur fabrication, et les rendant insusceptibles « de causer le moindre préjudice aux hommes »⁴.

La sécurité des êtres humains face aux robots.

Asimov a une vision très réaliste des robots, puisqu'il estime qu'« en tant que machine, un robot comportera sans doute des dispositifs de sécurité aussi complets que possible. Si les robots sont si perfectionnés qu'ils peuvent imiter le processus de la pensée humaine, c'est que la nature de ce processus aura été conçue par des ingénieurs humains qui y auront incorporé des dispositifs de sécurité. Celle-ci ne sera peut-être pas parfaite. (Mais la perfection est-elle de ce monde ?) Cependant elle sera aussi complète que les hommes pourront la réaliser »⁵. Asimov considère que ses Lois constituent un code moral contraignant encadrant fermement le robot pour lui interdire toute atteinte à la sécurité de l'humain. Il est à rapprocher des principes éthiques en cours d'élaboration en Europe pour rendre le robot et l'intelligence artificielle sûrs et acceptables pour l'humain. S'il convient d'évoquer sur ce point la résolution du 16 février 2017 sur les règles de droit civil sur la robotique⁶, c'est surtout le « Guide éthique pour une IA de confiance » du Groupe d'experts de haut niveau sur l'intelligence

artificielle du 8 avril 2019 qui estime que l'intelligence artificielle est digne de confiance si elle est licite, éthique, et robuste tant d'un point de vue technique que social⁷.

La Première Loi selon laquelle « *un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger* » présente une importance centrale pour remplir les objectifs de l'auteur concernant la sécurité humaine. Présentée comme un devoir du robot envers l'humain dans *L'homme bicentenaire*⁸, elle s'apparente davantage à une obligation contraignante pour la machine qu'à un simple devoir. L'intérêt de cette Première Loi est de rendre un robot acceptable par l'humain, car on sait qu'il ne pourra jamais le blesser. Elle présente donc une fonction préventive contre les atteintes à la sécurité de l'humain (I). Toutefois, comme le dispositif de sécurité est créé par l'homme, il connaît nécessairement des imperfections, ce qui génère de précieuses histoires chez Asimov. Dès lors, il convient d'en examiner sa fonction curative en cas d'atteinte à la sécurité de l'humain (II).

I. La fonction préventive de la Première Loi de la Robotique contre les atteintes à la sécurité de l'humain

Un dispositif de sécurité. Asimov explique dans sa préface que ses Lois forment un « dispositif de sécurité » pour protéger les humains. La Première Loi, qui prédomine sur toutes les autres⁹, remplit une fonction préventive en ce qu'elle a pour objet d'interdire au robot de porter atteinte à un être humain. La formulation de cette Loi est telle que chaque mot est d'une richesse qui permet à l'écrivain de décliner à loisir les récits autour des notions

³ *Runaround*, mars 1942.

⁴ Préface à son *Cycle des robots*.

⁵ Ibid.

⁶ Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL)).

⁷ « Lignes directrices en matière d'éthique pour une IA digne de confiance », Groupe d'experts de haut niveau

sur l'intelligence artificielle, Commission européenne, avril 2019. p. 6.

⁸ Chapitre 1, *The Bicentennial Man*, février 1976.

⁹ Les Lois d'Asimov forment une pyramide de Kelsen selon F. Defferard, « La science-fiction, source littéraire du droit. L'exemple des Trois Lois de la Robotique », in *Les fictions en droit*, F.X. Roux-Demare et M.-C. Dizès, dir., 2018, p. 11.

d'atteintes par action et par abstention de l'humain (A), et d'en ménager toute une série de nuances (B).

A. La formulation de la prohibition de toute atteinte par action et par abstention de l'humain

L'implantation de règles de comportement protectrices de l'humain. La formulation même de sa Première Loi, selon laquelle « *un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger* », relève d'une injonction sous une forme juridique du romancier afin de mettre en place une protection de l'humain résultant autant d'une action du robot, imposant alors une obligation de ne pas porter atteinte à l'homme (1) que d'une abstention du robot, impliquant alors une obligation d'assistance à l'humain (2).

1. L'obligation pour le robot de ne pas porter atteinte à l'humain par action

Une obligation contraignante pour les robots. Les robots d'Asimov sont suffisamment forts et intelligents pour être capables de mettre l'humain en danger, parfois de manière volontaire lorsque la Première Loi est altérée¹⁰. L'injonction de la Première Loi imposant au robot de ne pas porter atteinte à l'humain revêt donc une importance cruciale pour la sécurité humaine.

En l'état actuel des technologies, un robot est insusceptible de causer une atteinte volontaire à un être humain, c'est-à-dire par un effet de sa volonté consciente. Le *Cycle des robots* est très moderne étant donné que les règles de comportement du robot ne proviennent pas des robots eux-mêmes, car ce ne sont que des machines sans conscience, mais sont imposées

par l'homme au robot lui-même qui se voit alors contraint par ce code intégré dans son cerveau positronique.

Bien évidemment, dans la réalité, comme toute machine travaillant auprès de l'humain, le robot peut être une source d'atteinte à l'humain de manière accidentelle. En robotique industrielle, les dangers à redouter ont souvent une origine humaine, même si les sécurités mises en place sur et autour des robots, comme les limiteurs de vitesse, les détecteurs de présence ou les protecteurs, réduisent considérablement les risques. La directive Machines 2006/42/CE du 17 mai 2006 fixe d'ailleurs les exigences essentielles de santé et de sécurité pour les machines mises sur le marché européen afin de prévenir toute atteinte à l'humain¹¹.

Dans les nouvelles d'Asimov, l'humain est responsable de l'édiction et du respect de la norme. C'est pour cette raison qu'en cas de dysfonctionnement du robot, comme dans *Menteur*¹², les protagonistes décortiquent ce qui a pu mal se passer pour tenter de retracer la réaction du robot afin de la comprendre et de déterminer l'origine du problème. Il s'agit d'une forme de retour d'expérience destiné aussi à éviter qu'il ne se reproduise. Ce retour d'expérience est également au cœur des dispositifs de la directive Machines 2006/42/CE afin de prévenir les accidents. Avant de mettre sa machine sur le marché, le fabricant du robot doit ainsi délivrer une notice d'instructions qui relève notamment les risques qu'elle présente pour son utilisateur au regard d'une expérience antérieure ou d'une étude d'accidentologie¹³. De même, il repose sur l'employeur d'informer les salariés qui utilisent ces machines sur les retours d'expérience acquise permettant de supprimer certains risques (art. R. 4323-1, 4°, du C. trav.).

¹⁰ V. Le petit robot perdu (Little Lost Robot), mars 1947.

¹¹ Directive 2006/42/CE du Parlement européen et du Conseil du 17 mai 2006, relative aux machines et modifiant la directive 95/16/CE, JOUE n° 157 du 9 juin 2006, p. 24 et s., transposée dans le Code du travail par le décret n° 2008-1156 du 7 novembre 2008 relatif aux équipements de travail et aux équipements de

protection individuelle, portant transposition de la directive « machines », JORF n° 0262 du 9 novembre 2008, p. 17243, texte n° 10.

¹² *Liar!*, mai 1941.

¹³ § 1.1.2. ou encore § 1.7.4.2., Annexe I « Règles techniques en matière de santé et de sécurité », décret du 7 novembre 2008, préc.

Un robot hobbien. La règle asimovienne affirmant qu'« un robot ne peut porter atteinte à un être humain » n'est pas sans rappeler les infractions d'atteintes volontaires à la vie et à l'intégrité physique ou psychique de la personne réprimées par le Code pénal.

Parmi les atteintes volontaires, il faut spécialement songer au meurtre (art. 221-1 du C. pén.)¹⁴. Chez Asimov, la Première Loi constitue explicitement un rempart contre le meurtre d'êtres humains par les robots. Dans la nouvelle *Le petit robot perdu*¹⁵, la robopsychologue Susan Calvin explique que la domination des humains étant mal supportée par le robot qui leur est supérieur, seule la Première Loi permet de le maintenir dans son état d'esclave, car « sans elle, au premier ordre que vous donneriez à un robot, vous seriez un homme mort ». La Première Loi permet donc de laisser le robot à sa place d'esclave incapable de se rebeller. Mais on perçoit qu'en la supprimant ou en la réduisant, le robot n'hésiterait pas à attaquer l'homme, comme dans *Le petit robot perdu* où il finit par agresser physiquement un humain en l'attrapant violemment et en le faisant tomber. On réalise finalement que même si Asimov a voulu encore montrer les bienfaits de sa Première Loi et les dangers à la contourner, il a été tenté de faire frissonner le lecteur en plongeant dans le complexe de Frankenstein. En déformant la théorie du philosophe anglais Thomas Hobbes développée dans le *Léviathan*, on pourrait même suggérer que les robots d'Asimov constituent des robots hobiens car, ils veulent par essence du mal aux humains, démontrant alors que le robot serait un loup pour l'homme... sans la Première Loi d'Asimov.

A côté du meurtre, la question porte surtout sur les atteintes volontaires à l'intégrité de la personne, spécialement les violences (art. 222-7 à 222-16-2 du C. pén.). En jouant sur le mot « atteinte », Asimov envisage aussi bien une atteinte physique que psychologique. Il en va

également ainsi en droit puisque, même si le législateur n'a pas défini la notion de violence, l'incrimination figure dans la division « Des atteintes à l'intégrité physique ou psychique de la personne ». Dans la nouvelle *Menteur !*, le robot télépathe Herbie, capable de pénétrer les pensées humaines, applique strictement la Première Loi de la Robotique au point de mentir à tous ceux qui l'interrogent sur les pensées secrètes des autres. Ses mensonges sont une manière d'éviter de blesser les sentiments des êtres humains et de réduire leur estime de soi, ce qui constituerait, selon les propres affirmations du robot, une forme d'atteinte qui heurterait la Première Loi. Cela n'est pas sans rappeler le mensonge bienveillant de Saint Thomas d'Aquin. Toutefois dans *Menteur !*, la potentielle atteinte psychologique qu'aurait pu commettre le robot s'il avait choisi de ne pas mentir est autrement plus légère que celle habituellement retenue par les tribunaux. En effet, s'ils sanctionnent les agressions psychiques n'ayant aucune atteinte corporelle, ils estiment de manière constante que l'acte doit être de nature à impressionner vivement la victime ou à lui causer un choc émotif¹⁶. Finalement, « c'est donc bien dans l'ordre émotif, et non corporel, que se situe la violence psychique. Il n'est plus question de s'intéresser à la trace physique de l'agression, mais de se prononcer sur son ressenti psychologique, sur ce qu'elle peut représenter de frayeur, d'angoisse, de sensibilité négative »¹⁷.

La question des circonstances aggravantes. Le droit pénal prévoit une série de circonstances aggravantes qui appellent deux remarques s'agissant des robots. D'une part, la qualité de la victime pourrait jouer comme circonstance aggravante des infractions de meurtre ou de violences volontaires commises par les robots d'Asimov. Face à un robot fort et intelligent comme ceux du *Cycle des robots*, la circonstance aggravante liée à la « particulière vulnérabilité de

¹⁴ Art. 221-1 du C. pén. : « Le fait de donner volontairement la mort à autrui constitue un meurtre. Il est puni de trente ans de réclusion criminelle. »

¹⁵ Little Lost Robot, mars 1947.

¹⁶ Par ex., Cass. crim. 4 juin 2019, n° 18-84.720, *JurisData* n° 2019-009578.

¹⁷ Y. Mayaud, « Violences volontaires – L'action violente », *Rép. Dalloz droit pénal*, 2008 (actualisation en juillet 2019), n° 38.

la victime» devrait être aménagée pour tenir compte du fait que l'humain serait toujours en état de vulnérabilité face à cette machine, et qu'il irait de soi que cette situation serait apparente et connue du robot. D'autre part, pour sortir de l'univers romanesque, en l'état actuel des technologies, le fait qu'un robot soit, par exemple, programmé ou téléopéré par son propriétaire pour blesser une victime humaine pourrait être analysé par les tribunaux comme une circonstance aggravante tenant à l'usage ou la menace d'une arme, (art. 222-8, 10° du C. pén.)¹⁸. Même si le robot n'est pas, au sens du premier alinéa de l'art. 132-75 du C. pén., une arme conçue *ab initio* pour tuer ou blesser¹⁹, il est assurément susceptible de s'analyser en une arme par destination relevant de son alinéa 2²⁰. Les juges n'auraient certainement aucune hésitation à assimiler le robot à une arme par destination étant donné qu'ils admettent déjà à ce titre une automobile lancée volontairement contre la victime²¹, une tronçonneuse en fonctionnement²², et même une simple lime à ongles²³.

2. L'obligation pour le robot de ne pas porter atteinte à l'humain par abstention

Un robot prévenant. En imposant à ses robots de ne pas porter atteinte à l'humain en restant passif, Asimov les rend débiteurs d'un devoir d'assistance, qui est mis en relief dans la

nouvelle *Le petit robot perdu*. Il s'agit de l'histoire du robot Nestor modèle NS-2 appelé à travailler aux côtés d'humains dans une zone exposée aux radiations. Les utilisateurs avaient été contraints de supprimer l'obligation d'assistance de la Première Loi, car les robots non modifiés, face aux humains soumis aux radiations même sans danger pour eux, les empêchaient sans cesse de remplir leur mission pour les protéger.

Comment imaginer punir la passivité du robot asimovien ? En droit pénal français, la commission par abstention a longtemps posé problème. Ainsi, l'affaire de la séquestrée de Poitiers a montré que les violences volontaires consistent en des infractions de commission par un acte positif et non de commission par abstention²⁴. Historiquement, le législateur n'incriminait pas les abstentions fautives²⁵, seulement les infractions de commission par action, jusqu'à l'adoption de la loi du 25 octobre 1941²⁶. Asimov était donc dans l'air du temps lorsqu'il a affirmé qu'un robot ne peut, en restant passif, permettre qu'un être humain soit exposé au danger. Cette référence à l'abstention du robot semble également un écho au délit d'omission de porter secours à une personne en péril (art. 223-6, al. 2 du C. pén.)²⁷. Le péril consiste en un risque résidant dans l'exposition d'une personne à un risque physique, même s'il n'est pas en cours de

¹⁸ Art. 222-10 du C. pén. : « L'infraction définie à l'article 222-9 est punie de quinze ans de réclusion criminelle lorsqu'elle est commise : [...] 10° Avec usage ou menace d'une arme. »

¹⁹ En dehors des systèmes d'armes létaux autonomes (SALA), V. N. Nevejans, « La légalité des robots de guerre dans les conflits internationaux », *D.* 2016, pp. 1273-1278.

²⁰ Art. 132-75, al. 2 du C. pén. : « Tout autre objet susceptible de présenter un danger pour les personnes est assimilé à une arme dès lors qu'il est utilisé pour tuer, blesser ou menacer ou qu'il est destiné, par celui qui en est porteur, à tuer, blesser ou menacer. »

²¹ Cass. crim., 14 mars 1989, *Bull. crim.* 1989, n° 126 ; *RSC* 1989, p. 738, obs. G. Levasseur.

²² Cass. crim., 9 janv. 1986, *Gaz. Pal.* 1986, 2, p. 598.

²³ CA Riom, 15 janv. 2003, n° 02/00581, *JurisData* n° 2003-209222.

²⁴ CA Poitiers, 29 nov. 1901, *DP.* 1902, 2, 81, note Le Poitevin ; *S.* 1902, 2, 305, note Hémaré.

²⁵ Ph. Bonfils, « Entrave aux mesures d'assistance omission de porter secours », *J.-l. pén.*, Art. 223-5 à 223-7-1, Fasc. 20, 2017 – mis à jour en mars 2018, n° 23.

²⁶ L. 25 oct. 1941, *JO* 26 oct. 1941 ; *JCP G* 1942, III, 5053.

²⁷ Art. 223-6, al. 2 du C. pén. : « Sera puni des mêmes peines [cinq ans d'emprisonnement et 75 000 euros d'amende] quiconque s'abstient volontairement de porter à une personne en péril l'assistance que, sans risque pour lui ou pour les tiers, il pouvait lui prêter soit par son action personnelle, soit en provoquant un secours. »

Et dans une moindre mesure, on peut également penser à l'homicide ou aux blessures par imprudence (art. 221-6 et 222-19 du C. pén.), J. Pradel et M. Danti-Juan, *Droit pénal spécial. Droit commun – droit des affaires*, éd. Cujas, coll. « Référence », 5^e éd., 2010, n° 336.

réalisation²⁸, quelle que soit sa provenance²⁹, mais d'une certaine gravité³⁰. Il peut aussi bien concerner la vie, la santé ou l'intégrité physique³¹. Asimov est donc allé beaucoup plus loin que le droit français dans le *Cycle des robots* quand il estime que le robot est débiteur d'une obligation d'assistance même lorsque le danger est très faible, comme dans *Le petit robot perdu*.

Un robot très doué. Le délit d'omission de porter secours à une personne en péril (223-6, al. 2 du C. pén.), que nous venons d'évoquer, suppose que l'agent ait conscience du péril couru par la victime, puisqu'il doit en comprendre la gravité, mais, malgré tout, s'abstenir de lui porter secours³². Dans le *Cycle des robots*, les machines très intelligentes sont capables de détecter les situations dangereuses pour l'homme. Toutefois, dans la réalité, pour qu'un robot soit capable de mettre en œuvre ce devoir d'assistance, il devrait être en mesure d'apprécier la nature et l'intensité de l'atteinte, et faire également preuve d'une capacité d'analyse et de déduction. Cela impliquerait, par exemple, qu'un robot qui verrait une personne en train de fumer devrait lui ôter la cigarette des lèvres, car elle s'exposait à un risque de cancer du poumon. Si en l'état actuel des techniques, ce type de machine prévenante pour l'homme est encore loin de voir le jour, les robots d'Asimov sont non seulement pourvus de toutes les qualités intellectuelles pour comprendre qu'un humain est en danger, mais également de toutes les capacités techniques pour pouvoir réagir promptement. Il en va ainsi dans la nouvelle *Robbie*³³, où le robot nounou intervient à grande vitesse pour

sauver un enfant alors que personne n'avait eu le temps de réagir.

B. Les contours de la Première Loi de la Robotique

Les nuances de la Première Loi de la Robotique. La formulation de la Première Loi d'Asimov est cependant suffisamment floue pour permettre de générer d'excellentes histoires au regard des exceptions que l'auteur aménage (1) et des limites de sa Loi (2).

1. Les exceptions à la protection de l'humain contre des atteintes

Un robot altruiste. Dans la nouvelle *La preuve*³⁴, la robopsychologue Susan Calvin évoque une hypothèse d'école selon laquelle un robot face à un dément sur le point d'incendier une maison pleine d'habitants aurait pour réaction de l'arrêter, quitte à tuer l'humain pris de folie s'il y était contraint, car il doit éviter un péril plus grand. Dans cette situation, il devrait à la fois « *enfreindre la Première Loi pour obéir à cette même Première Loi, mais sur un plan plus élevé* ». Cette question, qui fait appel à notre devoir en tant qu'humain face aux périls courus par d'autres, est surprenante lorsqu'elle est abordée par Asimov. D'une part, la notion de péril plus grand ne figure pas dans la Première Loi et résulte donc d'une nouvelle interprétation d'Asimov. D'autre part, elle pousse encore plus loin le concept juridique d'état de nécessité, qui est un fait justificatif en droit pénal (art. 122-7 du C. pén.)³⁵. L'état de nécessité peut être invoqué dans un souci altruiste lorsque c'est autrui qui court un danger. Même si cette désignation vise souvent

²⁸ D. Rebut, « Omissions de porter secours – Entrave aux mesures d'assistance », *Rép. Dalloz droit pénal*, 2003 (actualisation en août 2018), n° 23.

²⁹ *Ibid.*, n° 28.

³⁰ Comme une menace pour la vie ou la santé de la personne, CA Rouen, 31 mars 1949, *D.* 1950, somm. p. 9.

³¹ Ph. Bonfils, *op. cit.*, n° 54.

³² Cass. crim., 26 avril 2006, n° 05-86.735, *JurisData* n° 2006-033676. – CA Caen, 17 septembre 2004, *JurisData* n° 2004-273306.

³³ *Robbie*, septembre 1940.

³⁴ *Evidence*, septembre 1946.

³⁵ Art. 122-7 du C. pén. : « N'est pas pénalement responsable la personne qui, face à un danger actuel ou imminent qui menace elle-même, autrui ou un bien, accomplit un acte nécessaire à la sauvegarde de la personne ou du bien, sauf s'il y a disproportion entre les moyens employés et la gravité de la menace. »

un proche ou un membre de la famille, la jurisprudence prend également en considération le danger couru par un parfait étranger³⁶. Ce danger pourra être d'ordre physique³⁷, comme dans *La preuve*, ou d'ordre simplement moral³⁸. Dans tous les cas, la réaction devra être nécessaire à la sauvegarde et mesurée quant aux moyens employés. Aussi, le fait pour un robot de tuer une personne pour éviter qu'elle en tue plusieurs autres en incendiant leur maison, comme dans l'exemple de la robopsychologue, serait certainement analysé par les tribunaux comme nécessaire et mesuré. Mais Asimov va plus loin que notre droit pénal français, puisqu'il n'hésite pas à demander à ses robots d'intervenir au péril de leur propre sécurité ou santé mentale, comme dans *La preuve*, où la robopsychologue Susan Calvin souligne bien que le fait d'avoir tué un humain rendrait fou le robot lui-même.

Un robot héroïque. La réflexion menée autour de l'exemple de l'incendie volontaire par un dément dans la nouvelle *La preuve* correspond également à la piste pénale de l'abstention d'empêcher un crime ou un délit contre l'intégrité corporelle (art. 223-6, al. 1^{er}, du C. pén.)³⁹. En droit français, serait punissable celui qui n'aurait pas empêché un incendie volontaire criminel susceptible d'entraîner la mort de plusieurs personnes⁴⁰. Les tribunaux sont très exigeants et estiment que la dissuasion verbale (ayant avorté) ne suffit pas⁴¹, car l'infraction doit être stoppée même en portant atteinte à l'intégrité corporelle ou à la vie de l'auteur de l'infraction empêchée⁴². Une fois encore, les nouvelles d'Asimov se

rèvelent particulièrement nuancées d'un point de vue juridique.

2. Les limites de la Première Loi de la Robotique

La situation de dilemme du robot. La situation de dilemme du robot est évoquée dans *La preuve*. Le robot ne peut agir sans enfreindre la Première Loi car, quelle que soit la solution adoptée par la machine, au moins un humain sera blessé ou mort. Si on sait que cette nouvelle a été publiée en 1946, on constate la qualité et le caractère innovant de la réflexion d'Isaac Asimov. C'est la prémisse d'un véritable questionnement philosophique, qui a fini par émerger chez les penseurs dans les années 60⁴³ sous la forme de l'expérience de pensée du dilemme du tramway. Il s'agit du conducteur imaginaire d'un tramway emballé qui est placé en situation de choix entre se diriger vers une voie où cinq hommes travaillent et une autre où se trouve un seul homme. Dans les deux hypothèses, sa décision aboutit à au moins un mort, mais dans l'un des cas, davantage de personnes seront sauvées. Le dilemme consiste donc à savoir qui sauver quand tout le monde ne peut pas l'être. Le dilemme du tramway est aujourd'hui invoqué pour réfléchir sur l'éthique des véhicules autonomes et permet de s'interroger sur de multiples variantes, comme la balance entre la vie du passager de la voiture autonome et celles de plusieurs piétons, par exemple. Toutefois, dans la situation actuelle, un robot qui devrait choisir entre deux alternatives contradictoires pourrait se trouver en situation de blocage. Dans une étude, des chercheurs ont programmé des règles éthiques de

³⁶ Comme l'excès de vitesse d'un chauffeur de taxi pour conduire à l'hôpital une cliente nécessitant des soins urgents, CA Paris, 24 avr. 2000, *JurisData* n° 2000-119275.

³⁷ Risque d'atteinte à la vie, à l'intégrité ou à la santé d'une personne, Cass. crim., 27 janv. 1933, *DP* 1933, 1, p. 489. – T. corr. Colmar, 27 avr. 1956, *D.* 1956, p. 500.

³⁸ C. Mascala, « *Faits justificatifs. – État de nécessité* », *J.-Cl. pén.*, Art. 122-7, Fasc. 20, 2012 (mis à jour en septembre 2019), n° 27.

³⁹ Art. 223-6, al. 1^{er} du C. pén. : « Quiconque pouvant empêcher par son action immédiate, sans risque pour

lui ou pour les tiers, soit un crime, soit un délit contre l'intégrité corporelle de la personne s'abstient volontairement de le faire est puni de cinq ans d'emprisonnement et de 75 000 euros d'amende. »

⁴⁰ Par ex., un époux a été condamné pour ne pas avoir empêché l'incendie volontaire commis par sa femme, Cass. crim. 17 déc. 1959, *D.* 1960, p. 151, note P. Bouzat.

⁴¹ Cass. crim. 26 mars 2002, n° 01-83.032.

⁴² D. Rebut, *op. cit.*, n° 140.

⁴³ Ph. Foot, « The Problem of Abortion and the Doctrine of the Double Effect », *Oxford Review*, n° 5, 1967.

comportement dans un robot contraint de respecter la Première Loi de la Robotique afin de tester ses réactions en situation de dilemme entre deux objectifs prioritaires consistant à sauver deux humains fictifs⁴⁴. Ils ont constaté que le robot parvenait dans un peu plus de la moitié des essais à sauver au moins un humain, mais restait bloqué dans tous les autres cas. Dans *La preuve*, le robot n'est pas bloqué, mais il devient fou en violant la Première Loi.

L'incomplétude de la Première Loi sans l'obligation d'assistance à l'humain. Les Lois d'Asimov sont énoncées par l'homme et programmées dans le robot pour prévenir les dommages causés à l'humain. Ce rôle préventif des accidents repose pour Asimov sur la société U.S. Robots, c'est-à-dire sur le fabricant. Cette solution est cohérente étant donné que c'est cette société qui édicte les Lois de ses propres machines. Lorsque l'on sort de ce schéma, la fonction préventive est incomplète. Ainsi, dans la nouvelle *Le petit robot perdu*, l'utilisateur du robot a supprimé du cerveau positronique à l'insu de tous l'obligation d'assistance à l'humain de la Première Loi pour lui permettre de travailler avec eux dans une ambiance dangereuse. Dans l'histoire, la suppression de cette obligation a eu pour effet de « dégrader » également la protection contre les atteintes par action⁴⁵. On constate que les deux énonciations de la Première Loi forment un tout inséparable. L'auteur en tire alors des conséquences particulièrement intéressantes. Il estime que la suppression de la protection contre les atteintes à l'humain par abstention a permis au robot de mentir, de manipuler les autres robots, et même de s'attaquer à l'homme. Cela signifie que si on touche à la Première Loi même pour partie, l'humain perd le bénéfice de la protection dans son entier.

