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, actuateurs, 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 humain[4] et l'obligation d'obéir aux ordres qui lui sont donnés par un être humain[5], 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, aspirateur[9], tondeuse à gazon[10]) 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'intime[12],

[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 paraitront 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 délinéation 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 Fœssel, *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 détections, 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 réentrainés 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 (LiDAR[41]) 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'évènements 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 « boite 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 évènements passés. La même année Philip K. Dick publie « Minority report » récit présentant une autre forme de chronoscopie destinée cette fois à anticiper les évènements 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 « Astouding 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. requérait 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 privée[51] ou du droit à un procès équitable[52] —impose précisément l'accès à l'enchaînement causal à l'origine de ces décisions. L'immixtion d'un intermédiaire algorithmique autonome et opaque[53] 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,

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 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'IBM[61] 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]

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'usager 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]

---

[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-doit-opter-pour-une-approche-ou-lhomme-reste-aux-commandes-affirme-le-cese

[70] Supra, not. n° 68, pt. 1.7 ; Journal Officiel de l'UE (JO UE) C 288, 31 aout 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. Dépassant 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 « imaginatif », 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'enferrer. 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 formatent 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'usager, 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