La question de savoir qui implantent les règles éthiques de la machine est importante aujourd'hui pour le développement des véhicules autonomes. Il semble difficile

d'admettre que cette implémentation doive être décidée par le seul fabricant, puisqu'il pourrait alors choisir arbitrairement l'éthique à introduire dans le système de sa voiture. Face à deux choix ayant une valeur opposée, le fabricant pourrait considérer que le passager de son véhicule devrait toujours être sauvé au détriment de la vie d'autres personnes, comme des piétons. Comme ce choix éthique ne saurait pas plus venir de l'utilisateur du véhicule autonome, on doit considérer qu'il devrait appartenir à la société tout entière par le biais de son législateur. De ce point de vue, les Lois d'Asimov sont intéressantes, car dans la nouvelle *La preuve*, le romancier exprime l'idée que les Lois qui sont incorporées dans le robot constituent les principes directeurs d'une grande partie des systèmes moraux humains de la société en général, de sorte qu'en respectant les Lois, le robot respecte les valeurs éthiques humaines. Bien entendu, si en l'état de l'art, les valeurs éthiques sont encore aujourd'hui difficiles à programmer dans le robot, la Première Loi pose un principe sur lequel on ne peut qu'être d'accord : l'éthique du robot n'a pas à protéger une personne spécifique, mais doit permettre à toute personne de l'être dans n'importe quelle situation, même en état d'indignité.

L'indifférence à l'indignité de la victime humaine. Dans la nouvelle *La preuve*, lors d'un échange avec la robopsychologue Susan Calvin, l'un des personnages estime que la question de la culpabilité humaine est indifférente à l'injonction d'agir faite au robot, car « *il n'appartient pas au robot de décider si un être humain mérite ou non la mort* ». Aussi, il estime qu'en ce cas, le robot protégerait de la même manière la personne qu'elle soit coupable ou innocente. En responsabilité civile, cela n'est pas sans faire penser à l'adage *Nemo auditur propriam turpitudinem allegans*, selon lequel il serait possible au nom de la morale d'écarter tant l'action que le droit à indemnisation des victimes indignes. La

⁴⁴ AFT Winfield, C. Blum and W. Liu, « Towards an Ethical Robot: Internal Models, Consequences and Ethical Action Selection », pp 85-96 in *Advances in Autonomous Robotics Systems*, Lecture Notes, Computer Science, Volume 8717, Eds. Mistry M,

Leonardis A, Witkowski M and Melhuish C, Springer, 2014.

⁴⁵ On ne dit pas comment dans la nouvelle...

jurisprudence se refuse à appliquer l'adage⁴⁶, qui est considéré comme étranger aux règles de la responsabilité civile. Malgré tout, la situation d'« immoralité » de la victime peut non seulement avoir une influence sur la nature de la responsabilité⁴⁷, mais surtout s'analyser en une faute susceptible de limiter, voire d'exclure son indemnisation⁴⁸. De plus, il apparaît que l'indignité de la victime dépend de la nature de l'atteinte à la victime. L'adage joue pleinement lorsqu'elle ne souffre que d'un préjudice purement économique⁴⁹, mais ne permet pas d'exclure l'indemnisation de ses préjudices corporels, quelle que soit sa faute ou sa fraude, en raison de l'existence de la règle de la réparation intégrale des préjudices corporels⁵⁰. En tous les cas, la Première Loi de la Robotique doit donc être appliquée indépendamment de la valeur morale individuelle de l'humain, car le robot n'a pas à émettre un jugement de valeur sur leur comportement⁵¹.

La Première Loi de la Robotique n'a pas qu'une fonction préventive. On peut aussi lui découvrir une fonction curative en cas d'atteinte à l'humain.

II. La fonction curative de la Première Loi de la Robotique en cas d'atteinte à la sécurité de l'humain

Une responsabilité exclusivement humaine. Asimov propose au lecteur une vision très réaliste de la robotique. Plus encore, il crée un

univers à la juridicité cohérente, puisqu'il décrit la nature juridique des robots, lesquels sont des objets de droit insusceptibles d'engager leur propre responsabilité (A), et les responsabilités imputables à un être humain (B).

A. Un robot objet de droit, insusceptible de responsabilité

Le robot est une machine. Malgré l'intelligence exceptionnelle de ses robots, Asimov ne les considère pas plus que des machines appropriées et mises à la disposition des humains. Elles ne constituent que des objets de droit (1). Il en découle très logiquement que le robot ne peut être responsable en cas d'atteinte à la sécurité d'un humain (2).

1. Un robot, objet de droit

Un robot sans âme. La science-fiction se plaît à mettre en scène des robots doués de libre arbitre, de conscience ou de volonté. Il en va ainsi des Hubots de la série suédoise *Real Humans*⁵² où la conscience est introduite dans quelques machines de manière illégale⁵³. Dans la nouvelle *L'homme bicentenaire*, Asimov donne à son androïde Andrew Martin une âme qui l'amène à réclamer sa liberté à son propriétaire, puis le statut de sujet de droit à un tribunal. En dehors de cette nouvelle, dans le *Cycle des robots*, la question de l'existence d'un « esprit » chez les robots reste posée. Dans la nouvelle *Le petit robot perdu*, Asimov évoque

⁴⁶ Cass. 1^{re} civ., 14 déc. 1982, RTD civ. 1983, p. 342, obs. G. Durry ; Cass. 2^e civ., 19 févr. 1992, JCP 1993, II, n° 22170, obs. G. Casile-Hugues ; Cass. 1^{re} civ., 17 nov. 1993, RTD civ. 1994, p. 115, obs. P. Jourdain ; Cass. 2^e civ., 2 nov. 1994, D. 1996, Somm. p. 28, obs. J. Mouly ; 10 juin 1999, Dr. et Patrimoine 1999, n° 74, p. 95, obs. F. Chabas.

⁴⁷ Le dommage subi par un voyageur sans titre de transport ferme à la victime l'exercice d'une action contractuelle, mais lui laisse tout de même la possibilité d'être indemnisée par une action en responsabilité quasi-délictuelle, Cass. 1^{re} civ., 17 nov. 1993, RTD civ. 1994, p. 115, obs. P. Jourdain.

⁴⁸ Pour la riposte de la victime à une agression, Cass. crim., 16 oct. 1975, RTD civ. 1976, p. 359, obs. G. Durry ; 31 oct. 1979, RTD civ. 1980, p. 575, obs. G. Durry ; pour

l'auteur de la faute dommageable, Cass. 1^{re} civ., 14 déc. 1982, RTD civ. 1983, p. 342, obs. G. Durry ; *contra*, Cass. com., 7 déc. 1982, RTD civ. 1983, p. 536, obs. G. Durry.

⁴⁹ G. Viney, P. Jourdain et S. Carval, *Traité de droit civil. Les conditions de la responsabilité*, sous la dir. de J. Ghestin, LGDJ, 2013, n° 274.

⁵⁰ D. Mazeaud, « La résistance de la règle morale dans la responsabilité civile », D. 2002, p. 2559, n°s 7 et 10.

⁵¹ A moins que la personne coupable d'un méfait soit sur le point de mettre en danger de nombreuses personnes, comme avec l'exemple de l'incendie volontaire...

⁵² *Real Humans : 100 % humain* (*Äkta människor*), série télévisée suédoise créée par Lars Lundström, 2012.

⁵³ Saison 1, épisode 8.

l'âme du robot : « *Qu'est-ce donc qui lui donne une âme d'esclave ? Uniquement la Première Loi !* » Toujours dans la même nouvelle, il évoque le subconscient des robots : « *Ils sentent, dans leur subconscient, que les hommes leur sont inférieurs.* » Dans la nouvelle *Raison*⁵⁴, il évoque « *les mystères de « l'âme » robotique.* » Cependant, le robot n'est pas responsable au sens moral. En effet, dans sa préface au *Cycle des robots*, l'auteur souligne que les robots ne pensent pas vraiment, ils ne font qu'« *imiter le processus de la pensée humaine* ».

Bien entendu, en dehors de la littérature et du cinéma, le robot ne dispose d'aucune des qualités supérieures de l'homme, comme la conscience ou le libre arbitre. Les chercheurs estiment qu'il est encore trop tôt pour discuter de la « possible émergence des fonctions humaines dans le robot : comme la conscience, le libre arbitre, le sentiment de dignité, les émotions et ainsi de suite »⁵⁵. Même si certains roboticiens tentent d'apprendre au robot à reconnaître les émotions humaines⁵⁶, ou à inspirer de l'empathie aux humains en copiant leurs mimiques ou leur gestuelle⁵⁷, il n'en demeure pas moins que la machine elle-même ne ressent aucune émotion et se trouve seulement programmée pour apprendre à les simuler. C'est le phénomène de l'empathie artificielle⁵⁸. De même, le robot a une forme d'intelligence dédiée à la tâche qu'il doit réaliser, mais reste incapable d'accomplir autre chose même très simple, tandis que l'humain est naturellement doté d'une intelligence étendue et d'une compréhension intuitive de son monde. Le robot aujourd'hui est incapable d'appréhender d'une manière profonde les

conséquences de ses actions. Il en découle qu'il ne saurait être considéré comme moralement responsable en cas d'atteinte à la sécurité de l'humain.

Le robot, objet de droit. Dans la préface du *Cycle des robots*, Asimov ne considère pas le robot autrement que comme une machine. Dans la nouvelle *La preuve*, l'auteur précise que cette machine ne bénéficie pas de la loi asimovienne assurant « *l'inviolabilité des personnes privées* » et ne peut rien posséder.

La pertinence de cette analyse juridique est telle qu'on la trouve aussi dans notre droit positif français. En effet, celui-ci considère que le robot est un objet de droit, puisque c'est une chose, par opposition aux personnes. Seules les personnes sont « susceptibles de jouir et d'exercer des droits »⁵⁹, car elles constituent des sujets de droit. Elles sont dotées de la personnalité juridique qui consiste en « l'aptitude à être titulaire actif ou passif de droits subjectifs »⁶⁰. La doctrine majoritaire considère que la personne est celle « à qui le droit objectif accorde des droits subjectifs réunis en un patrimoine »⁶¹. Lorsque cette personne est un être humain, elle bénéficie, à l'instar des personnes privées de la nouvelle d'Asimov, d'une forme de sanctuarisation de son corps, puisque celui-ci, « considéré comme juridiquement inviolable, est protégé contre toutes les atteintes que les tiers pourraient prétendre lui faire subir pour satisfaire des intérêts privés »⁶². La personne humaine est donc protégée contre toute atteinte au droit au respect, à l'inviolabilité et à l'intégrité de son

⁵⁴ *Reason*, avril 1941.

⁵⁵ [Notre traduction] G. Veruggio, dir., « EURON roboethics roadmap », EURON roboethics Atelier, Genoa 27 feb. - 3 march 2006, juillet 2006, PDF, p. 7, *Roboethics.org* [online], <http://www.roboethics.org/atelier2006/docs/ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf>

⁵⁶ L. Devillers, « Rire avec les robots pour mieux vivre avec », *CNRS Le journal* [en ligne], 19 juin 2015, <https://lejournel.cnrs.fr/billets/rire-avec-les-robots-pour-mieux-vivre-avec>

⁵⁷ P.-Y. Oudeyer, « Robotique : les grands défis à venir », *Futuribles* n° 399, mars-avril 2014, p. 20.

⁵⁸ A ce sujet, V., S. Tisseron, *Le jour où mon robot m'aimera. Vers l'empathie artificielle*, éd. Albin Michel, 2015.

⁵⁹ Y. Buffelan-Lanore et V. Larribau-Terneyre, *Droit civil. Introduction. Biens. Personnes. Famille*, Sirey, 17^e éd., 2011, p. 273.

⁶⁰ F. Terré, *Introduction générale au droit*, Précis Dalloz, 6^e éd., 2003, p. 177, n° 182.

⁶¹ X. Labbé, note sous TGI Lille, 28 sept. 1995, *D.* 1995, 29.

⁶² J. Penneau et E. Terrier, « Corps humain – Bioéthique », *Rép. Dalloz civ.*, juillet 2019 (actualisation : Septembre 2019), n° 3.

corps, en vertu de l'art. 16-1 du C. civ.⁶³. Toutefois, à côté de la personne physique, la qualité de sujet de droit est également accordée à la personne morale par le biais d'un mécanisme fictionnel qui a pour fonction de donner la vie juridique à une entité non vivante.

La nature juridique d'objet de droit du robot ne saurait être remise en cause au regard des évolutions législatives faisant de l'animal un être doué de sensibilité (art. 515-14 du C. civ.), puisqu'il reste un objet de droit. En effet, en énonçant que, « sous réserve des lois qui les protègent, les animaux sont soumis au régime des biens » (art. 515-14 *in fine* du C. civ.), le législateur refuse d'attribuer à l'animal la personnalité juridique. Comme seul un sujet de droit peut engager sa responsabilité civile, le robot, tout comme l'animal, ne peut être considéré comme responsable des dommages qu'il cause. À cet égard, Asimov a réussi son pari d'échapper au complexe de Frankenstein, puisque le robot est maintenu dans son rôle strict de machine en portant son carcan de soumission à l'homme.

Toutefois, la pertinence de la classification entre les objets et les sujets de droit semble remise en question par le Parlement européen dans sa résolution du 16 février 2017 concernant les règles de droit civil sur la robotique. En effet, il estime que « l'autonomie des robots pose la question de leur nature à la lumière des catégories juridiques existantes ou de la nécessité de créer une nouvelle catégorie dotée de ses propres caractéristiques et effets spécifiques » (point AC). Il invite donc à penser que le robot pourrait ne plus être qu'un simple objet de droit en raison du développement de ses capacités à prendre des décisions de manière autonome.

2. Un robot irresponsable

Un robot irresponsable pénalement. Selon la Cour de cassation, « toute infraction, même

non intentionnelle, suppose [...] que son auteur ait agi avec intelligence et volonté »⁶⁴. Pour être pénalement responsable, l'agent doit avoir été en mesure de comprendre et de vouloir ses actes. Dans le cas contraire, le législateur a créé une cause d'irresponsabilité pénale bénéficiant à celui qui était atteint, au moment des faits, d'un trouble psychique ou neuropsychique ayant aboli son discernement ou le contrôle de ses actes (art. 122-1, al. 1^{er}, du C. pén.). Dans une certaine mesure, cette situation est évoquée dans la nouvelle *Lenny* :

« Il ne l'a pas fait sciemment, riposta le Dr Calvin. Le cerveau de Lenny est déficient. Il ne pouvait pas connaître sa propre force ni la faiblesse humaine. En écartant le bras menaçant d'un être humain, il ne pouvait pas prévoir que l'os allait se rompre. Humainement parlant, on ne peut incriminer un individu qui ne peut honnêtement distinguer le bien du mal. »

Bien entendu, si le robot n'est pas responsable de cet accident en raison de ses déficiences, l'auteur n'a pas voulu pour autant le considérer comme responsable dans l'hypothèse où il aurait fonctionné correctement. En cela, il lui dénie toute qualité d'agent susceptible d'engager sa responsabilité pénale. Le caractère personnel de la responsabilité pénale (art. 121-1 du C. pén.)⁶⁵ impose de punir personnellement celui qui a commis l'infraction. Or, le robot ne saurait avoir conscience de ses agissements, puisqu'il ne constitue qu'une machine programmée pour obéir aux ordres de l'homme. Même si le droit français reconnaît aujourd'hui, à l'art. 121-2, al. 1^{er}, du C. pén., la responsabilité pénale des personnes morales, qui constituent à l'instar des robots des entités inanimées, il n'en demeure pas moins que les infractions commises doivent l'avoir été par leurs organes ou représentants, donc par une personne

⁶³ Art. 16-1 du C. civ. : « Chacun a droit au respect de son corps. Le corps humain est inviolable. Le corps humain, ses éléments et ses produits ne peuvent faire l'objet d'un droit patrimonial. »

⁶⁴ Cass. crim., 13 déc. 1956, n° 55-05772, *Bull. crim.* 1956, n° 840.

⁶⁵ Art. 121-1 du C. pén. : « nul n'est responsable pénalement que de son propre fait. »

physique, car « la personne morale n'est en France responsable que de façon indirecte »⁶⁶.

D'un point de vue plus général, aujourd'hui en robotique, la loi pénale ne saurait servir de modèle au codage du robot, car elle ne précise pas ce que l'agent peut faire et ce qu'il ne peut pas. On peut douter que le robot goûte le sel de la subtile absurdité des incriminations susceptibles à la fois de « s'analyser comme des autorisations contre paiement d'un prix autant que comme des interdictions sous peine d'avoir à payer ce prix »⁶⁷. Il y aurait fort à parier que la machine procède alors au calcul rapide des coûts-avantages d'une infraction dans une pure logique mathématique déshumanisée, et estime valable l'option de porter atteinte à l'humain s'il en avait, par exemple, les ressources financières.

Un robot irresponsable civilement. La responsabilité civile consiste à devoir réparer le dommage qu'une personne a occasionné à une autre. En la matière, l'irresponsabilité correspond à une cause de non-imputabilité morale. En effet, pendant longtemps, la responsabilité civile a exigé que l'acte soit imputable à l'agent. Il fallait alors se demander s'il avait agi « avec une liberté suffisante pour être déclaré responsable de ses actes »⁶⁸. Les personnes atteintes de troubles mentaux ou les *infans*, c'est-à-dire les enfants n'ayant pas encore atteint l'âge de raison, étaient ainsi considérés comme irresponsables civilement. Aujourd'hui, cette exigence d'imputabilité a disparu, de sorte que tant le dément, en vertu de l'art. 414-3 du C. civ.⁶⁹, que l'*infans*, par le biais de la jurisprudence⁷⁰, peuvent engager leur responsabilité civile. De même, s'agissant des personnes morales, en dépit de leur caractère fictif, leur responsabilité civile peut être engagée en raison de leur fait personnel⁷¹,

car « par l'intermédiaire de leurs organes, elles expriment une volonté (imputabilité morale) et agissent conformément à cette volonté (imputabilité physique). La différence avec les personnes physiques est que cette volonté est collective et qu'elle ne peut s'exprimer et se matérialiser que par la médiation d'êtres humains »⁷². Il en découle que le fait dommageable devant avoir été commis par un sujet de droit, le robot ne saurait être considéré comme responsable civilement.

B. Le dommage causé par un robot imputable à un être humain

L'humain est responsable de sa machine. La vision de la responsabilité par Isaac Asimov coïncide d'une manière étonnante avec l'application des règles du droit de la responsabilité civile qui est classiquement faite en France (1). Elle repose pour le romancier sur l'idée que la prise en charge de la sécurité des humains relevant exclusivement de l'homme, seul l'humain serait responsable de sa machine. Elle marque ainsi le fait que l'humain responsable ne saurait disparaître derrière la machine (2).

1. L'application des règles classiques du droit de la responsabilité civile

La responsabilité du fabricant du robot. Dans *le Cycle des robots*, Asimov identifie la société United States Robots comme le fabricant responsable des robots. La nouvelle *La preuve* comporte un paragraphe particulièrement évocateur :

« Vous savez que l'U.S. Robots est la seule à fabriquer des robots positroniques dans le système solaire [...]. Vous savez également que les robots

⁶⁶ J. Pradel, *Droit pénal général*, éd. Cujas, coll. « Références », 20^{ème} éd., 2014, n° 592.

⁶⁷ E. Dreyer, *Droit pénal général*, éd. LexisNexis, coll. Manuel, 4^{ème} éd., 2016, n° 90, p. 69.

⁶⁸ P. Jourdain, *Les principes de la responsabilité civile*, éd. Dalloz, coll. « Connaissance du droit », 8^{ème} éd., 2010, p. 72.

⁶⁹ Art. 414-3 du C. civ. : « 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. »

⁷⁰ Cass. ass. plén., 9 mai 1984, *Bull. civ. ass. plén.* n°s 2 et 3.

⁷¹ Cass. 2^e civ., 27 avr. 1977, *Bull. civ.* 1977, II, n° 108.

⁷² P. Jourdain, « Droit à réparation. Responsabilité fondée sur la faute. Imputabilité », *J.-Cl. civ.*, Fasc. 121-10, Art. 1382 à 1386, 2012, n° 9.

positroniques sont toujours loués, jamais vendus ; que la société en reste propriétaire et qu'elle est responsable de leurs actes. »

Nous sommes proches de la solution qui est admise aujourd'hui en droit de la responsabilité civile. Si un robot causait un dommage à un être humain en raison d'un défaut de sécurité provenant de son *hardware* ou de son *software*, la victime pourrait se prévaloir des dispositions des art. 1245 à 1245-17 du C. civ. pour engager la responsabilité du producteur du robot sur le fondement des produits défectueux⁷³, sauf si ce dernier pouvait invoquer le risque de développement⁷⁴. Dans ce dernier cas, la victime ne saurait alors se prévaloir d'un défaut de sécurité du robot, mais pourrait cependant encore attaquer le producteur sur le fondement de la garantie des vices cachés ou sur celui de la responsabilité pour faute⁷⁵.

En outre, bien qu'Asimov n'aborde pas vraiment les conséquences juridiques pour le fabricant en cas de violation de la Première Loi, la question a néanmoins été posée dans la nouvelle *Lenny* :

« Serez-vous d'accord pour décider sa destruction s'il [le robot] a violé la Première Loi ? »

Cette sanction ne s'adresse évidemment pas au robot lui-même, mais à l'humain qui l'aurait mal programmé, car le robot n'est pas autre chose qu'une machine.

La responsabilité de l'utilisateur du robot. Sauf dans la nouvelle d'ouverture *Robbie*, où le robot est au service de la petite Gloria en tant que robot nounou, tous les robots du *Cycle des robots* sont ultraspécialisés et dédiés à des tâches professionnelles dans des centres de production dans l'espace. On retrouve également aujourd'hui cette dichotomie puisque, à côté des robots dédiés aux

utilisateurs professionnels (robots collaboratifs, chirurgicaux, agricoles, ou de sécurité), il existe également de nombreuses sortes de robots destinés aux consommateurs (robots tondeuses, aspirateurs, ou compagnons). Aussi, quand un dommage n'est pas la conséquence d'un défaut de sécurité du robot, mais résulte de l'utilisation qui en est faite, la victime n'a pas d'autre solution que de diriger son action contre le propriétaire ou l'utilisateur du robot sur le fondement de la responsabilité du fait des choses⁷⁶.

Vers une remise en cause ? Cependant, une fois encore, la résolution du 16 février 2017 sur les règles de droit civil sur la robotique estime que « *plus un robot est autonome, moins il peut être considéré comme un simple outil contrôlé par d'autres acteurs (tels que le fabricant, l'opérateur, le propriétaire, l'utilisateur, etc.)* » (point AB). Pour la résolution, plus avant-gardiste qu'Asimov, cela impliquerait qu'un robot autonome ne puisse plus être qu'une simple machine-objet de droit, mais tende à devenir un agent autonome-sujet de droit. Le Comité économique et social européen (CESE) souligne l'absurdité d'une telle affirmation dans son avis du 31 mai 2017⁷⁷.

2. L'absence de disparition d'un responsable humain

L'absence de vide juridique en cas de dommage causé par un robot autonome et auto-apprenant. La résolution du 16 février 2017 sur les règles de droit civil sur la robotique s'interroge sur le risque d'une insuffisance des règles de responsabilité civile en présence de robots autonomes et/ou auto-apprenants. Elle estime que si un robot était en mesure de prendre des décisions de manière autonome, « *les règles habituelles ne suffiraient pas à établir la responsabilité juridique pour dommages causés par*

⁷³ N. Nevejans, *Traité de droit et d'éthique de la robotique civile*, LEH éditions, coll. « Science, éthique et société », 2017, n° 657 à 721.

⁷⁴ *Ibid.*, n° 705 à 707.

⁷⁵ *Ibid.*, n° 716 et s.

⁷⁶ N. Nevejans, *Traité de droit et d'éthique de la robotique civile*, op. cit., n° 722 à 752.

⁷⁷ § 3.33, Avis du CESE sur « L'intelligence artificielle : les retombées de l'intelligence artificielle pour le marché unique (numérique), la production, la consommation, l'emploi et la société », JOUE 31 août 2017, C 288, p. 1-9.

un robot, puisqu'elles ne permettraient pas de déterminer quelle est la partie responsable pour le versement des dommages et intérêts ni d'exiger de cette partie qu'elle répare les dégâts causés » (point AF). Elle affirme également que, malgré la directive sur les produits défectueux, « le cadre juridique actuellement en vigueur ne suffirait pas à couvrir les dommages causés par la nouvelle génération de robots, puisque celle-ci peut être équipée de capacités d'adaptation et d'apprentissage qui entraînent une certaine part d'imprévisibilité dans leur comportement, étant donné que ces robots tireraient, de manière autonome, des enseignements de leurs expériences, variables de l'un à l'autre, et interagiraient avec leur environnement de manière unique et imprévisible » (point AI). La résolution part donc du postulat non vérifié que ces robots rendront difficile la détermination du responsable en raison de la disparition du facteur humain, de sorte qu'il faudrait nécessairement faire évoluer le droit de la responsabilité civile.

Ces dispositions de la résolution sont si hardies que même Isaac Asimov n'est pas allé jusque-là, puisqu'il affirme que le fabricant U.S. Robots est responsable des actes des robots. De plus, en l'état de l'intelligence artificielle, les chercheurs estiment que le processus décisionnel des systèmes autonomes ne pourra jamais se passer de la présence humaine⁷⁸. Il en découle que même dans l'avenir, les tribunaux seront toujours en mesure de désigner l'humain responsable *in fine*, comme le concepteur, le fabricant, ou encore l'utilisateur.

Les dangers du recours à la personnalité juridique du robot. La résolution concernant les règles de droit civil sur la robotique évoque l'éventualité de « la création, à terme, d'une personnalité juridique spécifique aux robots, pour qu'au moins les robots autonomes les plus sophistiqués puissent être considérés comme des personnes électroniques responsables, tenues de réparer tout dommage causé à un tiers ; il serait

envisageable de conférer la personnalité électronique à tout robot qui prend des décisions autonomes ou qui interagit de manière indépendante avec des tiers » (§ 59, f). La résolution tente, pour ainsi dire, de faire ce que le robot Andrew Martin dans *L'homme bicentenaire* a mis deux cents ans à réaliser, c'est-à-dire à transformer un robot objet de droit en un robot sujet de droit. Dès le début de la nouvelle, Andrew est un robot différent des autres, qui finit, au fil des ans et des transformations corporelles, par devenir une véritable personne juridiquement reconnue après qu'il se soit doté d'un corps organique mortel.

La vision du Parlement européen n'est pas partagée par toutes les institutions. Le CESE s'oppose ainsi à « l'octroi d'une forme de personnalité juridique aux robots ou à l'IA et aux systèmes d'IA, en raison du risque moral inacceptable inhérent à une telle démarche »⁷⁹, et la Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST) souligne, dans son rapport de 2017, l'absurdité de qualifier les robots de personnes « puisqu'ils sont dépourvus de certaines autres qualités généralement associées aux êtres humains comme le libre arbitre, l'intentionnalité, la conscience de soi, le sens moral et le sentiment de l'identité personnelle »⁸⁰. Notre *Open Letter to the European Commission on « Artificial Intelligence and Robotics »*, signée par 285 chercheurs européens de différentes spécialités, marque également son opposition à la création d'une personnalité juridique du robot autonome⁸¹.

D'un point de vue juridique, si la résolution devait aboutir à faire du robot autonome et auto-apprenant une personne physique, cela aurait pour conséquence de lui donner tous les droits et libertés humains, notamment la protection contre les atteintes à sa dignité et à son intégrité physique, la protection de sa liberté individuelle ou l'instauration d'une

⁷⁸ V., par ex., W. Nothwang et ali., « The human should be part of the control loop? », *Resilience Week (RWS)* 2016, 214-220, DOI: 10.1109/RWEEK.2016.7573336.

⁷⁹ § 3.33, Avis du CESE, préc.

⁸⁰ § 201, Rapport de la COMEST sur l'éthique de la robotique, 14 septembre 2017,

https://unesdoc.unesco.org/ark:/48223/pf000025395_2_fre

⁸¹ 285 chercheurs européens en robotique, IA, droit, éthique, médecine, etc., <http://www.robotics-openletter.eu/>

égalité avec l'humain. Cette solution dépasse les prévisions des Lois de la Robotique d'Isaac Asimov dans le *Cycle des robots* puisque, s'il attribue des obligations aux robots, il ne leur confère aucun droit. En l'état actuel de la robotique autonome et auto-apprenante, les robots n'auraient que faire d'être des personnes physiques, faute de conscience, de volonté propre, ou de sentiments pour donner sens à cette élévation de leur statut juridique. Alors que *L'homme bicentenaire* nous fait réfléchir à ce qui nous définit en tant qu'humain, l'adoption d'une telle voie tendrait à flouter les frontières entre l'homme et la machine. Toutefois, il semble plus probable que la résolution suggère plutôt de faire du robot autonome une personne morale. Cependant, en tant que fictions, les personnes morales sont dépendantes d'humains qui leur donnent vie juridiquement. Or, le robot ne serait aucunement en mesure d'agir lui-même pour prendre des décisions complexes en tant qu'acteur juridique, de sorte qu'il faudrait faire appel encore une fois à des êtres humains pour le représenter. Ces montages juridiques, qui auraient nécessairement un coût pour le propriétaire du robot, profiteraient essentiellement au fabricant qui pourrait être déchargé du poids de sa responsabilité.

N. N.

Le robot, obéissant ? Réflexions juridiques et anthropologiques sur la deuxième loi de la robotique

Ugo BELLAGAMBA

Maître de conférences HdR

Centre de Recherche en Histoire des Idées et des Institutions politiques

Université d'Aix-Marseille

Écrivain de science-fiction

Introduction

1.- Vingt-sept ans plus tard. Sur la page de garde de l'édition poche, chez J'ai Lu, du *Robot qui rêvait*, parue en 1988 avec une magnifique couverture de Caza et des illustrations intérieures de Ralph McQuarrie¹, j'écrivais, d'une main émue, en 1992, à dix-neuf ans à peine : « *Isaac Asimov est mort le 6 avril 1992. Ses plus grandes œuvres sont le Cycle de Fondation et l'ensemble de ses écrits portant sur les robots. Il est l'un des fondateurs de la S.-F. moderne. R.I.P., Isaac* ». J'étais jeune, passionné, idolâtre. J'avais découvert Isaac Asimov quelques années plus tôt, et j'avais lu presque tout ce qui était disponible en France. Si quelqu'un venu du futur m'avait affirmé que le 26 avril 2019, soit vingt-sept années plus tard, je serai invité, moi, en tant qu'universitaire, historien du droit et des idées politiques, dans une Journée d'étude consacrée *entièrement* à l'étude des implications juridiques des *Trois Lois de la Robotique* d'Isaac Asimov, je ne l'aurais pas cru, même si mon visiteur s'était présenté en T.A.R.D.I.S., ou avait prétendu s'appeler R. Daneel. Je ne saurais donc occulter, ici, et même dans le cadre d'une publication, disons, académique, l'expression de mon émotion et de ma profonde reconnaissance aux organisateurs de cette journée à l'Université de Strasbourg, placée

sous l'égide du Centre d'études internationales de la Propriété intellectuelle, et sous la direction scientifique de Franck Macrez. Y participer fut un rêve éveillé, et une véritable épiphanie, me prouvant qu'en définitive, je ne m'étais pas trompé de discipline.

2.- De la fiction philosophique à la réalité économique. Mais, il y a plus important que ce pont temporel qui relie l'adolescent rêveur à l'universitaire réfléchi. C'est le basculement dans le quotidien, en France, en Europe, et presque partout dans le monde, après une époque où les robots n'effectuaient que des tâches répétitives au fond d'usines ultra-automatisées, vers ce qu'il convient d'appeler la robotique sociale, ou d'accompagnement, qui est déjà considérée par certains comme la troisième révolution industrielle. Une « *robolution* » même, pour reprendre l'amusante expression d'Arnaud Montebourg, en 2014, alors que ce dernier était le Ministre du Redressement Productif du gouvernement Ayraut sous la présidence de François Hollande et qu'il annonçait à la presse économique, le 4 mars, la création du premier fonds d'investissement exclusivement dédié à la robotique, « *Robolution Capital* »², et promettait des centaines de nouveaux emplois dans le domaine. Ainsi, les robots humanoïdes d'Asimov s'avancent vers nous, en surfant sur

¹ Isaac Asimov, *Le robot qui rêvait*, Paris, éd. J'ai Lu, 1988.

² Boris Manenti, « La France à l'aube de la Robolution », article en date du 4 mars 2014, in *Le*

Nouvel Observateur, publié en ligne : <https://o.nouvelobs.com/high-tech/20140304.OBS8430/la-france-a-l-aube-de-la-robolution.html>

la vague de la nécessité. Celle de l'aide à la personne.

3.- De la réalité économique à la réflexion juridique. Kate Darling, chercheuse au M.I.T. a trouvé les mots justes pour définir le robot social : « *un agent autonome disposant d'un corps physique (mobile) qui communique et interagit avec les humains, jusqu'à un niveau émotionnel, [et qu'il est important] de distinguer des ordinateurs inanimés autant que des robots industriels qui n'ont aucune aptitude à l'échange avec l'humain et ne peuvent mimer des comportements sociaux* »³. Et, afin de faciliter la mission d'aide, d'accompagnement et d'interaction de ces robots sociaux avec l'être humain, il faut leur trouver de nouvelles règles de comportement, ce qui rend, bien sûr, la question de leur obéissance aussi prégnante que celle de leur innocuité pour les personnes fragiles, vulnérables, parfois handicapées, qu'ils vont devoir aider au quotidien. Le retour d'Asimov était, en quelque sorte, programmé, et l'auteur l'avait, d'une certaine façon, prévu dans son approche narrative.

4.- Une intuition narrative. Dans la préface qu'il rédigea en 1986 pour les besoins du recueil *Le robot qui rêvait*, Asimov disait clairement, avec le bon sens rationaliste et le paternalisme quelque peu condescendant qui le caractérisaient si bien, pourquoi ses robots positroniques les plus complexes étaient humanoïdes : « *Nous disposons aujourd'hui d'appareils appelés robots, programmés par ordinateur, et employés dans l'industrie. Ce sont eux qui accomplissent de plus en plus fréquemment les travaux simples et répétitifs sur les chaînes de montage – ils soudent, fraisent, polissent, etc. ; et ils sont d'une importance croissante pour l'économie. Les robots sont devenus un domaine d'étude reconnu, et c'est le nom précis que j'ai inventé en 1942 qui le désigne : la robotique. Bien sûr, nous ne sommes qu'à l'aube de la révolution robotique. Néanmoins, la direction du mouvement est nette (...) L'avenir devrait nous permettre d'aller plus loin dans la direction que j'ai indiquée. Plusieurs grandes entreprises sont au travail sur des « robots*

ménagers » qui auront une apparence humaine, et accompliront certains travaux dont se chargeaient naguère les domestiques. Le résultat de tous ces efforts, c'est que je jouis d'une considérable estime de la part de ceux qui travaillent dans le domaine de la robotique »⁴. Et, aujourd'hui, sans qu'il soit nécessaire de retracer ici toute l'histoire de la robotique, il est aisé de comprendre à quel point l'excitation peut gagner le juriste, au-delà du seul jeu spéculatif consistant à imaginer des formulations et des syllogismes qui pourraient faire « descendre » les trois lois de la robotique, et, en particulier, ici, la deuxième, dans le champ des applications et des faits juridiques.

5.- Un critère d'acceptabilité. Bien sûr, ce qu'il faut envisager, lorsqu'on évoque l'obéissance du robot, c'est la manière dont celle-ci se comprend et s'exige en fonction d'un contexte. Si l'obéissance d'un robot qui n'exécute que des travaux de force, ou industriels, sur une lointaine colonie spatiale, et qui n'a de contact qu'avec un tout petit nombre d'êtres humains, est simple à obtenir, elle ne saurait être comprise de la même façon que l'obéissance plus subtile qui découle d'un tissu complexe d'interactions qu'aurait un robot humanoïde immergé dans une société civile. Et, cette question, cruciale autant pour le juriste que pour le philosophe, Isaac Asimov, se l'est naturellement posée à lui-même et y a répondu, dans une nouvelle, pour nous fondamentale, « Pour que tu t'y intéresses », écrite en 1974, et qui, dès ses premières lignes, aborde de front la problématique des interactions entre l'homme et la machine : « *La nécessité d'obéissance est permanente, [mais] si deux êtres humains donnent à un robot des ordres contradictoires, auquel le robot doit-il obéir ? (...) Nos robots n'ont été utilisés que dans des environnements particuliers : dans l'espace, où les hommes qui avaient des rapports avec eux étaient experts en la matière. Il n'y avait pas d'enfants, pas d'idiot, pas de criminels, pas d'ignares bien intentionnés. Mais même dans ce cas, il y a eu des occasions où des dégâts ont été causés par des ordres imbéciles ou irréfléchis (...) Sur la Terre, cependant,*

³ Riccardo Campa, "The rise of social robots: a review of recent literature", *Journal of Evolution & Technology*, The Institute of Ethics and Emerging Technologies,

Vol. 26-1, March 2016: <https://jetpress.org/v26.1/campa.htm>

⁴ Isaac Asimov, *Le robot qui rêvait*, Paris, op. cit., pp. 8-9.

il faudra que les robots aient un jugement »⁵. Une obéissance non plus spontanée, mais réfléchie, donc, c'est la condition *sine qua non* pour que les robots sociaux puissent être acceptés ; exister, en somme.

6.- Un plan simple. Ce qui nous amène à proposer un plan simple, humble et des plus logiques, dans la continuité de la problématique posée par l'auteur lui-même : comment l'on doit passer, en tant que juriste, de la question de l'obéissance sans discernement des robots industriels, simplement contraignante (I) à l'obéissance libératrice des robots sociaux, fondée sur la faculté de juger (II), ce qui nous amènera à convoquer les différentes formes de droit, de la loi au contrat, que l'histoire et la philosophie ont mises à notre disposition pour réfléchir à une société émergente, différente, exaltante, sans doute, mais qui doit savoir rester, *in fine*, humaine. Et, par ricochet, il faudra se demander dans quelle mesure cette approche asimovienne ne nous oblige-t-elle pas aussi à reconsidérer nos habituelles catégories juridiques. Car comment ne serait-elle pas « humaine », ou « personne », cette machine complexe, autonome, mobile et capable de discernement ?

I. De l'obéissance sans discernement...

7.- D'une problématique à l'autre. À partir d'une étude que j'avais déjà menée⁶, pour les besoins d'une journée d'étude sur *l'Homme Augmenté face au droit*, organisée à l'Université de Lille 2 en 2013, par le Centre de Recherches Droits et Perspectives du droit, sous la direction de Xavier Labbé, cette communication est l'occasion de montrer qu'en termes d'obéissance, le recours au statut juridique de l'esclave antique n'est pas la

meilleure approche (A), et qu'il faut se pencher sur les mécanismes logiques qui conditionnent l'usage d'un outil, puisque c'est ainsi que le robot est d'abord envisagé par Isaac Asimov, notamment dans les nouvelles du corpus qui mettent en scène les personnages de Powell et Donovan (B).

A. Les sources juridiques de l'obéissance robotique

8.- L'obéissance du robot n'est pas celle d'un esclave. Sur le plan de l'inspiration juridique, rien ne permet d'affirmer qu'Isaac Asimov aurait pu réfléchir sur le statut juridique de l'esclave antique, notamment en droit romain. Il est encore plus invraisemblable d'envisager qu'il aurait pu s'inspirer directement des dispositions du *Code Noir*, qui réglementaient l'esclavage dans les colonies françaises⁷. S'il peut être intéressant, comme je l'avais fait dans l'article susévoqué, de rapprocher la Première Loi de la robotique, qui prévoit qu'« *un robot ne peut porter atteinte à un être humain* », de la défense absolue faite aux esclaves de porter atteinte à l'intégrité physique de leurs maîtres, l'article 27 du *Code Noir* disposant que « *l'esclave qui aura frappé son maître, sa maîtresse, ou le mari de sa maîtresse, ou leurs enfants (...) sera puni de mort* », rien ne justifie le rapprochement de l'obéissance de l'esclave, fondée sur la peur et la contrainte, de celle du robot qui découle d'une programmation de son cerveau positronique. Dans le cas de la Première Loi, il s'agit d'une obligation de *ne pas faire*, alors que la Deuxième Loi, dans sa formulation, concerne principalement des obligations de *faire* quelque chose qui corresponde à l'ordre donné sans contrevenir à l'obligation précitée de ne pas nuire.

9.- Plutôt un dispositif de plus. De surcroît, et s'il fallait insister sur ce point, les robots ne sont pas tant des personnes qui, comme les esclaves

⁵ Isaac Asimov, « Pour que tu t'y intéresses », in *Le Grand Livre des Robots*, tome 1 : « Prélude à Trantor », Paris, éd. Presses de la Cité, 1990, pp. 476-477.

⁶ Ugo Bellagamba, « Du *Code Noir* de Louis XIV aux *Trois Lois de la robotique* d'Isaac Asimov », in *L'homme augmenté face au droit*, sous la dir. Xavier Labbé, Lille, Presses Universitaires du Septentrion, 2015, pp. 27-40.

⁷ Le *Code Noir* ou Édit du Roy servant de règlement pour le gouvernement et l'administration de Justice & la Police des Isles Françaises de l'Amérique & pour la Discipline et le Commerce des Nègres & Esclaves dans ledit Pays, donné à Versailles au mois de mars 1685, Chez la Veuve Saugrain, Paris, 1728.

antiques, subissent une réification par la norme, mais précisément des choses qui, à force de complexité, de puissance, et d'intelligence, donnent l'illusion, à leurs utilisateurs humains, qu'elles peuvent accéder à la conscience ou faire preuve de libre arbitre. Pour Isaac Asimov, il s'agit d'y réfléchir comme on le ferait d'un « *dispositif de plus* »⁸, dans la longue histoire des progrès techniques de l'humanité. La preuve en est que, même si l'un de ses principaux personnages, Susan Calvin, est « *robopsychologue* » et s'occupe de robots qui, dans leurs rêves, échappent au carcan des Trois Lois, Isaac Asimov définit le plus souvent les robots comme des machines, « *de l'acier, des fils électriques et un cerveau positronique (...) simplement programmé pour agir comme il le fait* »⁹. Il le rappelle dans un article tardif : « *j'ai traité les robots avec objectivité : produits par des ingénieurs, ils présentent des problèmes d'ingénierie (...) ce sont des machines, pas des métaphores* »¹⁰. Avec beaucoup d'humour, il contourne souvent la question de l'intelligence artificielle, comme dans sa nouvelle sur les voitures autonomes, « *Sally* », où il montre, un brin mutin, que ce sont les conducteurs humains qui projettent leurs sentiments sur les voitures « *automatiques* » et croient que « *les moteurs et les carrosseries ne sont pas deux choses séparées* »¹¹, comme s'il s'agissait de l'âme et du corps de ces robots à quatre roues. C'est l'être humain qui est victime de ses représentations, et c'est lui qui, par sa maladresse, est la source des dysfonctionnements de la machine.

B. Identification des blocages logiques

10.- L'art délicat du commandement. Le seul et véritable enjeu de la Deuxième Loi de la Robotique, dans la plupart des nouvelles ou

des romans d'Asimov, découle directement de la façon dont l'ordre a été donné au robot. L'erreur, pour Isaac Asimov, est toujours imputable à l'humain. Comprenez, par ricochet, que, dans le champ juridique, la loi n'est jamais aussi faillible dans ses effets que le législateur dans son travail de rédaction. Rien ne l'illustre mieux que les récits mettant en scène le couple d'ingénieurs Gregory Powell et Mike Donovan, dont le plus connu est sans doute « *Cercle vicieux* » (*Runaround*), qui date de mars 1942 et qui est la nouvelle dans laquelle les Trois Lois de la robotique sont explicitement formulées pour la première fois¹². Le contexte n'est pas celui de la cité, ou même de la civilisation galactique, mais bien celui d'une frontière lointaine, un monde encore sauvage, souvent désolé ou écrasé par un soleil trop proche : ici, Mercure. L'exploitation d'un minerai, le selenium, s'y fait par l'intermédiaire de robots puissants qui résistent mieux à la rudesse de l'environnement et aux radiations solaires que leurs maîtres humains ; mais, voilà que le robot SPD 13, surnommé « *Speedy* » ne revient pas de sa journée de travail. En se rendant sur place, nos infortunés pionniers trouvent un robot fou, qui tourne en rond autour d'un gisement, comme incapable de se décider entre exploiter le filon ou désobéir à l'ordre donné. Rapidement, la cause est identifiée : une fuite de gaz provenant du sous-sol de la planète risque d'endommager gravement le robot, et la Troisième Loi¹³, qui prévoit que les robots doivent préserver leur existence (car tous ont une valeur marchande importante) a été renforcée en raison de la valeur élevée des robots de la série SPD.

11.- L'envoûtant pouvoir de la déduction. Or, l'ordre donné au robot, par Mike Donovan, l'a été, souligne Powell, de façon trop « *ordinaire*,

⁸ Isaac Asimov, *Les Robots*, Paris, éd. J'ai Lu, 1976, p. 14 : « *Considérons un robot simplement comme un dispositif de plus. Il ne constitue pas une invasion sacrilège du domaine du Tout-Puissant, ni plus ni moins que le premier appareil venu. En tant que machine, un robot comportera sûrement des dispositifs de sécurité aussi complets que possible* ».

⁹ Isaac Asimov, « *Le meilleur ami du petit d'homme* », in *Le Grand Livre des Robots*, tome 1, op. cit., p. 17.

¹⁰ Isaac Asimov, *Gold. La fantascienza allo stato puro*, Bompiani, Italia, 1995, p. 185.

¹¹ Isaac Asimov, « *Sally* », in *Le Grand Livre des Robots*, tome 1, op. cit., pp. 21-24.

¹² Isaac Asimov, *Le Grand Livre des Robots*, tome 1, op. cit., p. 945 (extrait de la préface d'*Un défilé de Robots*).

¹³ « *Un robot doit protéger son existence dans la mesure où cette protection n'est pas en contradiction avec la première ou la deuxième loi* ».

sans le souligner en aucune façon »¹⁴, ce qui a pour conséquence de ne pas pousser le robot au sacrifice pour obéir, mais plutôt à rechercher « une position d'équilibre », d'où sa ronde perpétuelle. L'issue du texte est superbe : plutôt que de donner de nouveaux ordres au robot, Donovan et Powell choisissent d'accentuer l'effet de la Troisième Loi, en augmentant la dangerosité de l'environnement immédiat de Speedy, par une modification du taux d'oxyde de carbone qui a « une action corrosive appréciable »¹⁵ sur le métal. Powell utilise la force et la précision mécanique d'un autre robot, moins évolué que Speedy, qui obéit instantanément à son injonction et réussit parfaitement l'action requise, ce qui donne l'occasion à Isaac Asimov de démontrer l'efficacité d'un ordre bien formulé et bien prononcé. Mais la situation sur Mercure ne tourne pas à l'avantage des deux ingénieurs, et la cause est, une fois de plus, l'erreur humaine ; ici une simple erreur de calcul, « la concentration [en acide] n'étant pas suffisante »¹⁶ pour faire renoncer définitivement Speedy à obéir à l'ordre qu'il a trop mollement reçu.

12.- Jamais deux sans trois. Au final, c'est la Première Loi¹⁷ qui permet de débloquent la crise. À force de poursuivre la machine bloquée, Powell et Donovan finissent par se mettre en danger, et ce faisant, poussent Speedy à sortir de son cercle vicieux pour les secourir, comme d'ailleurs tentent de le faire aussi les autres robots présents, qui n'écoutent nullement l'ordre de ne pas intervenir qui leur est donné par leurs maîtres humains (qui commettent une nouvelle erreur, cette fois non de calcul, mais d'appréciation). Ce qui prouve par ricochet que, contrairement à ce que semble affirmer Franck Macrez dans son propre article, il n'y a aucune forme de hiérarchie kelsénienne entre les Trois Lois de la robotique, au contraire. Même si Isaac Asimov

relève qu'aucun robot ne peut résister « au potentiel de la Première Loi »¹⁸, toutes doivent bien être respectées *en même temps*. Le robot Speedy n'a pas simplement sauvé un ingénieur en danger : il a équilibré de façon rationnelle l'obéissance qu'il devait aux Trois Lois. Il a d'abord cherché à obéir à un ordre donné, tout en s'inquiétant de la préservation de son intégrité, pour continuer à fonctionner, jusqu'à ce qu'il se trouve confronté à la nécessité de sauver un être humain. La Première Loi n'a pas effacé les deux précédentes, elle les a justifiées. J'envisage les Trois Lois de la robotique à la manière des lois fondamentales de l'ancien Royaume de France, au sujet desquelles Jean Bodin rappelait, dans *Les Six Livres de la République*, en 1576, qu'elles étaient toutes « annexées et unies à la Couronne »¹⁹. Les principes de catholicité et de masculinité ne l'emportèrent jamais sur l'exigence de respecter l'indisponibilité et la continuité de la Couronne de France, comme l'a rappelé le Parlement de Paris dans l'arrêt Lemaistre du 28 juin 1593.

13.- Les limites de l'obéissance sans discernement. Ainsi, l'obéissance d'un robot qui n'est pas capable de mesurer lui-même l'absurdité d'une situation, l'existence d'un cercle vicieux, ne dépend en définitive que de la capacité de l'être humain lui-même à l'anticiper, au regard du contexte et d'adapter en conséquence le formalisme même de son commandement. D'ailleurs, à la fin de la nouvelle étudiée, la leçon a bien été comprise par les ingénieurs : « je l'ai envoyé à l'un des autres filons de selenium, avec l'ordre cette fois d'en ramener à tout prix »²⁰, dit Donovan à Powell à propos d'un Speedy, désormais parfaitement obéissant, soumis à toutes les contraintes que les maîtres humains peuvent formuler.

¹⁴ Isaac Asimov, « Cercle vicieux », in *Le Grand Livre des Robots*, tome 1, op. cit., p. 210.

¹⁵ *Ibid.*

¹⁶ *Ibid.*, p. 212.

¹⁷ « Un robot ne peut porter atteinte à un être humain, ni, restant passif, laisser cet être humain exposé au danger. »

¹⁸ Isaac Asimov, « Cercle vicieux », in *Le Grand Livre des Robots*, tome 1, op. cit., p. 214.

¹⁹ Jean Bodin, *Les Six Livres de la République*, Paris, Chez Jacques du Puys, Libraire Juré, 1576.

²⁰ Isaac Asimov, « Cercle vicieux », in *Le Grand Livre des Robots*, tome 1 op. cit., p. 215.

II. ... à la faculté de juger avant d'obéir !

14.- L'enfer pour le robot, c'est les humains.

Le problème reste que, sur la Terre, au cœur de la myriade de relations humaines, juridiques, économiques et sociales, les « cercles vicieux » potentiels seront infiniment plus nombreux et plus subtils. L'obéissance, l'articulation des Trois Lois, sera plus délicate. Ainsi, blesser un être humain, par exemple, pourra s'entendre sur le plan psychologique, et les robots devront, là encore, l'éviter. Comme ce pauvre modèle TN-3, surnommé « Tony », dans « Satisfaction garantie »²¹, une nouvelle qui date de 1951, et qui met en scène l'équivalent de ce que Kate Darling définit comme un robot social, voire, en jouant un peu sur le terme, un « compagnon ». Confronté à la solitude d'une épouse délaissée, Tony, au nom de la Première Loi, se découvre une mission inattendue : la séduire, au vu et au su de tous. Ce qu'analyse parfaitement la robopsychologue Susan Calvin. Peut-être le robot a-t-il suivi, au vrai, l'ordre informulé de sa maîtresse humaine, l'appel à une aventure ? Quoi qu'il en soit, l'illustration de la difficulté de socialisation du robot appelle à une réflexion sur la nécessité de son jugement moral, en premier lieu (A). C'est à partir de celui-ci que le robot pourra passer de l'obéissance « imposée », unilatérale, à une obéissance que nous qualifierons de « consentie », ce qui imposera une redéfinition de son statut social (B).

A. Le questionnement moral de la machine

15.- Le problème du robot menaçant. Du propre aveu d'Isaac Asimov, la longue nouvelle, ou novella, sur laquelle nous allons nous concentrer à présent, « Pour que tu t'y intéresses »²², initialement publiée dans le numéro de mai 1974 du *Magazine of Science Fiction and Fantasy*, et plusieurs fois reprise ensuite, est l'une des seules qui relève du

modèle « *Robot Menaçant* », et, pourtant, il la rapproche de « L'Homme bicentenaire », modèle de robot bienveillant s'il en est, comme si les deux constituaient, au fond, une justification de tout son travail narratif sur les robots. Elle est intéressante, dans le cadre de cette étude, car elle se concentre largement sur la Deuxième Loi et la nécessité de repenser en profondeur la question de l'obéissance du robot. Nous évoquerons aussi l'autre « borne » de cette réflexion, dans la deuxième et dernière sous-partie de cette étude.

16.- Une question de contexte bien plus que d'intelligence. La nouvelle précitée s'inscrit dans un contexte bien précis : celui du processus de socialisation des robots par la société U.S. Robots qui les commercialise depuis le début du cycle. La nouvelle s'ouvre *in medias res* sur une discussion animée, à la fois technique et théorique, entre Keith Harriman, le directeur de la recherche de U. S. Robots et un robot dénommé George Dix, qui semble bien décidé à changer les perspectives terrestres de la robotique, sous le regard énigmatique, voire lui aussi menaçant, d'une représentation holographique de Susan Calvin, qui tient évidemment de l'icône sacrée. C'est bien le changement de contexte qui est abordé par l'homme et la machine, après qu'ils ont rapidement mis de côté la question de l'intelligence artificielle. Isaac Asimov, une fois de plus, fait ici preuve d'une belle intuition, puisque, dans son univers diégétique, l'intelligence des superordinateurs n'a été utilisée que très ponctuellement, pour résoudre « le problème écologique qui menaçait la société humaine »²³ avant qu'ils ne se condamnent eux-mêmes au nom de la Première Loi de la robotique. En effet, si les intelligences artificielles s'étaient généralisées, nous dit Asimov, elles auraient fini par prendre en charge le destin de toute l'humanité, et auraient ainsi nui à son développement par excès de bienveillance et de surveillance.

²¹ Isaac Asimov, « Satisfaction Garantie », in *Le Grand Livre des Robots*, tome 1 op. cit., pp. 271-285.

²² Isaac Asimov, « Pour que tu t'y intéresses », in *Le Grand Livre des Robots*, tome 1 op. cit., pp. 473-495.

²³ *Ibid.*, p. 475.

17.- Des choix plus fréquents et moins graves.

Sur Terre, les choix à faire, pour un robot social, seront complexes, fréquents, mais pas nécessairement graves ou urgents. Ainsi, comme le souligne le directeur de la recherche, les dangers physiques sur Terre sont moins nombreux que dans l'espace : « *un robot peut exister depuis vingt ans sans jamais avoir eu à agir rapidement pour éviter qu'un être humain ne souffre, ou sans jamais s'être trouvé dans l'obligation de risquer sa propre destruction...* »²⁴. Mais, bien sûr, cela n'ôte rien au caractère permanent, crucial, de la loi d'obéissance, souligne l'auteur, puisque « *pendant tout ce temps, cependant, [le robot] ne cessera d'obéir aux ordres...* »²⁵. C'est donc bien, en toute logique, « *la Deuxième Loi [qui] sera plus profondément modifiée* »²⁶. Et les conséquences en sont bien comprises par Asimov : sauf en cas d'urgence, l'obéissance du robot sur Terre, en société, ne sera plus spontanée, mais réfléchie, répondant à certains critères.

18.- Vers des critères objectifs d'obéissance.

« *Aux ordres de qui* » le robot devra-t-il répondre ? Telle est la question et elle est fondamentale : confronté à des ordres contradictoires, « *le robot mettra plus de temps à obéir* », parce qu'il « *obéira de façon plus rationnelle* »²⁷. Et, pour ce faire, il faut bien sûr, comme en droit, identifier des critères objectifs d'obéissance. Tant que cela n'est pas fait, les cerveaux positroniques ne peuvent que se paralyser, s'enliser dans une impossibilité de choix, comme l'a fait celui de Speedy dans « *Cercle vicieux* ». Et, bien qu'il ne soit nullement juriste de formation, l'auteur s'engage dans une réflexion qui nous rapproche d'une typologie : le robot peut tenir compte, d'une part, des hiérarchies sociales, et, par exemple, se contenter d'obéir prioritairement aux détenteurs d'une autorité publique, ou, d'autre part, il peut tenir compte de la valeur morale des ordres qu'on lui donne, et par ricochet, évaluer le caractère vertueux des intentions de la personne qui le formule. Mais, en définitive, nous rappelle Asimov,

avec bon sens, l'autorité comme la vertu peuvent n'être qu'apparentes, et il est bien difficile pour un robot de le mesurer. Il faut donc qu'un robot social passe par un apprentissage de l'obéissance en situation.

19.- L'apprentissage de l'obéissance. Comme pour les enfants qui atteignent un jour l'âge de raison, cette fameuse faculté de jugement, entre la possibilité d'une action, le bénéfice qu'ils en attendent, et la sanction qui risque d'en découler, il faut que les robots puissent, en conscience, examiner la pertinence d'une action immédiate au regard de ses conséquences sociales, à plus ou moins long terme, et y souscrire ou pas. Asimov affronte l'analogie de front : « *Le robot n'a pas à former un jugement immédiat sur un être humain et la valeur de ses ordres. Il commence par obéir à tous les êtres humains, puis il apprend (...) C'est l'équivalent d'un enfant, au début, et on doit le surveiller constamment* »²⁸. Il s'agit bien ici d'un renversement situationnel tout à fait remarquable. C'est finalement à l'être humain de construire, avec le robot qui doit lui obéir, une relation fondée sur l'expérience et l'acceptation éclairée des règles de la société. En d'autres termes, il s'agit d'éducation.

20.- De l'esprit des lois robotiques. Et nous savons, ici, en tant que juristes, privatiste, publiciste, ou historien du droit et des idées politiques, qu'il n'y a pas de société ni d'État de droit sans une éducation à la norme juridique. Toutes les utopies, depuis celle de Sir Thomas More, toutes les grandes œuvres de philosophie politique, au premier rang desquelles je place, sans hésiter, *L'Esprit des Loix* du Baron de la Brède et de Montesquieu, l'ont scandé. Il faut que les enfants apprennent les lois par cœur, et pour cette raison, elles doivent être simples et peu nombreuses. Et les robots positroniques, enfants d'Asimov, en ont parfaitement conscience : ils ont appris les trois lois de la robotique, et ont fait l'expérience de la difficulté de leur application dans certaines circonstances. Ils en ont tiré les leçons, tant et si

²⁴ Ibid., p. 476.

²⁵ Ibid.

²⁶ Ibid., p. 477.

²⁷ Ibid.

²⁸ Ibid.

bien, que le robot a changé de nature : d'outil polyvalent employé dans les colonies de l'espace, il est peu à peu devenu le compagnon vigilant des êtres humains, d'abord dans un cadre purement domestique, comme « Robbie », dans la nouvelle éponyme, ou « Andrew », au début de « L'homme bicentenaire », puis en s'intégrant volontairement au milieu social, dans une perspective, disons, plus politique, au sens étymologique du terme, afin de partir à l'assaut d'une redéfinition de son rôle au cœur de la cité des hommes, sans remettre en cause les Trois Lois ; ainsi, « à la fin, c'est un membre à part entière de cette société » qui s'avance, sûr de ses droits et de ses libertés.

B. L'obéissance consentie du citoyen.

21.- Des mécanismes juridiques de la soumission volontaire à la loi. Pour comprendre la nature même de l'obéissance volontaire à la loi chez l'homme, et par ricochet, chez le robot qui est voué à intégrer la société, il suffit de remonter à Thomas Hobbes. La lecture du *Léviathan*, qui date de 1651, est, sur ce point édifiante : au chapitre XXVI, l'auteur examine la typologie des normes et distingue les lois « naturelles », qui sont l'œuvre de Dieu et qui existent de toute éternité, des lois « positives », qui sont le fruit des législateurs humains, au fil de l'espace et du temps. Mais, il y ajoute une troisième catégorie particulière, les « lois divines positives », qu'il définit comme, je cite, « celles qui, étant les commandements de Dieu, non de toute éternité, ni universellement adressés à tous les hommes, mais seulement à un certain peuple ou à certaines personnes, sont déclarées telles par ceux qui sont autorisés par Dieu à les déclarer »²⁹. Il faut comprendre, ici, bien sûr, les lois divines qui auraient été rédigées sous la dictée de la divinité, par les prophètes et les législateurs mythiques, comme c'est le cas notamment de Moïse, qui reçut, pendant l'Exode, les Dix Commandements de Dieu, sur le mont Sinaï, ou de Numa, roi des premiers temps de Rome.

²⁹ Thomas Hobbes, *Léviathan*, Paris, éd. Gallimard, coll. Folio Essais, n°375, pp. 430 et suiv.

³⁰ *Ibid.*, p. 433.

Dans ce cas, poursuit Thomas Hobbes, deux questions se posent : comment peut-on prouver la réalité d'une révélation ? Et, dans l'impossibilité rationnelle de le faire, sur laquelle l'auteur anglais ne s'attarde guère, comment peut-on accepter de se soumettre à ces lois sans être certain de leur origine divine ? La réponse à la deuxième question donne, je crois, la clef qu'il fallait pour comprendre pleinement la position d'Isaac Asimov quant à la loi d'obéissance. Hobbes écrit : « si la loi déclarée [comme une loi divine positive] n'est pas contraire à la loi de nature (qui est sans doute la loi de Dieu), et si l'on veut lui obéir, on se contraint soi-même par son propre acte »³⁰.

22.- De la soumission volontaire à l'acceptation sociale. Le robot qui interagit en société est donc celui qui va être capable de juger lui-même si l'ordre qu'on lui a donné, quelle que soit la personne qui l'a fait, est conforme aux raisons fondamentales pour lesquelles les Trois Lois de la robotique, et en particulier la deuxième, ont été forgées, et de décider lui-même s'il convient de s'y soumettre ou non. Cette faculté de se contraindre à l'obéissance dont il s'empare, et qui lui permet de faire la distinction entre l'obéissance aveugle et l'obéissance raisonnée, fait du robot un être social à part entière. Autant dire un être humain, même s'il n'en porte pas le titre : « les circuits de mon cerveau ressentent un besoin pressant de ne pas tenir compte de l'aspect extérieur dans le jugement d'un être humain, et ce besoin est plus fort que la distinction entre le métal et la chair »³¹, s'écrit le robot à la fin de la nouvelle étudiée.

Mais, reste la difficulté la plus élevée, qui n'est pas de nature juridique, mais plutôt d'essence politique : George Dix se demande quand la société humaine, faite d'êtres « biologiques », sera prête à accepter en son sein des « humains-de-notre-sort » ? Pour Asimov, l'acceptation viendra d'abord quand « une partie suffisante de la société » humaine sera robotisée. Puis, elle s'enracinera dès lors que les robots eux-mêmes consacreront du temps à « former une société à

³¹ Isaac Asimov, « Pour que tu t'y intéresses », in *Le Grand Livre des Robots*, tome 1, op. cit., p. 494.

part entière », ce qui obligera hommes et machines à coexister en bonne intelligence, dans le respect des « Trois lois de l'Humanique »³², qu'Isaac Asimov, comme surpris de sa propre audace, qui le pousse vers une redéfinition de l'humain, n'évoque qu'une seule et unique fois dans son œuvre, basculant plutôt vers la mise en œuvre de la fameuse Loi Zéro. C'est, au fond, la liberté civile, autant que la liberté politique, entendez celle de l'individu moderne comme celle du citoyen antique, qui sortent victorieuses de cette évolution de la deuxième loi.

Conclusion

23.- De la liberté d'obéir. Au final, le robot qui dispose de la faculté de juger n'est pas seulement destiné à prendre part à la vie citoyenne, il doit aller, et c'est logique, jusqu'à la reconnaissance de sa liberté en tant qu'individu. Le droit d'ester en justice, celui d'être propriétaire des fruits de son travail, et celui, bien sûr, d'exprimer ses opinions sur la scène publique. C'est « L'Homme Bicentenaire » qui transcende toutes les difficultés théoriques, et, grâce à l'inoubliable personnage du robot NDR, dit « Andrew », qui, en deux siècles d'existence, a su conquérir la liberté d'être humain. L'argument juridique avancé par Isaac Asimov, hélas, reste faible : « il n'y a aucune raison de refuser la liberté à un objet qui possède un cerveau assez développé pour

saisir le concept et souhaiter la condition »³³, écrit-il. Un tel argument, nous exposerait, s'il était pris à la lettre, à deux écueils : d'une part, accorder la liberté à nos bases de données juridiques, qui détaillent à la fois la définition et les conditions d'exercice de la liberté, et d'autre part, la refuser à celles et ceux qui, pour une raison ou une autre, n'ont plus la capacité de communiquer, ou d'exprimer valablement leur consentement. Le véritable argument, à la vérité, est encore plus simple, et il suffit de se tourner, après Thomas Hobbes, vers Jean-Jacques Rousseau. La seule véritable liberté du citoyen, c'est celle qui consiste à obéir à la loi qui découle du contrat social, nous dit à plusieurs reprises, le Citoyen de Genève, dans *Le Contrat Social* comme dans *Les Lettres écrites sur la Montagne*. Le robot Andrew mérite la liberté, simplement parce qu'il a choisi d'obéir à la loi des hommes, et de s'y soumettre, avec fierté, et jusqu'à sa dernière pensée.

Nos robots seront-ils libres, un jour ? Peut-être, si nous-mêmes le sommes suffisamment pour les intégrer volontairement au cœur de notre sociabilité. Mais, en serons-nous capables, ou, comme Susan Calvin, dans « Le robot qui rêvait »³⁴, lèverons-nous une arme pour qu'ils cessent d'exister s'ils osent remettre en cause leur programmation ?

U. B.

³² *Ibid.*, p. 495.

³³ Isaac Asimov, « L'Homme bicentenaire », in *Le Grand Livre des Robots*, tome 1, op. cit., p. 496.

³⁴ Isaac Asimov, « Le robot qui rêvait », in *Le Grand Livre des Robots*, tome 1, op. cit., pp. 23-32.

Le robot, entité autonome ? Entre norme et technique, contextes et limites de l'autonomie cybernétique

Jean-Marc DELTORN

Laboratoire E.A. 4375

Centre d'études internationales de la propriété intellectuelle
Université de Strasbourg

Introduction

1.- Le robot autonome au creuset de la science-fiction. L'image du robot qui surgit des récits de science-fiction est inséparable d'une certaine idée « d'autonomie ». Le robot calcule, se déplace, saisit un objet, dialogue avec l'humain sans que celui-ci ne doive en contrôler les opérations internes, ne doive en guider les mouvements, ni ne doive lui dicter ses réparties. Le robot semble agir en toute indépendance. Il détermine ses objectifs, les moyens nécessaires pour les atteindre, et manifeste l'expression de sa volonté. L'automate, le robot, l'intelligence artificielle, construits de toutes pièces, ordonnés par nos soins, semblent échapper ainsi, en partie au moins, au contrôle direct de leurs concepteurs.

Ressort dramatique de nombreux récits, du folklore au roman gothique, du Golem de la tradition juive à la *créature* de l'œuvre de Mary Shelley, la figure prométhéenne d'un être artificiel, créé et pourtant capable de décider de

lui-même et de « s'émanciper », a longtemps nourri l'imaginaire. L'avènement de la cybernétique lui a donné un second souffle. Non plus modelée d'argile ni cousue de chairs mortes, c'est un assemblage décidément technique qui anime désormais la créature. Éléments mécaniques, composants électroniques, senseurs, actionneurs, ordinateurs constituent à présent l'anatomie du « robot ». En deçà de ces rouages matériels, c'est une autre machinerie qui se déploie, un système conjoint de codes informatiques et de trames algorithmiques qui guident et animent l'automate, et donnent enfin vie au « cerveau positronique »¹ imaginé par Asimov.²

2.- Troisième loi et autonomie. C'est dans le récit « Cercle vicieux » qu'Isaac Asimov introduit sa troisième loi.³ Après l'obligation de ne pas porter atteinte à un être humain⁴ et l'obligation d'obéir aux ordres qui lui sont donnés par un être humain⁵, l'auteur propose de définir une troisième règle : « *Un robot doit protéger sa propre existence aussi longtemps qu'une*

¹ Il s'agit de l'unité centrale programmable, forgée d'iridium et de platine, conçue par Isaac Asimov pour animer les robots dans nombre de ses récits.

² Les avancées dans le domaine de l'apprentissage automatique, notamment des réseaux de neurones profonds, ont eu un impact considérable sur les développements robotiques récents. V. H.A. Pierson et M.S. Gashler. *Deep learning in robotics: a review of recent research*. Advanced Robotics, vol. 31, n° 16, 2017, p. 821-835.

³ En anglais « Roundabout », paru en mars 1942 dans la revue « Astounding Science Fiction » (la nouvelle, traduite par Pierre Billon paraît en France en 1967 dans le recueil « Les Robots » (sous le titre « Cercle fermé »), publié par OPTA).

⁴ La première loi dicte qu'« un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger. »

⁵ Selon la deuxième loi : « 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. »

telle protection n'est pas en contradiction avec la Première et/ou la Deuxième Loi. » Le robot, instruit de son « existence », a le devoir de se protéger sans pour autant enfreindre les exigences cumulées de la première loi et de la deuxième loi.⁶ La liberté de choix du robot s'exprime dans cet enchaînement de règles. Car si les trois lois imposent au robot un régime de contraintes, elles lui accordent une marge d'autonomie, un espace d'expression, dans lequel il peut décider seul. Une confrontation entre contraintes et objectifs, qui vire parfois à l'absurde, comme l'illustre Asimov dans « Cercle vicieux ». Sur Mercure, en 2015, les scientifiques Powell et Donovan, ont donné pour mission au robot SPD-13 d'aller chercher du sélénium. Mais le robot tarde à revenir. Partis à sa recherche, les deux protagonistes le découvrent enfin, tournant sans fin autour d'un gisement du minerai radioactif. Le robot ayant atteint un point d'équilibre entre la deuxième loi, lui imposant de suivre les ordres reçus et de prélever du sélénium, et la troisième, l'obligeant à se

protéger des radiations nocives émises par le minerai, ne savait comment trancher.⁷ La raison algorithmique du « cerveau positronique », soumise aux contraintes imposées par les trois lois, conduit à un cercle vicieux dans lequel le robot s'enferme. L'autonomie de la machine, si elle est supposée par la troisième loi ne se déploie que dans cet îlot somme tout bien étroit : face à une situation imprévue, entre capacités de décisions bornées et tabous impossibles à transgresser, la machine finit par « tourner en rond ».⁸ Alors peut-on parler d'autonomie pour le robot ?

3.- Du rêve au déploiement : le robot confronté au réel. Aujourd'hui le robot autonome n'appartient plus à la seule science-fiction, il fait partie de notre quotidien. Sorti du cadre de la littérature, son registre d'application se déploie dès à présent de l'espace domestique (le robot ménager, aspirateur⁹, tondeuse à gazon¹⁰) jusqu'au domaine industriel (les robots ont largement investi les usines, et les « cobots »¹¹ y occupent une place croissante). De l'utilitaire à l'intime¹²,

⁶ Pourtant Asimov ne s'interdit pas d'explorer d'autres configurations : dans la nouvelle « Le Robot qui rêvait » (titre original « Robot dreams » publié en 1986 et paru en France en 1988 aux éditions J'ai Lu), la structure fractale du cerveau positronique du robot, LVX-1 le conduit à rêver qu'il s'émancipe des deux premières lois et qu'il n'est seulement régi que par la troisième loi : « un robot doit protéger sa propre existence ».

⁷ Il faudra que Powell s'expose au danger radioactif pour briser la boucle. Redonnant priorité à la première loi afin de protéger la vie de l'ingénieur, le robot sortira enfin de son va-et-vient. Après avoir sauvé Powell, et avant de le renvoyer « au charbon », les ingénieurs décident de renforcer le poids de la seconde loi dans le cerveau positronique de SPD-13 pour éviter que la troisième loi ne vienne contrecarrer leurs ordres. Dans « Le petit robot perdu » (« Little lost robot », paru dans « Astounding Science Fiction » en 1947, publié en français en 1967 par OPTA), Asimov continuera d'utiliser la relation d'ordre entre les trois lois comme source de paradoxes.

⁸ La capacité des robots d'Asimov à décider par eux-mêmes est souvent source de curieux conflits. Dans « Menteur ! » (Publié en 1941 dans « Analog Science Fiction and Fact »), par exemple, un robot télépathe capable de lire les pensées, décide de mentir systématiquement pour éviter de froisser les humains. De même, dans « Un conflit évitable » (initialement

publié dans « Astounding Science Fiction » en 1950), les robots décident, suivant une logique aussi froide qu'imparable, de prendre le contrôle de l'humanité et de sacrifier certains humains pour préserver la majorité. (Les deux nouvelles paraîtront en France en 1967 chez OPTA au sein du recueil « Les robots » dans la traduction de Pierre Billon). Comme souvent chez Asimov, la logique imparable du robot (et, à travers elle, les règles dans lesquelles elle s'inscrit) entre en conflit avec le « sens commun ».

⁹ 25 millions de Roombas, le robot aspirateur produit par la compagnie « iRobot » (du nom de la nouvelle d'Asimov éponyme) équipent aujourd'hui les foyers à travers le monde (<http://www.iRobot.com>).

¹⁰ « Terra » est présentée par iRobot comme la première tondeuse véritablement « autonome » (<https://www.irobot.fr/terra>)

¹¹ Pour « collaborative robot », un robot conçu pour être en interaction directe et « collaborer » avec l'humain.

¹² Le robot accompagne l'humain dans la détresse, la maladie et la vieillesse (v. p. ex. M. Valentí Soler et al. *Social robots in advanced dementia*. *Frontiers in aging neuroscience* 7, p. 133, 2015 ; S. C. Chen, Jones, C., & Moyle, W. (2018). *Social robots for depression in older adults: A systematic review*. *Journal of Nursing Scholarship*, 50(6), p. 612-622.).

de la conduite automatique à la décision médicale, des assistants digitaux aux robots de service, du salon jusqu'au terrain militaire, les robots intègrent une panoplie de plus en plus vaste d'attributs et de fonctions que l'on supposait jusqu'à récemment l'exclusivité d'acteurs humains. Chacun pose à sa manière la question de l'autonomie de la machine.

L'ubiquité de robots capables de décider hors de l'emprise immédiate de leurs architectes, mais pourtant en interaction directe avec les humains, oblige à penser leur encadrement. Isaac Asimov a parmi les premiers envisagé la nécessité de réguler le robot, une obligation à laquelle ingénieurs, usagers et législateurs doivent à présent se confronter. Dans ce contexte, la question de l'autonomie – réelle ou imaginée – des robots apparaît incontournable, ne serait-ce que pour en préciser les contours. Ce travail de délimitation imposera d'abord de circonscrire le périmètre dans lequel l'autonomie des machines peut être définie (I) avant de considérer les moyens pratiques de la mesurer (II). Il nous faudra enfin envisager, au travers de quelques cas réels, les conséquences du déploiement en société de robots, « entités autonomes », et les moyens d'en mitiger les errements (III).

I.- Quels critères d'autonomie pour le robot ?

4.- La notion d'autonomie du robot ne peut être pensée *in abstracto* : elle ne prend sens que dans le contexte spécifique des systèmes techniques et des attributs (fonctionnels et applicatifs) auxquels elle s'applique (A). Dans ce contexte, son sens est sous astreintes : il y est déterminé par un faisceau de contraintes qui doit intégrer limitations techniques et impératifs

économiques dans lesquels le robot se déploie (B).

A. L'autonomie en contexte

5.- L'autonomie des personnes. L'autonomie est au centre d'une conception morale et philosophique qui lie les idées d'indépendance, de liberté et de contrôle. Être autonome, c'est d'abord pouvoir se gouverner soi-même, avoir la capacité de décider sans se soumettre de façon servile à une autorité extérieure.¹³ L'autonomie naît de la possibilité de se soumettre à sa propre loi : « l'obéissance à la loi qu'on s'est prescrite est liberté ». ¹⁴ Une loi choisie, mais quelle loi ? Kant précise : il n'y a d'autonomie que si l'individu suit les lois morales qu'il se fixe à lui-même.¹⁵ Or, ces lois ne sont « morales » que lorsqu'elles se conçoivent en tant que « législation universelle », c'est-à-dire qui puisse valoir pour tous. Ainsi « l'action sera morale si la règle qui y préside peut faire loi, c'est-à-dire faire monde. Par exemple, un monde qui adopterait le crime comme sa loi serait contradictoire ou impossible. C'est par là que le crime est immoral »¹⁶. L'individu autonome est donc son propre législateur. Il manifeste son autonomie si, « réfléchissant à sa conduite, il choisit volontairement et librement de se comporter de la façon qu'il juge être universellement la meilleure. Dans tout autre cas (si, par exemple, il suit les ordres qu'il a reçus, s'il obéit à la loi, s'il se conforme à son désir, etc.), il se comporte de façon hétéronome. »¹⁷

Le droit en donne une définition similaire : l'autonomie est le « pouvoir de se déterminer soi-même ; faculté de se donner sa propre loi »¹⁸ et, pour en limiter le registre à celui des

¹³ La racine grecque *αὐτονομία* définit l'autonomie comme « le droit de se régir par ses propres lois ».

¹⁴ Rousseau, *Du Contrat social*, I, p. 8. L'autonomie s'oppose ainsi au constat que « l'impulsion du seul appétit est esclavage » (Id.).

¹⁵ Cette loi propre que l'homme se fixe sur la base de sa seule volonté n'est universelle que lorsqu'il agit « selon la maxime qui peut en même temps s'ériger elle-même en loi

universelle » (Kant, *Fondements de la métaphysique des mœurs*, Le Livre de Poche).

¹⁶ Michaël Foessel, *Kant ou les vertus de l'autonomie*, *Études*, vol. 414, n° 3, 2011, p. 341-351.

¹⁷ Ronan Le Coadic, *L'autonomie, illusion ou projet de société ?* *Cahiers internationaux de sociologie*, vol. 2, n° 121, 2006, p. 317-340 (p. 319).

¹⁸ Gérard Cornu, *Vocabulaire juridique*, PUF, 2011. Le principe d'autodétermination, reposant sur l'article 8

individus, la rapproche de la notion d'indépendance, la « situation d'un individu qui exerce seul et en toute liberté les pouvoirs qui lui sont conférés »¹⁹. Une « liberté d'exercice » selon laquelle « chacun est maître de soi-même et exerce comme il le veut toutes ses facultés », qui évoque à son tour la notion de « capacité »²⁰, « l'aptitude de faire valoir par soi-même et seul un droit sans devoir être ni représenté ni assisté par un tiers »²¹. Indépendance, liberté d'exercice, capacité à agir forment ensemble un réseau de sens dans lequel la notion d'autonomie se déploie. Elle s'y décline sous diverses facettes, et selon plusieurs niveaux. Comme la liberté, l'autonomie du sujet peut s'exercer sous contraintes ; comme la capacité, elle peut être relative et limitée ; comme l'indépendance, elle suppose un espace dans lequel elle peut se déployer. Mais dans tous les cas, en philosophie comme en droit, l'acception classique de l'autonomie n'est conçue qu'en rapport aux personnes. Comment, autrement, mesurer la « volonté », comment penser le sens « moral », la notion de « devoir », s'il est question de biens ? Lorsqu'il s'agit de machines, et les robots ne sont que cela, il faudra chercher ailleurs, dans le domaine technique, les caractéristiques susceptibles de

définir une notion d'autonomie qui leur est véritablement applicable.²²

6.- L'autonomie technique du robot. Dans le domaine de la robotique, la notion d'autonomie repose d'abord sur la capacité pour un système d'accomplir des tâches sans requérir un contrôle humain permanent. C'est la capacité à « prendre des décisions (limitées) sur les tâches à exécuter, en fonction de perceptions et d'états internes, plutôt que de suivre une séquence d'actions prédéterminée basée sur des commandes préprogrammées »²³. L'Institute of Electrical and Electronics Engineers (IEEE) définit le robot comme un « système commandé par ordinateur programmé pour exécuter certaines tâches sans intervention humaine ».²⁴ Pour l'Organisation internationale de normalisation (ISO) le robot consiste en un « mécanisme programmable actionné sur au moins deux axes avec un degré d'autonomie, se déplaçant dans son environnement, pour exécuter des tâches prévues ».²⁵ Le robot y est décliné selon ses applications : il peut être « industriel » ou « de service ».²⁶ Les normes ISO définissent par ailleurs la notion d'autonomie comme la « capacité d'exécuter des tâches prévues à partir de l'état courant et des détectations, sans

de la Convention européenne des droits de l'Homme (CEDH) est notamment évoqué dans *Pretty c. Royaume-Uni*, 29 avr. 2002, CEDH req. n° 2346/02 : « Le pouvoir de décider de façon autonome ce qui convient le mieux à son propre corps est un attribut de la personne et de la dignité de l'être humain ». De même : *Evans c. Royaume-Uni*, 10 avr. 2007, CEDH req. n° 6339/50. : « la notion de vie privée, notion large qui englobe, entre autres, des aspects de l'identité physique et sociale d'un individu, notamment le droit à l'autonomie personnelle [...] ». Enfin, *Ternovsky c. Hongrie*, 14 déc. 2010, CEDH req. n° 67545/09 souligne : « the notion of personal autonomy is a fundamental principle underlying the interpretation of the guarantees of article 8 ».

¹⁹ Id.

²⁰ Id.

²¹ Id.

²² Nous ne discuterons pas ici de la question de la personnalité juridique des robots. Nous nous limiterons à une approche fonctionnelle de l'autonomie. Les robots (hors du cadre de la science-

fiction) ne se verront donc pas attribués « d'intention », de « volonté » qui leurs seraient propres, indépendamment des contraintes techniques imposées par leurs concepteurs (humains) : ils resteront, en droit comme en fait, des « choses ». Pour une critique raisonnée de la notion de personnalité juridique des robots, v. la lettre ouverte initiée par Nathalie Nevejans : <http://www.robotics-openletter.eu> ou encore J. Bryson et al. 2017, *Of, For, and By the People : The Legal Lacuna of Synthetic Persons*. *Artificial Intelligence and Law*, vol. 25, n° 3, p. 273.

²³ Matthias Scheutz & Charles Crowell, *The Burden of Embodied Autonomy: Some Reflections on the Social and Ethical Implications of Autonomous Robots*, Workshop on Robo-ethics at the International Conference on Robotics and Automation, p. 1, 2007.

²⁴ IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2. IEEE, 2017.

²⁵ ISO 8373:2012.

²⁶ Id.

intervention humaine ».²⁷ Il faut pourtant préciser cette définition et distinguer « l'automatisation » de « l'autonomie ». Un système automatisé fonctionne normalement sans intervention humaine, mais ne possède pas de capacité décisionnelle propre. Il ne fait que suivre pas à pas une séquence d'instructions, de règles, préalablement inscrites, de manière explicite, sous supervision humaine. Un système est dit autonome lorsqu'il émule le processus cognitif et permet une « prise de décision » indépendante de toute supervision humaine.

Au-delà des associations professionnelles ou des organismes de standardisation, la définition du robot a été posée dans le cadre des recommandations à la Commission concernant des règles de droit civil sur la robotique.²⁸ Le Rapport du 27 janvier 2017 propose d'établir une définition européenne commune des différentes catégories de robots « autonomes et intelligents » sur la base des trois caractéristiques suivantes : « la capacité d'acquisition d'autonomie grâce à des capteurs et/ou à l'échange de données avec l'environnement et l'analyse de données ; la capacité d'apprentissage à travers l'expérience et l'interaction ; la capacité d'adaptation de son comportement et de ses actes à son environnement. » : l'autonomie du robot est ainsi liée à la possibilité d'apprendre et de s'adapter en acquérant des informations issues de son environnement.

B. L'autonomie sous contraintes

7.- L'autonomie bornée de la machine. Alors que la seconde loi impose une relation d'ordre, un assujettissement du robot à l'humain, la troisième loi évoque la possibilité d'une émancipation : le robot devrait se préserver, protéger son existence. Pourtant, cette autonomie fait figure de leurre, car le robot reste *in fine* sous contrôle humain : l'étendue du domaine de décision de la machine est déterminée par un faisceau de contraintes

définies *ab initio* par les concepteurs du système. Ainsi, les propriétés d'un véhicule autonome sont régies par le jeu de capteurs dont il est équipé (radars, caméras, lasers), mais aussi par les capacités prédictives du modèle qui viennent interpréter les signaux des senseurs (un ensemble de réseaux de neurones, des modèles d'inférences, des règles codées « en dur ») autant que par les données d'apprentissages qui ont permis d'entraîner ce modèle. C'est dans ce domaine de représentations que le robot interprète, modélise, décide. Sa capacité d'expression est limitée par les attentes, les besoins exprimés par les concepteurs et par les conditions externes (le contexte) dans lesquelles il est déployé. Le véhicule autonome détermine un angle de rotation (tourner à droite ou à gauche), le contrôle de la vitesse (accélérer ou freiner) conforme au modèle et aux informations déduites des caméras, capteurs de distance, de vitesse qui l'équipent. C'est au travers des degrés de liberté qui lui ont été imposés que le robot exprime son « autonomie ». Le robot est autonome, certes, mais reste dans l'ombre de l'humain. L'autonomie de la machine s'exprime dans un environnement technique (capteurs, données, algorithmes, actuateurs) dans lequel le mode d'expression du robot est contraint. Ce n'est que dans ce contexte précis que la notion d'autonomie du robot doit être comprise.

8.- L'autonomie en trompe-l'œil du robot. Il faut donc préciser : l'autonomie du robot n'est pas du même ordre et ne se mesure pas selon les mêmes attributs, que l'autonomie—sociale ou cognitive—caractéristique de l'humain. Il ne faut pas mélanger les genres. Un rapport récent de la Commission européenne dirigé par Jim Dratwa, le rappelle : « L'autonomie au sens éthique du terme ne peut être attribuée qu'à des êtres humains. Il est donc quelque peu erroné d'appliquer le terme « autonomie » à de simples artefacts, même s'il s'agit de systèmes adaptatifs complexes très avancés ou même « intelligents ». La terminologie des systèmes

²⁷ Id., section 2.2.

²⁸ Résolution du parlement européen, 16 février 2017 contenant des recommandations à la Commission

concernant des règles de droit civil sur la robotique 2015/2013(INL).

« autonomes » s'est cependant largement répandue dans la littérature scientifique et dans le débat public pour désigner le degré le plus élevé d'automatisation et le degré le plus élevé d'indépendance vis-à-vis des êtres humains en termes d'« autonomie » opérationnelle et décisionnelle. Mais l'autonomie dans son sens originel est un aspect important de la dignité humaine qui ne doit pas être relativisé. »²⁹

Alors faut-il même considérer que les trois lois puissent s'appliquer aux robots autonomes ? La réponse de Colin Angle, PDG de iRobot est en ce sens éclairante de pragmatisme : « si le Roomba suit les trois lois c'est parce qu'il est conçu pour être intrinsèquement fiable et sûr [première loi], pour remplir sa fonction [deuxième loi] et pour être résistant [troisième loi]. Ce n'est pas la conséquence d'une « IA » ou le résultat d'une « intention » de la part du robot s'il suit les lois d'Asimov, mais c'est simplement parce que les trois lois sont alignées avec le fait d'être un bon produit robotique. »³⁰

Parler d'autonomie pour les robots doit donc être considéré avec prudence si c'est (seulement) d'une « automatisation du plus haut degré » ou d'un « alignement fonctionnel » en tant que produit commercial qu'il s'agit en pratique. Pourtant, force est de constater que la délégation progressive d'attributs cognitifs à la machine — qu'il s'agisse de perception (l'analyse d'une image ou d'un son), d'interprétation de situations données (la présence d'un objet dans l'image ou d'un message dans le signal audio), ou de la prise de décision qui en découle — autant que le

déploiement des robots en société, nécessite un encadrement strict.

II. - L'autonomie du robot - une question de mesure

9.- On l'a vu, les robots sont parmi nous. Or, tous les robots ne sont pas également autonomes : l'indépendance du robot aspirateur diffère de celle de la voiture autonome, l'impact du robot-chirurgien, du robot-militaire, est d'un autre ordre que celui d'un assistant vocal. De fait, l'autonomie de ces systèmes se décline sur différents niveaux qui déterminent le cadre légal susceptible de leur être appliqué (A). Une tabulation de l'autonomie du robot qui n'est pas sans montrer ses limites (B).

A. Quantifier l'autonomie

10.- **Véhicules autonomes et niveaux d'autonomie.** En 1953, Asimov décrit dans « Sally » un futur dans lequel seules les voitures sans conducteurs équipées de cerveaux positroniques sont autorisées sur les routes.³¹ Si nous n'en sommes certes pas encore là, depuis les premières compétitions organisées aux États-Unis en 2004 par le « Defense Advanced Research Projects Agency » (DARPA), les progrès ont été remarquables.³² Aujourd'hui le développement de véhicules autonomes est devenu une réalité industrielle. Un effort justifié en partie par la volonté de réduire le nombre de victimes d'accidents de la route, aujourd'hui un enjeu de société majeur.³³

²⁹ Statement on Artificial Intelligence, Robotics and "Autonomous" Systems - European Group on Ethics, in Science and New Technologies, Brussels, 9 March 2018.

³⁰ Lex Fridman, Entretien avec Colin Angle, PDG de iRobot, 19 sept. 2019 (<https://lexfridman.com/colin-angle>).

³¹ La nouvelle « Sally » est publiée en 1953 dans la revue « Fantastic ».

³² Les challenges DARPA ont constitué une plateforme de référence dans laquelle le développement de véhicules autonomes a été encouragé et testé. Alors qu'aucun participant n'était parvenu à bout du

premier « Grand Challenge » (le meilleur candidat ne parvenant à parcourir que moins de 12 km sur les 240 km de piste dans le désert du Mojave initialement envisagés), 3 ans plus tard, toutes les équipes complétèrent l'épreuve dans le temps imparti. La dernière version du DARPA challenge imposait pourtant de compléter un parcours de 96 km en milieu urbain en présence de circulation et d'obstacles, dans le respect du code de la route.

³³ Le nombre de décès s'élève à environ 1,25 millions par an dans le monde, auxquels s'ajoutent 20 à 50 millions de victimes, blessées ou handicapées (source : Association for safe international road travel

Encadrer légalement le déploiement de véhicules autonomes repose sur une évaluation des fonctionnalités qui leur sont dévolues. Entre absence d'autonomie et autonomie totale, la classification établie s'étend sur 6 niveaux, dans lesquels le véhicule gagne progressivement en indépendance.³⁴ Au premier niveau (0) l'humain contrôle toutes les fonctions de conduite sans aucune assistance par le véhicule (hormis les indications visuelles ou auditives, d'un capteur de proximité ou d'un GPS par exemple, laissant au pilote la responsabilité de leur interprétation). Le niveau suivant (1) délègue certaines fonctions de base à un système automatisé (p. ex. un régulateur de vitesse ou un système d'aide au freinage ABS), mais le conducteur conserve le contrôle global du véhicule. Le niveau 2 permet une autonomie partielle de plusieurs fonctions de conduite sous la supervision du conducteur (par exemple, le contrôle conjoint de la vitesse et de la direction du véhicule pour le maintenir sur une voie, ou un système intelligent d'aide au stationnement entrent dans cette catégorie). Au niveau 3, le système est capable d'assumer totalement la conduite, mais uniquement dans des situations prédéfinies. Le conducteur peut ainsi laisser le « pilote automatique » contrôler de manière autonome le véhicule sur certaines portions du réseau routier (les autoroutes par exemple) ou dans certaines conditions (lorsque le véhicule est pris dans un embouteillage et avance « pas à pas »). Le niveau 4, le véhicule est susceptible d'une autonomie complète (sans requérir le conducteur) dans des configurations limitées (un système de parking automatique dans lequel le conducteur laisse le soin au véhicule de se garer seul entre dans cette catégorie). La décision d'initier le système autonome reste néanmoins la responsabilité du conducteur. Enfin, le cas d'une conduite complètement autonome sans nécessiter à aucun moment

l'engagement du conducteur entre dans le niveau 5.

11.- Organiser l'autonomie, du robot-médecin au robot-combattant. Dans le domaine médical, autre champ d'application privilégié de l'intelligence artificielle et de la robotique, l'autonomie s'organise là encore selon plusieurs niveaux, allant (comme pour les véhicules) de l'absence d'autonomie (niveau 0), jusqu'à l'autonomie complète (niveau 5).³⁵ Les niveaux 0 et 1 comprennent le cas des robots télé-opérés (p.ex. Da Vinci) selon qu'ils suivent directement ou assistent partiellement le geste du chirurgien. Les niveaux 2 et 3 accordent une autonomie partielle de certaines fonctions au système pour des tâches spécifiques, sous le contrôle d'un opérateur. Au niveau 4, le robot peut prendre des décisions médicales, mais sous la supervision d'un médecin. Le niveau 5 accorde une autonomie totale au robot (qui peut effectuer, par exemple, une intervention chirurgicale complète). De même, dans le domaine militaire, une classification permet d'organiser les robots selon 10 niveaux d'autonomie (allant des systèmes téléguidés jusqu'aux ensembles de robots collaboratifs).³⁶

B. Les limites d'une mesure de l'autonomie

12.- Une taxonomie trompeuse. Des véhicules, au domaine médical ou au terrain militaire, l'autonomie du robot s'égrène au fil d'un transfert de fonctionnalités de l'humain vers la machine. Le caractère polysémique de la notion d'autonomie (selon le champ de connaissances, en philosophie, en droit ou en technique) se retrouve au sein même des applications industrielles et des taxonomies qui y sont associées. Même s'il existe bien des points communs entre ces classifications (le premier niveau relève toujours d'une absence

<https://www.asirt.org/safe-travel/road-safety-facts/>).

³⁴ Résolution du Parlement européen du 15 janvier 2019 sur les véhicules autonomes dans les transports européens, 2018/2089(INI).

³⁵ G.-Z. Yang et al., 2017. Medical robotics—Regulatory, ethical, and legal considerations for

increasing levels of autonomy. *Science Robotics*, 2(4), eaam8638.

³⁶ G. M. Kamsickas & J. N. Ward, 2003. *Developing UGVs For the FCS program*. In *Proceedings of SPIE*, volume 5083, Orlando, USA.

d'autonomie alors que le dernier décrit une « autonomie totale »), le passage d'un seuil d'autonomie à un autre suppose la définition de critères adaptés au secteur d'application du robot (l'expression de l'autonomie du robot-chirurgien diffère ainsi en général de celle du robot-combattant ou du véhicule autonome). Or, ces attributs, ces fonctionnalités qui devraient assigner, de manière objective, à la machine un certain niveau d'autonomie restent sujets à interprétation. Un robot pourrait être ainsi considéré autonome selon un modèle de classification (dans un secteur donné, selon une norme donnée), et pas dans un autre. Cette spécialisation de la notion d'autonomie est, en partie au moins, le reflet des processus techniques qui animent la machine. Car même les modèles d'apprentissage les plus évolués n'opèrent que dans un régime d'utilisation bien délimité. Les modèles AlphaGo et AlphaZero de DeepMind sont capables de battre Lee Sedol, le champion du monde de Go, et de produire des coups qualifiés par certains observateurs de « créatifs » ou « surprenants »³⁷, mais échoueraient piteusement à Candy Crush (à moins, bien sûr, d'être réentraînés à cette fin, ce qui exige une adaptation du modèle sur lequel ces systèmes sont bâtis). Ce n'est en rien diminuer l'exploit technique : il reste exceptionnel. Mais l'intelligence artificielle que manifestent ces systèmes reste strictement bornée au domaine d'expression pour lequel ils sont conçus. Le Roomba n'est autonome que dans le seul cadre de sa fonction : nettoyer le sol d'un logement. Hors de ce cadre, il est absolument « incapable ». C'est donc seulement à une autonomie conditionnée au régime d'utilisation du robot qu'il est fait référence dans les divers systèmes de classification.

Mais même à se limiter à un contexte applicatif donné, définir précisément les « configurations » qui déterminent l'affectation d'un niveau d'autonomie particulier au robot reste un problème ouvert. Comment s'assurer *a priori*

que le robot se trouve bien dans les conditions caractéristiques d'un niveau d'autonomie donné ? Là encore, face à l'apparence d'objectivité que présuppose l'existence même d'une échelle d'autonomie, la prudence s'impose. Un rapport du département américain de la défense le souligne, « ces taxonomies sont trompeuses... L'autonomie du système est un continuum allant du contrôle humain complet de toutes les décisions à des situations où de nombreuses fonctions sont déléguées à l'ordinateur, avec seulement une supervision humaine de haut niveau. »³⁸ Ce passage du continu au « discret » (c'est-à-dire aux niveaux d'autonomie caractérisant un système donné) sous-tend un ensemble de critères, largement arbitraires, parfois idéalisés, auxquels la complexité et la variété des situations réelles manquent souvent de se conformer.

13.- Une taxonomie floue. Le robot est-il autonome ? Quand bien même on se placerait dans le cadre limité des taxonomies proposées par l'ISO et l'IEEE, la réponse pourrait être dans certains cas, oui *et* non. Le même rapport du département américain de la défense précise en effet : « De multiples fonctions techniques peuvent être nécessaires à un moment donné, certaines peuvent nécessiter la présence d'un humain dans la boucle alors que d'autres pas. Ainsi, à n'importe quelle étape d'une mission, il est possible qu'un système se trouve simultanément sur plusieurs niveaux d'autonomie. »³⁹ Le robot est plus ou moins « autonome » pour certaines de ses fonctions, suit un mode plus ou moins « automatisé » pour d'autres. Le régime de l'autonomie du robot pourrait donc s'exprimer dans une forme de « logique floue ». Ces diverses tentatives de normalisation illustrent bien à quel point la classification d'un robot en termes de niveaux d'autonomie doit être nuancée. Et puisque la mesure de l'autonomie des systèmes robotiques semble résister à une définition positive, non ambiguë, c'est par l'exemple qu'il

³⁷

<https://www.newyorker.com/science/elements/how-the-artificial-intelligence-program-alphazero-mastered-its-games>

³⁸ Dept of Defense Science Board, Task Force Report: The Role of Autonomy in DoD Systems (2012) <https://fas.org/irp/agency/dod/dsb/autonomy.pdf>

³⁹ Id.

faut peut-être l'aborder. Il est donc utile d'en présenter quelques applications pratiques pour mieux cerner le domaine d'expression de l'autonomie des robots aujourd'hui, quelles conséquences pratiques elles impliquent et comment – dans la mesure du possible – envisager leur encadrement.

III. – Conséquences de l'autonomie du robot

14.- Des véhicules autonomes aux outils d'aide à la décision, de la salle d'opération au tribunal, les « robots » accomplissent aujourd'hui nombre de prouesses imaginées par Asimov dans son œuvre fictionnelle. Ces avancées technologiques, aussi spectaculaires qu'elles soient, ne sont pourtant pas sans dangers (A) et imposent dès à présent de repenser la relation de l'humain face au robot autonome (B).

A. Prodiges et périls du robot autonome

15.- Le véhicule autonome. « Vers 21 h 58, le dimanche 18 mars 2018, un véhicule d'essai d'Uber Technologies, Inc., basé sur un modèle Volvo XC90 2017 modifié et fonctionnant avec un système de conduite autonome en mode automatique, a heurté un piéton sur Mill Avenue, à Tempe, dans le Comté de Maricopa en Arizona. », c'est ainsi que débute le rapport préliminaire du National Transportation Safety Board (NTSB) décrivant l'accident dont fut victime Elaine Herzberg, le premier cas enregistré de décès d'un piéton dû à un véhicule autonome. L'analyse des données enregistrées par le véhicule permet de retracer les étapes du processus de décision qui a conduit à l'accident.⁴⁰ Six secondes avant l'impact, alors que le véhicule se déplaçait à près de 70 km/h, le système de télédétection par laser (LiDAR⁴¹) détecte un obstacle. La

qualité de la détection n'est pas optimale et le classificateur l'interprète successivement comme un « objet inconnu », puis comme un « véhicule » et enfin comme une « bicyclette ». Or à chacune de ces catégories le système associe un comportement particulier (en termes de vitesse supposée et de trajectoire) et propose, en conséquence, une réponse spécifique : alors que les secondes s'égrènent, l'algorithme hésite, la logique « autonome » du système de décision « oscille » entre ces différentes options et peine à décider (à la manière du robot SPD 13 dans la nouvelle d'Asimov « Cercle vicieux », pris au piège entre plusieurs options incompatibles). Il faut ainsi attendre 1,3 seconde avant l'accident pour que le système s'accorde sur une prédiction sûre : l'imminence de l'impact, et décide de la stratégie adéquate : un freinage d'urgence. Or cette opération doit normalement être effectuée par le conducteur du véhicule, car selon Uber, « les manœuvres de freinage d'urgence ne sont pas autorisées lorsque le véhicule est sous contrôle informatique, afin de réduire le risque de comportement erratique du véhicule »⁴². En bonne logique, le système anticipant l'imminence de l'impact aurait donc pu à ce stade alerter le conducteur et lui déléguer son autonomie. Mais, Uber précise, dans de telles configurations, « c'est au conducteur du véhicule d'intervenir et de prendre des mesures adéquates. Le système n'est pas conçu pour alerter l'opérateur. »⁴³ Difficile de comprendre la logique de ce choix. Une telle situation d'urgence – dans laquelle le temps de réaction laissé au conducteur est par trop bref – est précisément celle dans laquelle le système pourrait opter pour une décision autonome. Mais la cause de l'accident n'est pas le fait de ce seul choix, aussi surprenant soit-il. C'est l'ensemble de la chaîne de décision, de la configuration des capteurs (un seul LiDAR

⁴⁰ L'attribution des responsabilités en cas d'accident nécessite de déterminer précisément la séquence d'événements qui l'ont précédé. Lorsque les décisions sont le fait d'un système algorithmique, la Commission européenne prévoit ainsi d'imposer l'inclusion dans les véhicules autonomes d'un enregistreur, une « boîte noire », permettant d'archiver les étapes prises par un système autonome afin d'aider à identifier les causes d'un éventuel accident (Communication of the

European Commission on automated mobility, COM(2018) 283, 17 mai 2018).

⁴¹ LiDar est l'acronyme anglais de « light detection and ranging ».

⁴² National Transportation Safety Board (NTSB) Preliminary Report HWY18MH010.

⁴³ Id.

monté sur le toit du véhicule, au lieu des sept qui équipaient auparavant la flotte de véhicules autonomes mis en service par Uber), à l'interprétation de l'obstacle (l'assignation des signaux en provenance des capteurs à une catégorie donnée : « inconnu », « véhicule », etc.) et à la stratégie associée (éviter l'obstacle, en modifiant la trajectoire du véhicule, en accélérant ou en décélérant) jusqu'au protocole de contrôle du système de freinage, qui est en cause. Ces opérations reflètent en effet une cohorte de choix humains en amont du déploiement du véhicule autonome : les économies liées à la sélection des capteurs, le cadre du modèle d'apprentissage qui permet *in fine* la classification des obstacles, la décision d'interdire un freinage automatique en cas d'urgence et de ne pas alerter le conducteur. L'autonomie du robot est somme toute bien relative si elle n'est que le reflet algorithmique d'un faisceau de contraintes, d'objectifs et de choix techniques, prédéfinis par des individus au service desquels la machine opère. L'autonomie supposée des systèmes ne saurait en ce sens servir de prétexte à décharger de toute responsabilité ceux-là mêmes qui ont établi ces contraintes et fixé ces objectifs.

16.- Le robot et le juge. « L'objectivité et l'intégrité de Multivac, le juge dans cette affaire, étaient telles qu'aucun avocat de la défense ou de procureur n'était requis, seules suffisaient la présence de l'accusé et la présentation des preuves [...] ». ⁴⁴ Isaac Asimov imagine dans « La Vie et les Œuvres de Multivac » déléguer à un superordinateur le rôle de trancher les litiges et de dire le droit. Dans une autre nouvelle, « Les cendres du passé » ⁴⁵ publiée en 1956, Isaac Asimov décrit l'usage du « chronoscope », un outil permettant de visualiser les événements

passés. La même année Philip K. Dick publie « Minority report » récit présentant une autre forme de chronoscopie destinée cette fois à anticiper les événements à venir, plus précisément à prédire les crimes.

Là encore, la science-fiction semble avoir pressenti les développements technologiques les plus récents. La connaissance des données passées pourrait-elle bientôt suffire pour anticiper les risques de délits ? « L'intelligence artificielle » pourrait-elle bientôt assister les juges dans leurs décisions ? Sans aller encore jusqu'à prévoir les crimes, certains systèmes algorithmiques autonomes, sortes de robots policiers, sont dès à présent à l'œuvre. ⁴⁶ En aval de la chaîne pénale, l'algorithme, le robot, font déjà partie de la boîte à outils dont se sert le juge pour estimer, notamment, le risque de récidive et déterminer les peines. Dans une affaire récente, la Cour suprême du Wisconsin a ainsi statué que l'évaluation de la probabilité de récidive par un système algorithmique « autonome » pour aider à la détermination de la peine par un tribunal de première instance ne violait pas le droit de l'accusé à une procédure régulière, et ce, même si la méthodologie utilisée pour produire l'évaluation n'était révélée ni au juge ni au défendeur. ⁴⁷ Eric Loomis, avait été condamné en 2013 après avoir plaidé coupable pour conduite d'une voiture volée et délit de fuite. ⁴⁸ Le juge de première instance avait utilisé, pour déterminer la peine, un outil de profilage permettant d'estimer automatiquement la probabilité de récidive, ⁴⁹ risque estimé particulièrement élevé dans le cas de M. Loomis. Le juge suivit les recommandations de COMPAS et M. Loomis fut condamné en conséquence à 6 ans d'emprisonnement. ⁵⁰ Eric

⁴⁴ « The lifes and times of Multivac », publiée en 1975 dans le New York Times Magazine (et en français dans le recueil « L'homme bicentenaire » édité par Denoël en 1978).

⁴⁵ Publiée en français en 1976 à la Librairie des Champs-Élysées (l'original « The dead past » avait paru en 1956 dans « Astounding Science Fiction »).

⁴⁶ L. Bennett Moses & J. Chan, 2018. Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society*, 28(7), p. 806-822.

⁴⁷ Loomis 881 N.W.2d 749, Wis. 2016, point 754.

⁴⁸ Précisément : « attempting to flee a traffic officer and operating a motor vehicle without the owner's consent. » (Id.)

⁴⁹ Il s'agit du système « COMPAS », pour « Correctional Offender Management Profiling for Alternative Sanctions », logiciel développé par Northpointe Inc.

⁵⁰ Le rapport produit par Northpointe indique que l'algorithme avait attribué à Éric Loomis « un risque

Loomis interjeta appel au fondement que la décision avait été basée, en partie au moins, sur un algorithme, un processus autonome (c'est-à-dire fonctionnant indépendamment et hors du contrôle du juge), dont la décision, pour le moins opaque, n'avait été accompagnée d'aucune explication : seule la probabilité de récidive avait été fournie aux différentes parties. En effet, l'algorithme étant protégé par le secret des affaires, ni le défendeur, ni le juge n'avaient eu les moyens d'examiner la formule utilisée pour examiner la valeur de cette « probabilité ». La Cour suprême du Wisconsin confirma cependant la décision. Le respect de la propriété intellectuelle de Northpointe Inc. requerrait de ne pas dévoiler l'algorithme. Par ailleurs, bien que secret, le calcul du score obtenu par le procédé automatique n'est qu'un des nombreux éléments participant à la décision rendue par le juge, qui reste seul responsable du verdict et de la peine. Une décision qui peut laisser perplexe cependant, car comment justifier la peine si celle-ci repose (même partiellement) sur un score algorithmique obtenu de manière autonome par un algorithme dont on ne peut vérifier la logique ? De fait, la notion « d'explication » au cœur de nombreux domaines juridiques—les droits fondamentaux l'invoquent ainsi au profit de la protection de la vie privée⁵¹ ou du droit à un procès équitable⁵²—impose précisément l'accès à l'enchaînement causal à

élevé de violence et un risque élevé de récidive ». Le juge prit bien en compte ces recommandations et déclara au défendeur durant le procès : « vous êtes identifié, selon l'évaluation Compas, comme un individu présentant un risque élevé pour la collectivité » (Adam Liptak, New-York Times, 1/5/2017).

⁵¹ Les articles 13(2)(f) et 14(2)(g) du Règlement UE 2016/679, accorde à toute personne sujette à une prise de décision automatisée sur la base de ses données personnelles le droit d'obtenir « des informations utiles concernant la logique sous-jacente [au traitement automatisé] », ces informations devant être communiquées « d'une façon concise, transparente, compréhensible et aisément accessible, en des termes clairs et simples » (Art. 12 (1)). Ce même principe est inscrit à l'article 4 la loi pour une République numérique.

⁵² Règle essentielle du procès civil, toute décision de justice, tout jugement, doit être motivé « en fait et en

l'origine de ces décisions. L'immixtion d'un intermédiaire algorithmique autonome et opaque⁵³ brise cette chaîne de déduction et interdit de répondre à cet impératif de transparence de la décision. Le cas d'Eric Loomis n'est pas isolé, alors que les outils prédictifs se popularisent dans le domaine judiciaire, le nombre d'erreurs croît. En 2016, Glenn Rodríguez, un détenu de l'Eastern Correctional Facility, dans l'État de New York, s'est ainsi vu refuser par erreur une libération conditionnelle parce qu'une erreur typographique s'était glissée dans le fichier informatique fourni à l'algorithme COMPAS pour calculer son risque de récidive.⁵⁴

17.- Le robot-combattant. L'opacité de certains systèmes autonomes justifie certainement de remettre en question leur application indifférenciée. Si l'usage du « robot autonome » dans les domaines judiciaires ou pénaux doit être soumis à examen, c'est aussi le cas des applications militaires. Dans ce registre, la référence aux « armes autonomes » a déjà, en particulier, soulevé de nombreux débats. Bien que l'on puisse aisément imaginer des applications pour lesquelles le déploiement de systèmes autonomes sur le terrain militaire contribue à sauver des vies (pensons au déminage, aux secours d'urgence ou à l'évacuation des blessés), leurs applications offensives forcent l'interrogation. Face à ce constat, de nombreuses personnalités,

droit » (Art. 455 du code de procédure civile). Cette condition impose d'une part au juge d'exposer et de justifier son raisonnement juridique, liant les faits d'espèce à la règle de droit applicable et permet, d'autre part, pour justiciable, le maintien d'un principe d'équité. Le droit fondamental à un procès équitable est inscrit à l'article 6 CEDH. L'article 45 précise en outre que « les arrêts, ainsi que les décisions déclarant des requêtes recevables ou irrecevables, sont motivés. »

⁵³ Un algorithme opaque à plus d'un sens dans ce cas : de par sa construction algorithmique (on sait aujourd'hui à quel point certains modèles d'inférence, tels que les réseaux de neurones profonds, se prêtent mal aux exigences d'« explicabilité ») et de par sa protection par le secret d'affaires.

⁵⁴

<https://www.nytimes.com/2017/06/13/opinion/how-computers-are-harming-criminal-justice.html>

scientifiques et militants, ont tenté d'inciter les gouvernements et les institutions internationales à envisager une interdiction préventive des armes autonomes, capables de sélectionner et de prendre à parti des cibles sans intervention humaine. Une situation d'urgence, car ces « robots tueurs » ne sont plus désormais – là encore – du seul ressort de la science-fiction. Une lettre ouverte initiée par des dizaines d'experts en intelligence artificielle le souligne : « Le développement de l'intelligence artificielle est tel que le déploiement de tels systèmes ne prendra pas des décennies, il est - pratiquement sinon légalement - réalisable en quelques années. Les enjeux sont des plus élevés : les armes autonomes ont été décrites comme « une troisième révolution dans l'art de la guerre », après la poudre à canon et les armes nucléaires. »⁵⁵ Ce même sentiment de défiance à l'encontre des armes autonomes a été affiché par Antonio Gutierrez, secrétaire général des Nations Unies, qui déclarait en mars 2019 que « les machines autonomes qui ont la capacité de décision d'attenter à la vie sans intervention humaine sont politiquement inacceptables, moralement répugnantes et devraient être interdites par le droit international. »⁵⁶ L'autonomie du robot est donc dans ce cas, ex ante, sujet de vie ou de mort et la question (ouverte, dont il faudra débattre) se pose aujourd'hui : le robot autonome doit-il être banni du terrain de la guerre ?

18.- Le robot-médecin. Autre domaine de prédilection des applications robotiques, autre

domaine dans lequel les questions de vie ou de mort sont susceptibles de se poser, le champ de la santé est l'un des premiers à bénéficier des avancées de l'automatisation. Tous les robots médicaux ne sont pourtant pas autonomes. Par exemple, le robot « Da Vinci » entré en service dès le début des années 2000 est essentiellement une extension du geste du chirurgien.⁵⁷ À l'autre extrémité du spectre, dans le domaine de l'analyse automatique des données épidémiologiques ou de la radiologie, les résultats de l'algorithmique, et les succès récents dus à l'apprentissage automatique ont catapulté les systèmes d'aide à la décision au-devant de la scène médicale. Les réseaux neuronaux identifient certaines maladies avec des taux d'erreurs plus faibles que les meilleurs praticiens,⁵⁸ d'autres systèmes anticipent automatiquement les pathologies avant que les médecins n'en soient alertés.⁵⁹ Bien que seuls les humains soient autorisés à pratiquer la médecine,⁶⁰ les algorithmes se voient dès à présent déléguer un nombre croissant d'opérations qui participent à l'élaboration de diagnostics cliniques.

Pourtant, en dépit de ces promesses, de nombreux obstacles continuent de paver la voie des applications médicales de l'IA. Le système Watson Health d'IBM⁶¹ en a récemment donné un exemple. Watson, la plateforme de traitement automatique du langage naturel d'IBM, est devenue en 2011 un des porte-flambeaux du renouveau de l'IA après avoir battu les candidats (humains) au jeu « Jeopardy ! »⁶². Elle a été dérivée selon

⁵⁵ *Autonomous weapons: an open letter from AI & robotics researcher* (<https://futureoflife.org/open-letter-autonomous-weapons>). Une initiative similaire a été lancée en 2012 par une coalition de 93 organisations non gouvernementales et a déjà obtenu le soutien de 28 états : <https://www.stopkillerrobots.org>

⁵⁶ <https://news.un.org/en/story/2019/03/1035381>

⁵⁷ Une fonction du robot est de minimiser les risques d'erreur humaine, en guidant la trajectoire, en atténuant par les vibrations ou en évitant le contact avec certains organes.

⁵⁸ Par exemple, dans le domaine de l'oncologie : A. Esteva et al. *Dermatologist-level classification of skin cancer with deep neural networks*. *Nature* 542, n° 7639, 2017, p. 115. Dans le domaine cardiologique : A.Y. Hannun et al., 2019. *Cardiologist-level arrhythmia*

detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, vol. 25, n° 1, p.65.

⁵⁹ Dans le cas de la maladie d'Alzheimer : G. Lee et al., 2019. *Predicting Alzheimer's disease progression using multi-modal deep learning approach*. *Scientific reports*, vol. 9, n° 1, p. 1952. Pour les risques cardiovasculaires : R. Poplin et al., 2018. *Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning*. *Nature Biomedical Engineering*, vol. 2, n° 3, p. 158.

⁶⁰ Art. L. 4131-1, Code de la Santé Publique.

⁶¹ <https://www.ibm.com/watson/health>

⁶² Le jeu consiste à trouver le plus rapidement possible la question correspondant à une « réponse » présentée aux candidats.

plusieurs domaines d'application et sa version spécialisée « Watson for Oncology » est aujourd'hui utilisée par des centaines d'hôpitaux à travers le monde pour recommander des traitements aux patients atteints de cancer. Or l'efficacité de tels systèmes reposant sur un apprentissage automatique est déterminée par l'accès à une quantité suffisante des données d'entraînement de qualité suffisante.⁶³ Lorsque ces valeurs expérimentales sont trop rares pour permettre d'entraîner un modèle, il est possible de faire appel à une technique dite « d'augmentation des données » dans laquelle de nouveaux cas – non plus réels, mais synthétiques cette fois – sont générés automatiquement. Confronté au manque de données cliniques, « Watson for Oncology » dut recourir à une telle approche. Le modèle final, destiné à établir un pronostic dans les cas de cancers et à proposer des traitements adaptés, fut donc dérivé d'un corpus d'entraînement dominé par des données simulées. Résultat : nombre des recommandations formulées par Watson se sont avérées erronées, voire dangereuses, comme la suggestion de prescrire du Bevacizumab à des patients présentant des saignements sévères, une contre-indication explicite de ce médicament.⁶⁴

B. L'humain face au robot

19.- Robots « autonomes », mais humains « aux commandes ». L'autonomie, au sens technique tout du moins, des robots « nouvelle

génération » n'est plus affaire de littérature. Ces quelques exemples viennent illustrer à la fois combien les systèmes autonomes (mécaniques ou algorithmiques) se sont déjà imposés dans notre quotidien. Certes, les effets d'annonces – médiatiques ou publicitaires – doivent être modulés : les véhicules totalement autonomes (au sens technique, toujours) sont encore en phase de développement, RoboCop ne patrouille pas encore les rues, Multivac ne préside pas seul au déroulement des procès. Mais la tendance ne peut être niée. Niveau après niveau des nomenclatures établies, la machine se voit assigner des fonctions décisionnelles toujours plus nombreuses. Ce courant transverse, qui affecte tous les domaines de la société, n'est pas dû aux seules innovations technologiques.⁶⁵ Il est aussi porté par une volonté délibérée de déléguer à des plateformes techniques, des machines, des tâches qui nécessitaient jusqu'alors le recours à l'humain, que l'on sait faillible : la machine permettrait de gagner du temps, de gagner en productivité, d'éviter l'erreur de jugement ou l'accident.

Or ce transfert de ces prérogatives de l'humain à la machine n'est pas sans coût. On l'a vu, d'Eric Loomis à Elaine Herzberg, les systèmes autonomes montrent leurs limites. Comme dans les nouvelles d'Asimov, tout rationnels qu'ils soient, les robots se trompent, butent, tournent en rond. La question se pose alors de l'encadrement de ces outils.

Si certaines décisions échappent au registre de compétences de la machine, c'est à l'humain de

⁶³ Ce problème est d'autant plus aigu dans le domaine de la santé, pour des raisons à la fois pratiques (les données sont le plus souvent collectées par les cliniciens dans un cadre hospitalier et ne sont que rarement agrégées dans des bases de données partagées, encadrées par des principes d'interopérabilité) ou encore légales (les données médicales sont notamment classées par le RGPD (Art. 9) dans la catégorie des données personnelles « sensibles »). Elles sont soumises à un régime protecteur qui requiert un nombre de précautions supplémentaires par rapport aux simples données personnelles).

⁶⁴ C. Ross & I. Swetlitz, *IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments*,

internal documents show. In Stat News <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/> (2018). Topol, E. J., 2019. *High-performance medicine: the convergence of human and artificial intelligence*. Nature medicine, vol. 25, n° 1, p. 44-56.

⁶⁵ La démultiplication des capacités de calcul des ordinateurs (illustrée par la « loi de Moore ») ; l'omniprésence des capteurs, de la numérisation, des réseaux, des moyens de stockage a permis l'agrégation de vastes corpus de données ; l'apprentissage automatique a capitalisé sur ces avancées pour reproduire (imiter) certains processus cognitifs, qu'il s'agisse de perception (p.ex. la vision ou l'audition) ou de décision (p.ex. classifier des données).

servir de garde-fou. Face à l'autonomie du robot, l'humain doit conserver la main. Cette nécessité est soulignée par des textes récents. Le règlement 2016/679 (RGPD) l'évoque lorsqu'il accorde la possibilité aux personnes de ne pas être l'objet d'une décision entièrement automatisée si cette dernière a un effet juridique ou un impact significatif (Art. 22 RGPD). On la retrouve également dans le cadre du projet de loi relatif à la bioéthique n°2187. Dans les documents préparatoires, le Conseil d'État préconise en effet de consacrer explicitement l'interdiction d'un diagnostic établi uniquement par un système d'intelligence artificielle autonome, sans intervention d'un médecin.⁶⁶ L'article 11 du projet de loi visant « à sécuriser la bonne information du patient lorsqu'un traitement algorithmique de données massives [“intelligence artificielle”] est utilisé à l'occasion d'un acte de soins » garantit une intervention humaine et oblige d'informer le patient de l'utilisation d'un traitement algorithmique.⁶⁷ Plus récemment encore, un rapport du Comité économique et social européen souligne la nécessité de conserver le contrôle sur les systèmes autonomes.⁶⁸ Catelijne Muller, membre de la commission en charge du rapport, rappelle ainsi : « en matière d'IA, nous avons besoin d'une approche où l'homme reste aux commandes (« human-in-command approach ») et où les machines restent des machines que les hommes ne cessent jamais de contrôler ».⁶⁹

⁶⁶ Une référence à l'article 10 de la loi du 6 janvier 1978 « informatique et liberté », qui prévoit qu'aucune décision produisant des effets juridiques à l'égard d'une personne ne peut être prise sur le seul fondement d'un traitement automatisé des données.

⁶⁷ « Lorsque, pour des actes à visée préventive, diagnostique ou thérapeutique, est utilisé un traitement algorithmique de données massives, le professionnel de santé qui communique les résultats de ces actes informe la personne de cette utilisation, et des modalités d'action de ce traitement. » (Art. 11).

⁶⁸ Building Trust in Human-Centric Artificial Intelligence. Comité économique et social européen, INT/887-EESC-2019, 25 oct. 2019.

⁶⁹ <https://www.eesc.europa.eu/fr/news-media/press-releases/intelligence-artificielle-leurope->

20.- Du robot au cobot, entre contrôle et collaboration. Si le développement des systèmes autonomes est voué à se poursuivre, comment un médecin pourra-t-il contrôler (accepter ou remettre en question) le diagnostic proposé par la machine et l'expliquer au patient ? Comment le juge pourra-t-il interpréter les probabilités issues de l'algorithme pour informer sa décision ? Comment le conducteur d'un véhicule, l'utilisateur d'un robot, pourra-t-il décider en connaissance de cause face à l'autonomie de la machine ? En somme, comment « rester aux commandes » ?

Pour éviter que ces annonces ne restent des vœux pieux, il faudra s'interroger sur les modalités pratiques de mise en œuvre d'un contrôle effectif des systèmes autonomes. Le rapport du CESE souligne en ce sens que le développement d'une intelligence artificielle de confiance « suppose le contrôle de l'humain sur la machine et l'information des citoyens quant à ses usages. Les systèmes d'IA doivent être explicables ou, lorsque cela n'est pas possible, des informations doivent être fournies aux citoyens et aux consommateurs sur leurs limites et leurs risques. »⁷⁰

Le Rapport de la commission allemande sur « l'éthique des données » propose de catégoriser les systèmes algorithmiques selon une échelle de risques, requérant pour les niveaux intermédiaires une transparence accrue et allant jusqu'à l'interdiction totale.⁷¹

doit-opter-pour-une-approche-ou-l'homme-reste-aux-commandes-affirme-le-cese

⁷⁰ Supra, not. n° 68, pt. 1.7 ; Journal Officiel de l'UE (JO UE) C 288, 31 août 2017, p. 1 ; JO UE C 440, 6 déc. 2018, p. 1.

⁷¹ Gutachten der Datenethikkommission der Bundesregierung, oct. 2019. https://datenethikkommission.de/wp-content/uploads/191015_DEK_Gutachten_screen.pdf. Dans le contexte de la littérature de science-fiction, Frank Herbert évoque dans *Dune* l'épisode du « Jihad Butlérien » (la « Grande révolte ») qui, décrétant « tu ne créeras point de machine à l'image de l'esprit humain », conduit à l'interdiction des ordinateurs et de toute forme d'intelligence artificielle (Frank Herbert, *Dune*, Robert Laffont, 1977).

Alors que ces systèmes autonomes occupent une place croissante dans de nombreux domaines, que les robots industriels font place aux « cobots »⁷², il faut d'ores et déjà se préparer à faire face à ces nouvelles interfaces « homme-machine ». Devant l'urgence de ce constat, de nouvelles formations sont envisagées qui permettront aux utilisateurs de mieux comprendre les processus algorithmiques dont dépendent leurs pratiques. Pour Pierre-Antoine Gourraud, praticien hospitalier à l'université de Nantes, « il devient urgent de former nos jeunes, pour ne pas laisser partir une génération de médecins sans armes pour comprendre, expliquer et démystifier l'utilisation de l'IA en santé. »⁷³ Une « École de l'intelligence artificielle en santé » (EIAS), rattachée au centre hospitalier de l'Université de Montréal (CHUM) a ainsi été créée en novembre 2018 pour accompagner les équipes médicales. Fabrice Brunet, directeur du CHUM souligne l'importance d'un tel accompagnement pour le patient comme pour le corps médical : « si on veut utiliser tous les bienfaits de l'IA, il faut que les personnes se l'approprient et qu'ils ne la voient plus comme une menace, mais comme une manière contrôlée de mieux soigner. »⁷⁴ Le même constat pourrait certainement s'appliquer aux praticiens du droit pour lesquels l'intermédiation algorithmique est destinée à se populariser.

L'autonomie accrue des systèmes de décisions ne saurait se justifier si elle se fait hors du contrôle—ou pire, aux dépens—de ses utilisateurs. La formation des cliniciens, des juges, des ingénieurs pourrait permettre une juste délimitation de leur périmètre d'application. Préparer les usagers de tels outils « autonomes » à un maniement raisonné, intégrer les bénéfices de tels systèmes tout en en comprenant les écueils est un enjeu de

société majeur auquel le système éducatif doit dès à présent se confronter.

Conclusion

21.- Isaac Asimov nous invite au fil de ses récits, à explorer – et à remettre en question – le caractère absolu des lois de la robotique. Face aux impondérables, confrontées à une réalité souvent rétive, les lois achoppent. Plus qu'une solution, un cadre normatif de résolution des conflits, le jeu de règles imposées au robot est d'abord un outil narratif, le moyen de nouer l'intrigue autour d'un paradoxe dont il est la source. Les « Trois Lois » servent alors à révéler les limites de la rationalité « hors sol » du robot : s'il est bien doté d'une logique sans faille, il faut en convenir, le cerveau cartésien de la machine manque bien souvent de « sens commun », voire de « bon sens ».

Mais il ne s'agit plus seulement de littérature : la réalité a rattrapé la fiction. Dépasant l'artifice dans lequel elle était cantonnée, la réflexion sur l'encadrement du robot initiée par Asimov a pris récemment une dimension pratique. Car le robot est aujourd'hui objet du quotidien. Il interagit avec un environnement physique, social, économique, il se confronte chaque jour davantage à l'humain. Les spéculations sur les conditions de régulation des robots ne sont donc plus d'ordre académique ou intellectuel, elles ont investi la sphère publique, animent le forum politique, provoquent les débats, éthiques ou juridiques.

22.- Dans ce contexte, souvent passionné, le robot ne laisse pas indifférent. Il est le support de craintes existentielles (« et s'il nous remplaçait ? ») autant que d'attentes messianiques (« pourvu qu'il nous remplace... »). On se plaît à l'imaginer doté d'une conscience, d'intention, de volonté.⁷⁵ On

⁷² V. A.M. Djuric, R.J. Urbanic & J.L. Rickli, 2016. *A framework for collaborative robot (CoBot) integration in advanced manufacturing systems*. SAE International Journal of Materials and Manufacturing, vol. 9, n° 2, p. 457-464 ; M.M. Veloso et al. 2015, *CoBots: Robust Symbiotic Autonomous Mobile Service Robots*. In IJCAI, p. 4423.

⁷³ Le Monde, 12 oct. 2019.

⁷⁴ Id.

⁷⁵ En ce sens, Kate Darling souligne : « La projection de qualités d'apparence humaines commence par une tendance générale à sur-attribuer l'autonomie et l'intelligence à la manière dont les choses se comportent, même si elles ne suivent qu'un algorithme

lui prête autant de travers (le robot, l'IA, est tout à tour « criminel », « misogyne », « raciste ») que de qualités (il est ici « à l'écoute », là « imaginaire », là encore « artiste »). On en viendrait presque à vouloir lui attribuer une « personnalité ».⁷⁶

Il serait malvenu de s'en surprendre : le robot, tout autant objet culturel que création technique, se prête par nature au glissement anthropomorphique. N'est-il pas, au fond, une entité hybride, une « chimère » au croisement de deux mondes : une chose – mécanique et algorithmique – certes, mais une chose dédiée à remplir des fonctions bien humaines ; un substitut. Que l'on puisse y voir (y « projeter ») certaines de nos facultés cognitives, rien là donc que de très normal : le robot est façonné au miroir de nos compétences. S'il évoque l'humain, c'est qu'il en est un simulacre. (Alan Turing, en posant le problème de l'intelligence artificielle sous l'angle d'un « jeu de l'imitation »⁷⁷ ne s'y était pas trompé.) Mais aux faux-semblants de ce « miroir-automate »⁷⁸ il serait malvenu de s'enfermer. Car attribuer en propre à la machine un élément d'intentionnalité, ce serait éluder la part humaine qui s'y cache et l'anime ; ce serait oublier combien il y a, finalement, d'humain dans l'automate. Le robot n'est en effet ici qu'un outil, un levier, une extension de l'intention de ses concepteurs. Pour paraphraser Bergson, le robot c'est d'abord « de la mécanique plaquée sur du vivant »⁷⁹.

23.- Le robot fait illusion. S'il fait figure « d'entité autonome », il ne faudrait donc pas tomber dans le piège, ni de la sémantique ni des premières impressions. Bien sûr, considérer les robots comme « autonomes »

semble tomber sous le sens : n'est-ce pas là après tout leur « raison d'être » que de prendre en charge seuls un ensemble d'opérations répétitives, fastidieuses ou complexes ?⁸⁰ En agissant sans contrôle direct, l'automate gagne son indépendance. Mais à quelle autonomie fait-on référence ? Pas à celle du philosophe ni à celle du juriste. C'est seulement (modestement) d'une indépendance fonctionnelle, technique, qu'il s'agit.

Et si l'on creuse plus profond dans le maillage algorithmique, la nature de cette autonomie s'érode davantage. Car le robot n'exprime son ersatz d'indépendance que dans un périmètre strict. Derrière l'autonomie affichée de la machine c'est en effet un réseau de décisions, d'intentions et au fond, de responsabilités, bien humaines qui forment et orientent les actions du robot. Le choix d'utiliser un capteur plutôt qu'un autre, la sélection de l'architecture d'apprentissage, l'agrégation d'une base d'exemples pour entraîner un modèle d'inférence détermineront, tous ensemble, les actions que la machine sera susceptible de produire. Le robot n'est libre d'atteindre les objectifs qui lui ont été fixés, que dans ce seul espace de représentations ; espace circonscrit, restreint, et somme toute arbitraire.

Les auteurs d'un rapport du département de la défense américain ne s'y trompent d'ailleurs pas : « tous les systèmes autonomes sont, à un certain niveau, supervisés par des opérateurs humains. Leur programmation détermine les limites à l'autonomie du robot au sens des actions et des décisions qui lui sont déléguées. »⁸¹

24.- Pourtant, nous l'avons montré, de la santé à la justice, des transports à la guerre, on

simple. » (K. Darling, *Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects*, in: *Robot Law*, Ed. Ryan Calo, A. Michael Froomkin and Ian Kerr, Edward Elgar, 2016, p. 213-231).

⁷⁶ V. supra, not. n° 22.

⁷⁷ A.M. Turing, *Computing Machinery and Intelligence*, *Mind* 59, 1950, p. 433-460.

⁷⁸ En référence à l'ouvrage de Gérard Chazal, *Le miroir automate – introduction à une philosophie de l'informatique*, Champ Vallon, 1995. (V. en particulier, p. 143-165).

⁷⁹ Bergson, dans sa formule classique, fait référence au rire : « Le rire, c'est du mécanique plaqué sur du vivant » (Henri Bergson, *Le rire*, Flammarion, p. 29).

⁸⁰ Le terme « robot » qui découle de « robota », tiré de la pièce de Karel Čapek « *Rossumovi univerzální roboti* » (1920), signifie en tchèque « travail forcé », « corvée ».

⁸¹ Dept. of Defense Science Board, Task Force Report: *The Role of Autonomy in DoD Systems* (2012) <https://fas.org/irp/agency/dod/dsb/autonomy.pdf>

accorde aujourd'hui à la machine un régime d'indépendance inédit. Les conséquences d'une telle délégation d'autonomie ne sauraient être ignorées. Elles peuvent être désastreuses ; elles sont parfois tragiques. Elles engagent à la plus grande prudence. Norbert Wiener, un des pères de la cybernétique, le notait déjà en 1960 : « nous ferions mieux d'être tout à fait sûrs que l'objectif que l'on met dans la machine est celui que nous désirons vraiment »⁸² (certitude dont on sait qu'elle restera bien souvent vœu pieux). Les récits d'Isaac Asimov du Cycle des robots illustrent à merveille la mise en garde du scientifique.

Alors, puisque codifier les lois de la robotique en langage machine ne suffit pas, pour profiter pleinement des avantages offerts par ces nouveaux outils, il faudra donc laisser à l'humain, donneur d'ordre autant qu'utilisateur, un droit de regard sur l'autonomie du robot. Il faudra, en fin de compte, et c'est peut-être là le paradoxe, en contrepoint de l'automatisation, s'efforcer de laisser « l'humain aux commandes ».

J.-M. D.

⁸² Norbert Wiener, *Some Moral and Technical Consequences of Automation*. Science, vol. 131, n° 3410, 1960, p. 1355-1358

Le robot, gardien de l'humanité. La Loi zéro dans les cycles des Robots et de Fondation d'Asimov.

Anthony VALLAT

Enseignant en sociologie et en philosophie
École Supérieure de la santé de Lausanne

Introduction

1.- L'importance d'Isaac Asimov. Je remercie les organisateurs de ce colloque sur les lois de la robotique d'Asimov et le Droit de m'avoir invité à présenter cette contribution. En effet, j'ai une histoire singulière avec Isaac Asimov. Encore lycéen, je me destinais aux études d'ingénieur en physique à l'EPFL, jusqu'à ce que je lise *Fondation*¹. Sous le choc d'une sorte de révélation, je me suis alors orienté vers des études de sociologie, espérant en apprendre davantage sur la psychohistoire. Hélas, à l'issue de ma licence, je n'ai pas été recruté par la Seconde Fondation. J'ai tenté de rembourser autrement ma dette à Asimov en publiant un essai en 2014 aux éditions ActuSF, *Utopie et Raison dans le Cycle Fondation d'Isaac Asimov*.

Il y a quelque chose de vraiment remarquable dans l'œuvre d'Asimov. C'est le point de départ de mes travaux sur Asimov. Si *Fondation* a été lu par des millions de lecteurs, c'est que cette œuvre a touché à quelque chose d'essentiel. Mon hypothèse est qu'à travers son œuvre, Asimov propose une pensée forte sur l'utopie, et que c'est une explication de son succès.

Cette communication va porter sur la question du robot gardien de l'humanité et la Loi zéro imaginée par Asimov. La voici : « Un robot ne doit causer aucun mal à l'humanité ou, faute d'intervenir, de permettre que l'humanité souffre d'un mal.² » Elle apparaît en 1985 dans

le roman *Les Robots et l'Empire*, pour compléter, ou dépasser, les trois Lois de la robotique. Pour rappel : « 1. Un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ; 2. 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 ; 3.- Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi. »

Je vais montrer comment la Loi zéro s'inscrit dans la dynamique de l'œuvre d'Asimov, et quels enjeux philosophiques sont interrogés (les robots, la raison, le langage, le dialogue, la prise de décision, l'humanité, l'utopie, ..).

I. Le Cycle de Fondation, une réflexion sur l'utopie

2.- Corpus sélectionné. Tout d'abord un mot sur la composition du Cycle de Fondation.

Notre lecture analytique sera basée sur le corpus suivant :

Fondation (*Foundation*, Gnome Press, 1951), volume qui inclut les nouvelles *Les Psychohistoriens* (1951), *Les Encyclopédistes* (1942), *Les Maires* (1942), *Les Marchands* (1944) et *Les Princes Marchands* (1944) ;

Fondation et Empire (*Foundation and Empire*, Gnome Press, 1952), volume qui inclut les nouvelles *Le Général* (1945) et *Le Mulet* (1945) ;

¹ *Foundation*, Gnome Press, 1951

² *Les Robots et l'Empire*, vol. 2, p.145, *Robots and Empire*, Doubleday, 1985

Seconde Fondation (*Second Foundation*, Gnome Press, 1953), volume qui inclut les nouvelles *La Quête du Mulet* (1948) et *La Quête de la Fondation* (1949) ;

le roman *Fondation Foudroyée* (*Foundation's Edge*, Doubleday, 1982) ;

le roman *Les Robots de l'Aube* (*Robots of Dawn*, Doubleday, 1983) ;

le roman *Les Robots et l'Empire* (*Robots and Empire*, Doubleday, 1985) ;

le roman *Terre et Fondation* (*Foundation and Earth*, Doubleday, 1986) ;

le roman *Prélude à Fondation* (*Prelude to Foundation*, Doubleday, 1988) ;

le roman *L'Aube de Fondation* (*Forward the Foundation*, Doubleday, 1992) ;

La présence de *Les Robots de l'Aube* et *Les Robots et l'Empire* peut surprendre, mais je montrerai en quoi leur apport au Cycle est décisif.

Pour toutes les citations, je me référerai aux deux volumes parus chez Denoël en 2006. Le premier, *Fondation*, regroupe les volumes *Fondation*, *Fondation et Empire* et *Seconde Fondation*. Le second, *Fondation Foudroyée*, regroupe *Fondation Foudroyée* et *Terre et Fondation*. Pour *Les Robots de l'Aube*, je me référerai aux deux volumes publiés en 1984 par J'ai Lu et, pour *Les Robots et l'Empire*, les deux volumes parus en 1986 chez le même éditeur.

3.- L'utopie comme projet littéraire d'Asimov.
Commençons par une citation remarquable.

« Le contrôle de soi et de la société a été abandonné au hasard, ou aux vagues tâtonnements de systèmes éthiques essentiellement basés sur l'intuition, l'inspiration et l'émotion. Il en résulte qu'aucune culture dont le coefficient de stabilité excède environ 55 % n'a jamais vu le jour, avec pour corollaire une

affreuse détresse humaine.³» (Isaac Asimov, *Seconde Fondation*).

Très tôt, Asimov a conscience des malheurs de l'humanité⁴. C'est ainsi qu'à travers le Cycle de Fondation, il va proposer une série d'utopies possibles, ou plutôt une série de paradigmes rendant possible l'utopie.

À titre d'exemple, décrivons ici le tout premier paradigme, présent dans le premier volume de Fondation. Tout d'abord, il faut qu'un chercheur particulièrement brillant achève l'élaboration d'une sociologie définitive, la fameuse psychohistoire⁵. Cette science a pour vocation de décrire et de prédire le plus précisément possible l'évolution d'une société, en calculant les probabilités qu'ont les différents avènements de se réaliser. Ensuite cette science pourrait ainsi prédire comment orienter l'histoire de l'humanité pour l'amener à l'utopie. Et finalement, il faudrait des hommes raisonnables et de bonne volonté pour utiliser cette science pour amener l'humanité à l'utopie. Ces trois éléments sont le point de départ de *Fondation*, un premier paradigme d'une utopie possible.

Bien entendu, ce n'est pas si simple, et Asimov va mettre en scène une série de coups de théâtre pour montrer que ce paradigme n'est pas suffisant. Dans les nouvelles et romans suivants, il va enrichir le premier paradigme, et proposer de nouveaux paradigmes décrivant les conditions de possibilités de l'utopie.

Je ne vais pas ici détailler l'ensemble des paradigmes qui se succèdent dans le Cycle de Fondation. Cela excéderait l'ambition de cette communication⁶. Pour cet exposé sur la Loi zéro, je vais débiter par l'examen de *Fondation Foudroyée*⁷.

³ I. Asimov, *Fondation*, coll. Lunes d'encre, Denoel, 2006, p. 773.

⁴ Notons que les premières nouvelles qui composent le Cycle de Fondation ont été rédigées pendant la Deuxième Guerre Mondiale.

⁵ Asimov reconnaît d'ailleurs dans l'article Psychohistoire (in recueil *Mais le Docteur est d'Or*, Presses Pocket, 1996, Gold, 1995), qu'il aurait été plus

approprié de nommer cette science la « psychosociologie », mais qu'à l'époque de ses vingt-deux ans, il était fasciné par les livres d'histoire.

⁶ On trouvera dans mon essai *Utopie et Raison dans le Cycle Fondation d'Isaac Asimov*, éditions ActuaSF, 2014 une description analytique et critique des paradigmes liés à l'utopie dans le Cycle de Fondation.

⁷ *Foundation's Edge*, Doubleday, 1982.

4.- Fondation foudroyée. *Fondation Foudroyée* est le quatrième volume du Cycle, roman écrit après une pause de 29 ans depuis la parution de la trilogie initiale, et paru en 1982.

À la fin de *Fondation Foudroyée*, une utopie possible est proposée, l'utopie Gaia. Gaia est une entité planétaire, une planète où les humains sont dans une forme de communion. Ils communiquent entre eux par télépathie, et savent ce que les autres veulent ou pensent. S'installe ainsi une forme d'utopie, que l'on pourrait qualifier sommairement de démocratie directe absolue.

À la fin du roman, Gaia, par le truchement d'une entité nommée Joie, raconte aux personnages principaux qu'elle est structurée et contrainte par les trois Lois de la robotique. Nous nous retrouvons donc non pas avec un robot soumis aux trois Lois, mais toute une société planétaire soumise aux trois Lois.

Gaia explique ensuite qu'elle est devant un choix décisif. Soit rester une utopie locale, c'est-à-dire une utopie localisée sur une seule planète. Soit s'étendre dans toute la Galaxie et englober ainsi tous les humains de la Galaxie dans une communication télépathique les uns avec les autres, projet qu'elle nomme Galaxia. Gaia doit opter pour l'une ou l'autre alternative.

Gaia va demander à un personnage humain, Golan Trevize⁸, qui a été sélectionné au début du roman par Gaia, de choisir à sa place.

En effet, Gaia explique qu'elle ne peut pas procéder elle-même à cette décision, parce qu'elle est bloquée par les trois Lois. Voilà où on en est à la fin du roman. Et coup de théâtre final, on découvre que Gaia est contrôlée par des robots !

C'est inattendu ! Jusqu'alors il n'y avait jamais eu de robots dans le Cycle de Fondation, aucun robot depuis les premiers textes de 1941. Le cycle des robots et celui de Fondation étaient jusqu'alors totalement disjoints.

Nous apprenons donc que c'est à cause des trois Lois que les robots ne peuvent pas décider de l'avenir de l'humanité. Mais dans cette fin de roman, il n'y a pas de justifications ou d'explication du pourquoi les trois Lois empêchent la prise de décision par Gaia.

Asimov n'est visiblement pas satisfait de cette fin un peu abrupte. Il se propose de développer les différentes questions en suspens. Pour cela, il va rédiger deux nouveaux romans, *Les Robots de l'Aube*⁹ et *Les Robots et l'Empire*¹⁰.

Ces deux romans ont été publiés entre *Fondation Foudroyée* et *Terre et Fondation*¹¹. Pour l'anecdote, Asimov précise dans sa biographie¹² qu'il a préféré écrire *Les Robots de l'Aube* et *Les Robots et l'Empire* que *Fondation Foudroyée* ou *Terre et Fondation*.

5.- Les Robots de l'Aube. Posons le cadre des *Robots de l'Aube*¹³.

L'action se situe plusieurs milliers d'années avant la Chute de l'Empire, quelques années après *Les Cavernes d'acier*¹⁴ et *Face aux feux du soleil*¹⁵. Il s'agit de la troisième enquête de l'inspecteur terrien Elijah Baley, accompagné pour la troisième fois par le robot Daneel. Le cadre est le début de l'expansion spatiale humaine. On trouve d'un côté une Terre surpeuplée, des milliards d'habitants, et de l'autre cinquante planètes colonisées et terraformées. Y habitent les Spatiaux, peu nombreux, qui vivent jusqu'à quatre cents ans. Ils sont assistés par une multitude de robots soumis aux trois Lois. Les Terriens n'ont ni la possibilité, ni le droit d'émigrer sur ces

⁸ Golan Trévize est présenté comme un individu qui pratique une très fine épistémologie, basée sur un exercice rigoureux de la Raison.

⁹ *Robots of Dawn*, Doubleday, 1983.

¹⁰ *Robots and Empire*, Doubleday, 1985.

¹¹ *Foundation and Earth*, Doubleday, 1986.

¹² *Moi, Asimov*, Denoël, 1996, I, Asimov, Doubleday, New York, 1994.

¹³ Pour toutes les citations, nous nous sommes référés aux éditions suivantes : *Les Robots de l'Aube*, paru en 1984 chez J'ai Lu en deux volumes (n° 1602 et n° 1603,) et *Les Robots et l'Empire*, paru en 1986 chez J'ai Lu en deux volumes (n° 1996 et n° 1997).

¹⁴ *The Caves of Steel*, Doubleday, New York, 1954.

¹⁵ *The Naked Sun*, Doubleday, New York, 1957.

cinquante mondes. Ils n'ont pas non plus les moyens de coloniser de nouvelles planètes. Précisons qu'il n'y a pas de robots positroniques sur Terre à ce moment-là.

Les Robots de l'Aube raconte comment Elijah Baley parvient, grâce à son enquête, à libérer les Terriens de l'isolement et ainsi à déclencher la seconde vague de colonisations extrasolaires. Il est aidé par les robots R. Daneel et R. Giskard. La révélation finale est la suivante : Baley découvre, par son seul raisonnement, que R. Giskard a développé des pouvoirs mentaux télépathiques. Il est capable de lire les émotions et pensées des humains et des robots, et est capable de les modifier. Ce pouvoir est arrivé suite à un accident technologique purement fortuit¹⁶.

6.- Les Robots et l'Empire. *Les Robots et l'Empire* se situe deux cents ans plus tard. Elijah Baley est mort, mais les robots Daneel et Giskard sont encore en activité. Il s'agit de l'un des rares récits d'Asimov où le lecteur accède au récit par le point de vue des robots.

Daneel et Giskard découvrent un complot des Spatiaux qui ont prévu d'isoler la Terre et de laisser les milliards de Terriens végéter dans leurs cités gigantesques en leur interdisant d'accéder à l'espace extrasolaire.

Ce complot est déjoué. Les Terriens obtiennent finalement la suprématie territoriale sur la Galaxie, et les Spatiaux se retrouvent isolés et voués au déclin. Avant de mourir, Giskard transmet ses pouvoirs à Daneel qui devra mettre en œuvre les projets utopiques qu'ils ont tous deux conçus pour l'humanité.

II. – Utopie et Lois de la robotique

7.- Les mondes des Spatiaux, des utopies grâce aux robots ? C'est par l'élaboration des mondes des Spatiaux qu'Asimov parvient à relier les thèmes de l'utopie et des trois Lois de

la robotique. Examinons comment cela se présente.

Les mondes Spatiaux, planètes terraformées pour être idylliques, sont des sortes d'utopie. Par exemple, voilà comment est présentée Aurora dans les *Les Robots de l'Aube* par le Dr Fastofle, le constructeur de Daneel et Giskard. :

« Le résultat est un monde qui produit tout ce dont nous avons besoin, en tenant compte de nos désirs... Voulez-vous que je vous dise vers quel idéal nous avons tendu ?

— Je vous en prie, dit Baley.

— Nous avons travaillé pour créer une planète qui, dans son ensemble, obéirait aux trois Lois de la robotique. Elle ne fait rien qui blesse les êtres humains, par action ou omission. Elle fait ce que nous voulons qu'elle fasse, du moment que nous ne lui demandons pas de faire du mal à des êtres humains. Et elle se protège, à des moments et dans des lieux où elle doit nous servir ou nous sauver même au prix d'un mal fait à elle-même. »¹⁷

Fastofle présente donc Aurora comme une planète obéissant aux trois Lois. Qu'est-ce que cela veut dire ?

Il y a un grand nombre de robots sur les planètes des Spatiaux. Les robots sont littéralement formatés par les trois Lois de la Robotique auxquelles ils ne peuvent échapper. Elles forment une sorte de devoir, un impératif. Quelle activité exerce un robot ? Il doit se préoccuper des êtres humains qui sont ontologiquement proches de lui. En vertu de la première Loi¹⁸, les robots sont programmés pour chercher à établir le bien-être des humains qui sont à proximité d'eux, et faire en sorte que rien ne leur arrive de mal. Et donc chaque humain entouré d'un grand nombre de robots est enveloppé dans une bulle de confort et de sécurité.

¹⁶ Le lecteur apprend cette révélation à travers une tranquille causerie entre Giskard et Baley (p. 246, volume 2). Comme souvent chez Asimov, nombre de scènes de révélations se passent dans un dialogue paisible, entre gens de bonne compagnie, confortablement assis et ravitaillés.

¹⁷ *Les Robots de l'Aube*, vol. 1, p. 136-137.

¹⁸ Pour rappel : un robot ne peut porter atteinte à un être humain, ni, restant passif, permettre qu'un être humain soit exposé au danger.

Cette bulle correspond à une définition minimale de l'utopie. Au minimum, on peut définir l'utopie comme une société dont les membres subissent le moins de mal possible. Chaque Spatial est donc au centre d'une micro-utopie locale. Toutes ces bulles additionnées transforment ces planètes en conglomerat de bulles d'utopies.

8.- De l'utopie locale à l'utopie globale. Les utopies générées par les robots ne sont que des utopies locales, elles ne s'étendent qu'aux humains en contact avec eux. Cela ne pose pas de problème dans les mondes spatiaux ; avec une forte densité de robots associée avec une faible densité d'humains, cette superposition des utopies individuelles englobe toute la planète, comme le dit plus haut le Dr Fastofle¹⁹. Mais ces utopies n'englobent qu'une fraction de l'humanité, celle susceptible d'être sous la surveillance bienveillante des robots.

Mais voilà qu'une rupture se produit dans cette architecture de l'effet utopique du confort social amené par les robots. C'est cette rupture qui va créer les conditions d'apparition de la Loi zéro.

Pour les robots, la question se pose en permanence : comment, par des actions ou des inactions, prévenir au mieux les dangers qui menacent les hommes qui sont à leur contact ?

Les enjeux changent lorsque le rayon d'action des robots est modifié. C'est ce qui est raconté dans *Les Robots de l'Aube*. Suite à une involontaire modification technique, le robot Giskard se retrouve doté de pouvoirs mentaux de type télépathique : il peut lire les pensées et émotions des humains et les modifier. Son rayon d'action augmente. Par rapport à la première Loi, les humains dont il doit s'occuper sont beaucoup plus nombreux que s'il n'était pas télépathe. Il a désormais les moyens d'agir sur beaucoup d'êtres humains, et donc a beaucoup plus de gens sous sa responsabilité.

Donc à lui se pose le problème de l'extension de l'utopie locale qu'il peut générer.

Voilà un enjeu au cœur de ces deux romans, *Les Robots de l'Aube* et *les Robots et l'Empire* : la transformation d'une addition d'utopies locales en une utopie englobant tous les individus.

Voyons comment cela se déroule.

Dans *Les Robots de l'Aube*, une menace apparaît à laquelle Giskard est obligé de faire face. Il découvre par télépathie un complot visant à empêcher les huit milliards de Terriens d'entamer la colonisation de nouvelles planètes, complot dirigé par le Docteur Amadiro, politicien et ingénieur de haut vol.

Giskard découvre ce plan. Avec son potentiel d'action dû à ses pouvoirs, et l'impératif de protection des humains, Giskard est contraint de se prononcer. Doit-il agir ou laisser faire ? Le projet d'Amadiro apporte-t-il un mieux pour les humains en général, ou est-il néfaste ? S'il est certes positif pour les Spatiaux, il est cependant négatif pour les Terriens. Mais quelle partie de l'humanité compte le plus ?

9.- Robots et prise de décision. Giskard doit choisir entre l'inaction – laisser les Spatiaux développer des utopies locales à travers la Galaxie, avec des milliards de Terriens végétant dans une forme de dystopie –, ou l'action – chercher à établir une utopie globale pour toute l'humanité. Il décide de partager cette responsabilité avec R. Daneel.

Les romans *Les Robots de l'Aube* et *Les Robots et l'Empire* racontent leur quête des meilleures décisions à prendre pour l'humanité.

Ainsi apprend-on, à la fin de *Les Robots de l'Aube*, que Giskard a réussi à prendre les bonnes décisions, et à empêcher Amadiro de mettre son plan à exécution. Et grâce également aux manœuvres d'Elijah Baley, les Terriens vont réussir à lancer une vague de colonisations.

Les Robots et l'Empire se déroule quelque deux cents années après *Les Robots de l'Aube*. Les nouvelles colonies fondées par les Terriens menacent la suprématie des Spatiaux et

¹⁹ Dans *Terre et Fondation*, les protagonistes visitent quelques-uns des mondes spatiaux des milliers

d'années plus tard, et le thème de l'utopie locale y sera à nouveau interrogé.

Amadiro (encore lui !) ourdit un vil dessein, celui de tuer les milliards d'habitants de la Terre en la rendant radioactive. Daneel et Giskard vont partiellement déjouer ce plan. Giskard décide tout de même de rendre la Terre radioactive, mais très progressivement, pour laisser cent cinquante ans aux Terriens pour partir dans l'espace. Sa motivation est que les humains se détachent de la Terre, et donc que l'expansion spatiale se fasse sans retour en arrière. Cette décision prise, Giskard meurt et transmet ses pouvoirs télépathiques à Daneel, désormais chargé d'amener les humains à l'Utopie.

10.- Mortelle prise de décision. La fin est très forte : Giskard meurt *parce qu'il a pris la décision de rendre la Terre radioactive en prenant consciemment le risque, certes très faible, de faire mourir des millions d'humains. Il meurt parce qu'il n'est pas sûr de sa décision.* Son cerveau positronique ne survit pas à une telle violation des Lois de la robotique.

Comment les deux robots en sont-ils arrivés à prendre ces décisions si difficiles ? Pour réussir cela, ils suivent tout un cheminement, une vraie quête intellectuelle. Le ressort central de ces deux récits se situe là.

Dans les quatre premiers volumes du *Cycle de Fondation*, certains êtres humains réussissent à prendre des décisions importantes pour la destinée de l'humanité, que ce soit Hari Seldon, Salvor Hardin, Preem Palver ou Golan Trevize, sans que leur cerveau se paralyse.

Par contre, lorsque Giskard prend la décision finale, soit laisser la Terre devenir radioactive pour que l'Utopie s'installe des milliers d'années plus tard, il en meurt !

Notre hypothèse est qu'Asimov révèle ainsi une différence fondamentale entre humains et robots.

11.- Le devoir de Giskard : un dilemme : respecter les trois Lois ou faire ce qu'il faut faire. Au départ, Giskard est confronté à une forme d'impasse : la meilleure chose à faire

entre parfois en contradiction avec les trois Lois. S'il choisit de faire ce qu'il faut faire, et donc de combattre le plan du Docteur Amadiro, il occasionnera fatalement du mal à des humains, en l'espèce Amadiro, et donc Giskard contreviendrait aux trois Lois.

On assiste à ses réflexions pour sortir de cette impasse.

La notion de « devoir » est un premier élément de la différence entre robots et humains. Cela se révèle dans une conversation²⁰ entre Giskard et Elijah Baley. Pour Giskard, le devoir est « ce qui doit être fait ». Elijah Baley lui répond que pour les robots, le devoir est d'obéir aux Lois de la robotique. Il ajoute que les êtres humains ont aussi des lois auxquelles ils doivent obéir. Giskard pose ensuite un constat et une question fondamentale : « Monsieur, je ne peux pas contrevenir aux Lois. Pouvez-vous désobéir aux vôtres ? »²¹

Réponse de Baley : « Je peux choisir de ne pas faire mon devoir, mais je ne le choisis pas, et c'est parfois la pulsion la plus forte. »

Les robots sont donc obligés d'obéir aux trois Lois, alors que les humains peuvent choisir de ne pas obéir à leurs propres lois.

Le problème pour Giskard est d'harmoniser le « faire ce qu'il faut » avec son devoir, son obéissance inconditionnelle aux trois Lois. Pour ce faire, il ne dispose que de la faculté de raison.

12.- Solution : harmoniser le respect des trois Lois et faire ce qu'il faut faire. La première étape de Giskard est de définir « ce qu'il faut faire ».

Cette étape, autorisée par les trois Lois, est de réfléchir à ce qui peut être le meilleur pour l'humanité. Ce projet utopique est un raisonnement, une construction intellectuelle. Tant qu'il ne prend pas de décision au sujet de son application, Giskard est autorisé à simplement y réfléchir.

²⁰ Comme souvent chez Asimov, les représentations mentales des personnages se révèlent à travers les dialogues (*Les Robots de l'Aube*, vol. 1, p. 141).

²¹ Op. cit., vol. 1, p. 142.

Il dévoile son projet à Elijah Baley à la fin des *Robots de l'Aube*. Dans un premier temps, laisser les Terriens coloniser la Galaxie, sans robots pour les assister. Cela engendrera quantité de maux de toutes natures, mais « à la longue, à la fin, les êtres humains bénéficieront d'avoir travaillé par eux-mêmes²² ». Dans un deuxième temps, « peut-être un jour – un jour lointain dans l'avenir – les robots pourront de nouveau intervenir ».

Intervenir, oui, mais comment ? En utilisant la psychohistoire.

Pour Giskard, cette science sociale sera nécessaire un jour. Il la croit possible, car il pense qu'il existe des lois gouvernant le comportement humain, sur le modèle des trois Lois de la robotique, mais différente parce qu'« elles peuvent être si peu contraignantes qu'elles n'ont guère de sens, à moins que ces énormes populations ignorent le fonctionnement de ces lois²³ ». Dans le roman *Les Robots et l'Empire*, Giskard poursuit l'élaboration de cette psychohistoire en dialoguant avec Daneel.

Voilà qui justifie la lecture des *Robots de l'Aube*. Asimov y tisse des fils narratifs et thématiques qui rejoignent les révélations de *Fondation Foudroyée*, en montrant comment les robots ont amorcé le projet utopique, ici le Plan Seldon.

13.- Robots et prison des trois Lois. Mais Giskard est toujours devant son impasse : certes il a le droit de réfléchir à « ce qu'il faut faire », mais la mise en application de ce projet implique des décisions impossibles à prendre à cause des trois Lois. Cela sera l'enjeu du roman *Les Robots et l'Empire*.

Cette impasse se résoudra d'abord par l'emploi d'une faculté commune aux robots et aux humains : la Raison. Les deux robots emploieront leur raison, à travers de longs dialogues entre eux, qui se révèlent indispensables pour qu'ils puissent avancer dans leurs constructions intellectuelles. Cela montre l'importance qu'Asimov attribue au

langage dans le processus de clarification de la pensée.

Au début du roman, suite à une modification mentale qu'il a dû opérer sur Gladia (la Spatiale héroïne de *Face aux feux du soleil* et amante de Baley dans *Les Robots de l'Aube*), Giskard déclare que les trois Lois sont incomplètes. Il lui apparaît que la modification de l'esprit de Gladia « était la meilleure réponse possible dans la limite des trois Lois, mais pas la meilleure réponse possible dans l'absolu²⁴ ».

À la question de Daneel sur cette meilleure réponse possible, Giskard affirme qu'il ne peut l'exprimer tant qu'il est lié par les trois Lois. Et quand Daneel lui affirme qu'il n'y a rien au-dessus des trois Lois, Giskard met en avant une autre différence fondamentale entre humains et robots : « Si j'étais humain, je parviendrais à voir au delà des Lois. »

Selon Giskard, la solution serait de modifier les trois Lois. Mais « je ne peux les modifier – parce qu'elles se trouvent sur mon chemin²⁵ ».

Giskard est structuré par un paradigme, les trois Lois, qui surdéterminent son cerveau. Il doit donc utiliser sa raison pour redéfinir ce paradigme. Cela devient impératif lorsqu'il détecte dans l'esprit de Mandamus, un comparse d'Amadiro, l'existence d'un complot visant à tuer des milliards d'humains. Hélas pour connaître les détails du complot, il devrait endommager son esprit. Giskard ne peut pas le faire, prisonnier des trois Lois. À ce stade du récit, Giskard et Daneel sont au désespoir. Plus tard, ils débarquent sur Solaria, accompagnant Gladia. Ils seront attaqués par des robots. Grâce à ses pouvoirs télépathiques, Giskard triomphe sans peine.

Giskard et Daneel y font une découverte cruciale : la possibilité de redéfinir le paradigme des trois Lois. En effet, les robots Solariens qui les ont attaqués ont été programmés pour ne considérer comme humains que les natifs de Solaria.

²² Op. cit., vol. 2 p. 253.

²³ Op. cit., vol. 2 p. 253.

²⁴ *Les Robots et l'Empire*, vol. 1, p. 23.

²⁵ Op. cit., vol. 1, p. 24.

C'est Daneel qui avance le raisonnement : « *Si*²⁶ l'on pouvait cerner les instructions par des définitions et des instructions, *si* l'on pouvait détailler les instructions d'une manière suffisamment rigoureuse, *serait-il* aussi possible de tuer un être humain pour une raison sérieuse que de sauver la vie d'un autre être humain ? »²⁷

Giskard répond que c'est possible. Daneel s'appuie là-dessus pour poursuivre son raisonnement. Il en déduit alors que les trois Lois ne sont pas absolues. Elles pourraient être modifiées.

Giskard supplie Daneel ne pas aller plus loin.

14.- Emploi de la Raison et danger de mort. En effet, en raisonnant sur les trois Lois, Giskard et Daneel sont en danger de mort ! Leur cerveau positronique peut se paralyser. Pour eux, robots, utiliser la raison est une source de danger mortel : une construction mentale mal étayée, et ils meurent ! Voilà pourquoi ils ont besoin d'être deux à raisonner. Ce n'est pas le cas pour les êtres humains, qui peuvent se tromper sans risquer la mort.

« Encore un pas, ami Giskard. Le camarade Elijah l'aurait franchi.

– C'était un être humain. Il le pouvait. »²⁸

Les humains peuvent sans risque remettre en question les principes qui structurent le comportement. Encore une nouvelle différence entre humains et robots !

Et Daneel brise la prison des trois Lois en faisant un pas supplémentaire : « Ne serait-il pas possible que nous-même, *peut-être*, dans certaines conditions précises, puissions mod... »²⁹

Ici, Daneel s'arrête pour éviter la paralysie mortelle.

III. La Loi zéro : une transgression des trois Lois au service de l'humanité

15.- Loi zéro et Humanité. Voici maintenant un passage clé de *Les Robots et l'Empire*, qui va aboutir à la proclamation de la Loi zéro.

Plus tard dans le roman, Daneel sera contraint de poursuivre son raisonnement sous peine que lui et Giskard soient détruits, et que l'humanité sorte du long chemin vers l'utopie.

Giskard est confronté à Vassilia, fille du Dr Falstofle, roboticienne et comparse d'Amadiro. Celle-ci s'est enfermée dans une pièce avec nos deux robots. Nous assistons à un combat de langage et de raison. Celui qui argumentera le mieux vaincra. Si Daneel n'argumente pas suffisamment bien, lui et Giskard mourront et l'humanité n'accédera pas à l'utopie.

L'enjeu du combat sera de convaincre Giskard. En effet, Vassilia donne l'ordre à Giskard d'anéantir l'esprit de Daneel. Giskard est prisonnier de la deuxième Loi³⁰, qui l'oblige à obéir à un humain si la première Loi³¹ n'est pas en cause. C'est le cas ici : aucun être humain n'est menacé. Giskard doit obéir, prisonnier des trois Lois. Daneel doit donc modifier les trois Lois qui structurent Giskard. Il doit le faire par un raisonnement, de telle sorte que Giskard puisse s'automodifier en suivant ce raisonnement.

Vassilia lui donne l'ordre de se taire, s'appuyant également sur la deuxième Loi.

Daneel désobéit en invoquant la Loi zéro, une Loi qui supprime la première Loi, et donc aussi la deuxième.

Daneel décrit le cheminement de son raisonnement. Il commence par constater l'incomplétude des trois Lois. Selon lui, cela vient de ce que l'humanité est plus importante qu'un seul être humain. Il énonce alors la Loi

²⁶ Les italiques sont ceux du texte original.

²⁷ Op. cit., vol. 1, p. 214.

²⁸ Op. cit., vol. 1, p. 215.

²⁹ Op. cit., vol. 1, p. 215.

³⁰ Pour rappel, « 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 ».

³¹ Pour rappel, « un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ».

zéro : « Un robot ne doit causer aucun mal à l'humanité ou, faute d'intervenir, de permettre que l'humanité souffre d'un mal. »³² Pour étayer cela, Daneel s'appuie sur une métaphore qu'Elijah Baley lui a confiée avant de mourir : « Daneel, garde l'esprit fermement fixé sur la tapisserie et ne te laisse pas affecter parce qu'un fil s'estompe. Il y en a tant d'autres, tous précieux, chacun jouant son rôle... »³³

Daneel affirme donc que « la tapisserie de la vie est plus importante qu'un seul fil³⁴ ».

16.- L'humanité est-elle un concept abstrait ?

À cette question, Vassilia expose une objection majeure : l'humanité n'est qu'un concept, une abstraction ; elle n'a pas de valeur ontologique.

« Les trois Lois de la Robotique concernent les êtres humains en tant qu'individus et les robots en tant qu'individus-robots. Mais qu'est-ce que « l'humanité » sinon une abstraction ? Peux-tu toucher l'humanité ? Tu peux blesser ou éviter de blesser un être individuel et comprendre le préjudice ou l'absence de préjudice. Peux-tu voir le préjudice ou l'absence de préjudice causé à l'humanité ? Peux-tu le comprendre ? Peux-tu le montrer du doigt ? »³⁵

Elle demande ensuite à Giskard d'arbitrer le débat. Celui-ci confirme la validité de l'argument de Vassilia. Il se justifie en invoquant l'histoire de l'humanité. De grands crimes ont souvent été accomplis en se justifiant par ce concept d'humanité. Et c'est parce que l'humanité est une abstraction qu'on peut si aisément en appeler à elle pour tout justifier.

C'est donc une victoire dans ce débat pour Vassilia. Daneel tente un ultime argument pour convaincre Giskard, en lui demandant s'il est conscient des conséquences de son choix. Giskard confirme qu'il l'est, mais que « le danger dont tu parles ne constitue pas une certitude, mais découle d'une déduction. Et

l'on ne peut fonder nos actions au mépris des trois Lois pour autant³⁶ ».

Giskard est toujours prisonnier des trois Lois, et c'est le juste exercice de sa raison qui le retient prisonnier.

Vassilia fait appel à quatre robots et leur ordonne de démanteler Daneel. Et surprise, Giskard agit. Il paralyse les robots, endort Vassilia et efface une partie de sa mémoire.

Plus tard, en sécurité, Giskard et Daneel reviennent sur le débat.

Giskard s'explique. Il n'a pas été convaincu par le raisonnement de Daneel, mais le raisonnement a offert assez d'éléments convaincants pour réinterpréter la Première Loi, lui faisant bénéficier d'une marge de manœuvre.

Ils en déduisent que « penser suffisamment » à ce concept de Loi zéro peut affaiblir la Première Loi. Il y a donc une certaine force de la raison.

Pourtant, tous deux voient le manque de présence ontologique de l'Humanité comme un obstacle, qui leur interdit de s'affranchir définitivement de la Première Loi.

Cependant, ils persévèrent : « C'est là la difficulté, ami Daneel, avec ta Loi zéro. La Première Loi traite d'individus spécifiques et de certitudes. Ta Loi zéro traite de groupes abstraits et de probabilités. »³⁷

Daneel fait appel à l'expérience de Giskard : « Ne peut-on imaginer qu'il existe dans la Galaxie le bourdonnement de l'activité mentale de l'humanité tout entière ? Comment, dans ce cas, l'humanité serait-elle une abstraction ? C'est quelque chose de palpable. Rapproche cela de la Loi zéro et tu verras que l'extension des Lois de la Robotique est justifiée... justifiée par ta propre expérience. »³⁸

Ici, Daneel fait référence à ce qu'a ressenti Giskard pendant un discours de Gladia,

³² Op. cit., vol. 2, p.145.

³³ Op. cit., vol. 1, p. 276.

³⁴ Op. cit., vol. 2, p.145.

³⁵ Op. cit., vol. 2, p. 145.

³⁶ Op. cit., vol. 2, p. 146.

³⁷ Op. cit., p. 192.

³⁸ Op. cit., p. 194.

lorsqu'il a senti, grâce à ses pouvoirs mentaux, l'activité de la foule tout entière.

Giskard accepte le raisonnement. Encore une fois, un robot a eu besoin de la confirmation d'un autre robot pour accepter un bouleversement des principes. Daneel a désormais modifié son paradigme structurant en y incorporant la Loi zéro comme loi suprême. Cela se révèle dans le récit lorsqu'il choisit de protéger Giskard plutôt que Gladia lors d'un attentat.

Nous assistons au triomphe de la raison chez Daneel et Giskard, qui se sont auto-modifiés grâce à son usage rigoureux. Et cela grâce également au langage et au dialogue. En effet, le langage naturel a comme particularité fondamentale que les mots offrent une latitude dans l'interprétation de leurs définitions, contrairement au langage mathématique. Et c'est le dialogue qui leur permet de coconstruire les définitions qui permettent de transgresser les trois Lois. Je développe dans mon essai pourquoi le style d'Asimov, parfois décrié, est pourtant fidèle et cohérent à la philosophie du langage esquissée ci-dessus.

17.- Pour que tu t'y intéresses. Pour être complet sur cette thématique de la fécondité du dialogue, signalons la nouvelle *Pour que tu t'y intéresses* (*That thou art mindfull of him*), parue en 1974. Elle présente l'intérêt de faire un lien entre les premières nouvelles de robots et *Les Robots de l'Aube*. La problématique de la marge de manœuvre des robots au sujet de l'interprétation d'une des Lois de la Robotique y est déjà évoquée. Cette marge de manœuvre est ici conceptualisée comme la capacité de « jugement ». Le directeur de U.S. Robots, successeur de Susan Calvin, est face à un problème : un robot aura beaucoup de difficulté à obéir à la deuxième Loi s'il est confronté à un grand nombre d'êtres humains susceptibles de lui donner des ordres contradictoires. Il va ainsi demander à son plus récent robot, George 10, de réfléchir à une solution technique. Ce dernier demande à être accompagné dans sa recherche par un autre robot, George 9, pour prévenir la paralysie

mortelle par le dialogue rationnel. Ici, les deux George préfigurent en quelque sorte Daneel et Giskard.

Dans ce récit, y est également évoqué le thème de la redéfinition par les robots de ce qu'est un être humain, avec comme conséquence que les deux George se considèrent finalement comme supérieurs aux humains et se permettent d'envisager d'être les futurs maîtres de l'histoire humaine. Ainsi cette nouvelle présente déjà une réflexion d'Asimov sur des thématiques plus amplement approfondies par la suite.

18.- Application de la Loi zéro, une voie vers l'utopie globale. Une fois la Loi zéro acceptée, reste le problème de l'application de cette Loi. Désormais, les deux robots sont contraints d'installer l'humanité dans une utopie globale.

C'est ainsi que lors de la confrontation finale, Giskard devra procéder à des décisions critiques, hélas, sans psychohistoire à disposition.

Les deux robots se retrouvent sur Terre face à Amadiro. Celui-ci s'apprête à rendre la Terre mortellement radioactive, provoquant des milliards de morts. Il est neutralisé par Giskard. Giskard prend ensuite la première décision qui mettra l'humanité sur la voie de l'utopie. Il décide que cette radioactivité mortelle mettra plutôt cent cinquante ans à s'installer. Cela laissera le temps aux Terriens de partir dans l'espace et d'abandonner la Terre. Mais cette décision entraîne la mort de Giskard. Il meurt parce qu'il y a des incertitudes à la base de sa décision. Il meurt parce que la raison l'exige. Sa décision n'est pas complètement justifiée par des arguments formellement inattaquables, car il n'a pas la psychohistoire à sa disposition.

Avant de mourir, il transmet ses pouvoirs à Daneel.

Et Daneel se retrouve seul en charge de la responsabilité de l'humanité ; une solitude et une responsabilité écrasantes. Je ne dévoilerai pas ici ce qui adviendra de Daneel dans les romans suivants, *Terre et Fondation*³⁹, *Prélude à*

³⁹ *Foundation and Earth*, Doubleday, 1986

*Fondation*⁴⁰ et *Aube de Fondation*⁴¹, afin de laisser au lecteur le plaisir de la découverte des dernières réflexions d'Asimov.

19.- Différence entre humains et robots, et éthique kantienne. Les humains ordinaires ne sont pas comme les robots. Si les humains prennent une mauvaise décision, leur cerveau ne se paralyse pas, ils n'en meurent pas. Alors que les robots, s'ils prennent une mauvaise décision, qui trahit une Loi de la robotique, leur cerveau se fige, et ils meurent.

La philosophie de Kant offre des concepts qui permettent de décrire cela⁴². Kant partage l'esprit en deux espaces différents⁴³. Il y a l'espace phénoménal, soumis à la loi de la causalité et sensible aux données empiriques. Et l'espace nouménal qui transcende l'espace phénoménal, et est totalement indépendant des données empiriques. C'est là que s'exerce l'autonomie de la volonté, de manière inconditionnelle. La volonté peut y exercer la prise de décision en toute liberté, hors de toute la loi causale qui s'exerce dans le monde phénoménal, par exemple les humeurs, sensations ou sentiments. Dans cet espace, c'est la bonne volonté qui est responsable de la justesse de la décision éthique. L'esprit peut donc choisir entre deux principes ou deux valeurs en toute liberté. Kant considère que la nature fondamentale de l'homme est sa liberté de s'imposer des devoirs.

Dans les cycles des Robots et de Fondation, les humains jouissent de cet espace nouménal transcendantal, comme on l'a vu dans les dialogues entre Baley et Giskard et entre Giskard et Daneel. Ils peuvent désobéir aux lois ou à la morale. Et ils peuvent obstinément persister à vivre dans des sociétés où règne une forte détresse humaine.

Alors que les robots sont prisonniers d'une réalité phénoménale qui leur impose une loi causale, un déterminisme. D'une part leur cerveau est structuré ontologiquement par les Lois de la robotique. D'autre part l'exigence de l'exercice juste de la raison est aussi inscrite ontologiquement dans leur cerveau, car tout manquement entraînerait leur mort. C'est cette combinaison qui conduit Daneel et Giskard à la proclamation de la Loi zéro, et au devoir impératif d'amener l'humanité à l'utopie⁴⁴.

Faudra-t-il nécessairement que surviennent les robots asimoviens pour que les humains aient la moindre chance de vivre dans une utopie globale ?

Conclusion

20.- Robots, Utopie, Raison et Science. À travers cet exposé, nous avons décrit comment Asimov met en scène la Loi zéro, et les enjeux philosophiques qui y sont liés, à travers les concepts de la raison, le langage, le dialogue, la prise de décision, l'humanité, l'utopie, et cet espace de liberté intérieure qui semble nous différencier des robots.

Dans les deux romans examinés, l'utopie est définie dans la Loi zéro : « Un robot ne doit causer aucun mal à l'humanité ou, faute d'intervenir, de permettre que l'humanité souffre d'un mal. »

L'utopie, c'est de ne souffrir d'aucun mal. Vision pour le moins sèche et désenchantée !

Quelle est la part de la raison et de la science dans ce paradigme de l'utopie proposé ici ?

Daneel et Giskard ne peuvent survivre que s'ils parviennent à concilier complètement un paradigme idéal, la Loi zéro, avec l'humanité réelle. Le danger mortel d'un mauvais exercice

⁴⁰ *Prelude to Foundation*, Doubleday, 1988

⁴¹ *Forward the Foundation*, Doubleday, 1992

⁴² Pour ce commentaire, je me réfère au vocabulaire et à la description proposés par Jeanne Hersch dans *L'Etonnement philosophique*, 1981, coll. folio/essais, Gallimard. Les pages qu'elle consacre à l'éthique kantienne résonnent profondément avec le dispositif spéculatif que sont les robots asimoviens.

⁴³ Notre exposé sur Kant peut paraître lapidaire et simplifié, eu égard à la complexité de la philosophie de

Kant, mais une démonstration plus approfondie excéderait les ambitions de cette communication.

⁴⁴ Il serait intéressant de décrire aussi les contraintes des robots par les concepts d'éthique de la responsabilité et d'éthique de la conviction, tels que développés par Max Weber. Les robots semblent devoir assumer ces deux éthiques de façon absolue dans chacun de leurs actes.

de la raison les guette toujours. Une décision mal justifiée sur le devenir de l'humanité, et c'est la mort.

Respecter la Loi zéro sans mourir implique un instrument pour peser le coût humain des actions, une science dure qui guide les décisions ; une sociologie dure, la psychohistoire. Dans ce paradigme, la science est la seule voie de survie des robots. En effet, l'exigence de la juste conduite de la raison implique une attitude scientifique, une démarche qui doit faire correspondre au mieux les modèles et le réel. Sans elle, ils meurent.

Dans cette œuvre, nous ne sommes pas dans le *sense of wonder*. Ici, aucune mystique qui puisse faire rêver un lecteur en mal de spiritualité, nuls voyages sur des mondes extraordinaires foisonnants d'une vie extraterrestre luxuriante. Mais nous assistons à l'opiniâtre travail intellectuel de deux esprits aux prises avec des principes et des concepts, cherchant une assise ontologique à leurs constructions mentales. L'action rebondit principalement grâce aux dialogues, où les arguments sont potentiellement mortels.

Et, je crois qu'on reconnaît là une valeur centrale qui a toujours guidé Isaac Asimov : la quête du savoir et de la raison.

Par la voix d'Hari Seldon, peut-être Asimov a-t-il livré la clé d'une vie vouée à la recherche du meilleur pour l'homme.

« Pourquoi, se demanda-t-il, tant de gens passent-ils leur vie à éviter de trouver les réponses aux questions - et en premier lieu à éviter les questions ? N'y a-t-il pourtant rien de plus excitant que de chercher des réponses ?⁴⁵ »

A. V.

⁴⁵ *Prélude à Fondation*, p. 260, *Prelude to Foundation*, Doubleday, 1988.

Conclusion : l'œil du praticien

Gilles VERCKEN

Avocat aux barreaux de Paris et Strasbourg
Cabinet Vercken et Gaullier

Introduction

Prenons l'hypothèse suivante : recevant un client, l'avocat l'interroge sur la manière dont il peut répondre à son besoin.

Le client lui tend les lois d'Asimov, et lui pose la question : comment dois-je me comporter au regard de ces « lois » ? Comment les appliquer ?

Que répondre ?

Nous envisagerons les réponses possibles dans une première partie.

Mais les lois d'Asimov sur le « robot » impliquent d'élargir la vision, et de poser la question de l'impact sur notre pratique de ce qui est dénommé l'intelligence artificielle, terme générique dévoyé, qui ne rend pas compte de sa polysémie. Ce sera notre seconde partie.

I. L'analyse du praticien

Les Lois d'Asimov énoncent de nombreuses dispositions qui, une fois envisagées par le praticien à travers un prisme de technique juridique, vont éveiller en lui deux séries d'interrogations. Ces questionnements vont concerner d'une part, le destinataire de ces Lois (A.) et, d'autre part, le contenu même de ces Lois (B.).

A. Les incertitudes liées au destinataire de ces Lois

La première étape de réflexion posée par les Lois d'Asimov¹ consiste en l'identification claire du destinataire desdites règles. En apparence, ces dernières sont adressées au « robot ».

Bien que cette identification paraisse non équivoque, il faut alors s'interroger. Des règles peuvent-elles s'appliquer à un robot, ce qui amène à la question en amont : qu'est-ce qu'un robot ? Il convient d'en envisager une définition juridique.

Aux yeux de la loi – et du praticien – la réponse est claire, et sans appel : le robot est un objet, et uniquement un objet.

Toute théorie tentant d'accueillir l'idée du robot en tant que sujet de droit, comme s'efforce notamment de l'affirmer certains, en proposant qu'il puisse jouir de la personnalité juridique, n'a aucun fondement juridique à ce jour.

Au-delà de ce premier constat de droit positif, l'admission d'un tel postulat – le robot comme sujet de droit – , poserait un réel problème philosophique. Nos sociétés et notre droit ont été construits par et pour l'homme. La protection de l'humanité est la pierre angulaire de notre société, et nous renvoyons aux écrits de Jean-Claude Guillebaud², qui expliquent parfaitement la nécessité irréductible de consacrer l'être humain comme la base de l'existence même, non par volonté

¹ Ces lois ont été énoncées pour la première fois dans la nouvelle d'Asimov, « Runaround » (« Cercle Vicieux »), Edition Astounding Science Fiction, mars 1942.

² Jean-Claude Guillebaud, « Le principe d'humanité », Editions du Seuil, 2001, 384 pages.

d'ethnocentrisme, mais comme fondement même de notre humanité.

De cette manière, admettre dans un système social juridique construit autour des rapports humains qu'un objet créé par l'homme puisse le concurrencer – ou à tout le moins disposer de prérogatives équivalentes aux siennes – serait un non-sens au plan du raisonnement juridique et contraire aux valeurs fondamentales de la société.

Une telle dangerosité prend également forme à travers deux autres phénomènes autour du « robot ».

Dans un premier temps, il est de nos jours fréquent de constater que, dès lors qu'un nouveau champ d'expérimentation social et juridique voit le jour, les théories juridiques les plus diverses et les plus farfelues émergent, par volonté sans doute d'obtenir une certaine publicité.

Un second phénomène consiste quant à lui en une exagération absolue du sujet prétendument nouveau. Ainsi, force est de constater qu'aborder la question du robot et de l'intelligence artificielle à travers le panorama offert par la presse conduit bien souvent à une vision tout à fait erronée de ce phénomène. En effet, bien qu'apportant une grande fraîcheur par sa naïveté, ce dernier s'avère démenti par la réalité des faits qui sont soumis à l'analyse, comme celle de l'illusion d'une intelligence artificielle supplantant celle de l'homme, alors qu'elle est créée par lui, et que, pour l'heure, le fonctionnement des systèmes résulte des contours donnés en amont par l'humain.

En résumé, les lois d'Asimov ne peuvent que s'adresser à l'être humain.

B. Les questionnements liés au contenu des Lois

Une fois identifié le destinataire, les règles doivent s'appliquer. La deuxième question concerne le contenu des règles édictées par les Lois d'Asimov. Les règles fixées déterminent finalement – et assez simplement – des caractéristiques du produit et de la responsabilité de celui qui l'accompagne.

« 1. Un robot ne peut porter atteinte à un être humain, ni, en restant passif, permettre qu'un être humain soit exposé au danger ;

2. 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 ;

3.- Un robot doit protéger son existence tant que cette protection n'entre pas en conflit avec la première ou la deuxième loi »

La première règle peut être perçue par le praticien que comme un concentré des articles 1240 et suivants du Code civil, énonçant notamment que « [t]out fait quelconque de l'homme, qui cause à autrui un dommage, oblige celui par la faute duquel il est arrivé à le réparer ». Cette première « loi » rappelle ainsi les fondamentaux de la responsabilité civile et on peut lui donner ce sens. Mais elle ne s'adresse qu'à l'homme : le robot ne peut être utilisé pour nuire à autrui, et vous devez préserver les biens et les personnes.

En revanche, la deuxième Loi suscite, quant à elle, beaucoup plus d'intérêt puisqu'elle semble exempte de toute signification pratique, car elle est inutile si elle s'adresse à l'homme, qui devrait alors obéir à lui-même.

En effet, plus qu'une loi *per se*, cette deuxième Loi induit une compréhension du robot comme étant un produit, en faisant référence à la caractéristique même d'un produit, selon laquelle ce dernier doit répondre aux attentes de l'être humain. Ainsi, le robot doit-il par définition répondre aux instructions dictées par l'homme, sous réserve, qu'elles ne soient pas incompatibles avec la première Loi, c'est-à-dire qu'elles ne portent atteinte ni aux personnes ni aux biens.

À travers l'œil du praticien, l'introduction de cette deuxième Loi semble dénuée de sens dans la mesure où elle opère un renvoi à la première règle qui, *de facto*, n'appelait pas à plus de précisions.

Ainsi, au-delà de l'extraordinaire exercice littéraire en lequel consiste cette règle et du potentiel fictionnel que fait naître la notion

d'incompatibilité³ qui en résulte, la rédaction de cette deuxième Loi ne semble pas trouver écho dans l'application concrète à laquelle aspire le praticien.

Enfin, la troisième Loi suscite tout autant de perplexité dans une perspective pratique. En effet, il semble difficile pour un praticien d'appréhender la pertinence de cette troisième règle, qui consiste en une prérogative d'autoprotection du robot. Le robot, à travers cette règle, paraît envisagé en tant que sujet devant assurer sa propre sécurité. Cependant, la priorité de la première Loi d'Asimov, selon laquelle le robot doit avant tout assurer la sécurité et l'intégrité de l'humain vient en contradiction avec la troisième Loi, dans de nombreuses situations aisées à envisager. La troisième règle vient en contradiction avec la première dans la mesure où elle s'oppose à la conception du robot comme un objet, un produit avant tout au service de l'homme, que celui-ci puisse donc décider de détruire s'il le souhaite, ou de programmer le robot pour qu'il le fasse. Ou alors il faut lire la règle comme interdisant l'obsolescence, et prônant l'éternité ? De telles contradictions génèrent des complications pratiques et des paradoxes évidents.

II.- La mise en œuvre des Lois par le praticien

L'avènement des phénomènes de robotique et plus largement de l'intelligence artificielle, font naître des difficultés inédites dans la pratique juridique. Dans son quotidien professionnel, le praticien a pour principales missions de sécuriser et de valoriser les activités de son client en utilisant les règles et outils de droit à sa disposition.

La spécificité de l'intelligence artificielle, pour un praticien, réside dans un constat simple : le résultat appréhendable est issu de la mise en

œuvre d'un système dont l'être humain n'avait pas conscience de ce résultat, même s'il a mis en œuvre les moyens permettant ce résultat.

Comment prendre en compte les « robots intelligents » aux deux stades essentiels de l'intervention du praticien : le conseil (A.) ou la voie contentieuse (B.).

A. La phase du conseil

Dans la phase du conseil, la principale difficulté du praticien va être de définir la charge des obligations de chaque partie au contrat, ainsi que de déterminer à qui incombera la charge de la preuve en cas de fait dommageable, et ce au moyen de la définition des obligations, et des clauses de garantie ou de clauses d'assurance.

À titre d'exemple, cette difficulté s'est présentée dans le cas d'un client proposant une application mettant en place un système automatique « auto-apprenant » sur la manière dont une famille gère la température des différentes pièces de la maison. L'application en question était fonction de paramètres tels que les habitudes de vie de la famille, les moments de l'année, de l'heure du jour ou de la nuit, et devait permettre de fixer la bonne température en fonction de ces multiples critères. Pour la mise en œuvre complète du système, interviennent de nombreux acteurs : les occupants, le locataire, le propriétaire, le fabricant de la maison, les fournisseurs du radiateur, du logiciel dans le tableau électrique, le fabricant des différents capteurs disposés dans la pièce et son fournisseur, le fournisseur du pont multimédia, le fournisseur de l'application. Dans ce cas précis, les principales difficultés ont été de définir d'une part les responsabilités de chacun en cas de dommage et d'autre part de prévoir des clauses de garantie en cas de responsabilité engagée pour un tel fait, et ce pour une raison simple qui est au cœur de la problématique de la robotique : quel est le fait ou plutôt les faits

³ La notion d'incompatibilité a été traitée à de nombreuses reprises dans les nouvelles d'Isaac Asimov. A titre d'exemple : « *Runaround* », mars 1942 et « *Le Cycle des Robots* », Tome 2 « *Les Cavernes*

d'acier », New-York (Etats-Unis), Edition Doubleday, juin 1954.

générateurs du dommage, dans une réalité où les causes peuvent être multiples et relever de l'intervention de plusieurs acteurs, et, pour partie, être généré par le « robot » (pour simplifier).

Ces problématiques se multiplient face à l'essor de cas complexes comme celui précité. Il convient donc, au-delà de la définition des obligations de chacun, de déterminer les personnes responsables à l'origine d'un éventuel dommage en réglementant par contrat la charge de la preuve, les moyens de la preuve, sa force probante (convention de preuve) et/ou en désignant le responsable, et d'envisager des systèmes pour se prémunir contractuellement contre les dommages causés. Les clauses d'assurance sont alors aussi le complément essentiel de ces responsabilités désignées.

Trois axes donc : définition de la charge et des modes de preuve, désignation d'un responsable, assurance et garantie.

B. La phase contentieuse

Durant la phase contentieuse, le principal défi va être aussi de déterminer la charge de la preuve de la responsabilité et définir le fait générateur du dommage. Cependant, face à un litige effectif et en cas de carence dans la rédaction du contrat, la spécificité de ces contentieux repose sur la technicité des faits, et la difficulté d'identifier les causes des dommages. Notamment, des expertises pointues et extrêmement onéreuses seront requises pour définir qui a fait quoi, identifier les faits générateurs et les liens de causalité, ce qui pose aussi le problème droit d'accès à la justice⁴. L'existence de systèmes autonomes renforcera cette difficulté d'identification des causes et conséquences.

Mais il est primordial de bien définir au préalable l'impact potentiel de ces nouveaux problèmes pour l'homme. Il ne serait pas

pertinent de céder en premier lieu à la tentation de la création d'un nouveau droit avant que les phénomènes auxquels nous devons répondre soient avérés, et qu'il soit démontré que le cadre existant n'est pas suffisant pour répondre aux cas pratiques.

Quelques mots sur le droit d'auteur en guise de conclusion : dans le domaine du droit d'auteur, la question du statut des créations « assistées par ordinateur » dans les domaines de la musique, des arts visuels et graphiques est posée depuis très longtemps. La réponse juridique, applicable aussi à l'intelligence artificielle, est connue.

Lorsqu'on ne peut pas établir de lien de cause à effet entre un choix fait par l'homme et la forme extérieure, alors un droit d'auteur ne peut être envisagé.

Le droit apporte donc une réponse a priori négative sur la protection par la propriété intellectuelle de formes non générées par l'homme. Est-elle pertinente ? C'est un autre débat, et il est à notre sens parfois inutile de tordre le droit pour l'amener à s'appliquer à une situation pour laquelle il n'a pas été conçu.

Il convient alors de recentrer les débats autour de l'éventuelle nécessité de créer de nouvelles protections juridiques. Faut-il créer un nouveau droit face à une nouvelle situation ? Cette dernière solution n'est pas à prioriser ni à préconiser, mais elle doit être analysée. La tentation est forte, et compréhensible, de créer de nouveaux droits de propriété sur de nouveaux objets, autonomes au regard du droit d'auteur (cf. l'exemple des droits voisins de l'éditeur de presse).

Entre le conformisme éventuel (les robots ne changent rien) et l'angélisme qui peut nous saisir à l'arrivée de nouveaux objets (les robots changent tout), l'emprunt d'une voie médiane, de curiosité critique et attentive est à favoriser : le droit en pratique peut à ce jour répondre aux robots !

⁴ Accès garanti à l'article 6 de la Convention Européenne des Droits de l'Homme (CEDH) qui dispose que : « 1. Toute personne a droit à ce que sa cause soit entendue équitablement, publiquement et dans un délai raisonnable, par un tribunal

indépendant et impartial, établi par la loi, qui décidera (...) des contestations sur ses droits et obligations de caractère civil (...).