





# **L'intelligence artificielle**

*Dans toutes ses dimensions*

**Collection Europe & Asie**  
dirigée par JI Zhe et Michel MAZOYER

La mondialisation actuelle, qui voit simultanément l'essor de l'Asie et la recomposition de l'Europe, s'accompagne dans les sociétés humaines de changements fondamentaux dans le rapport au temps et à l'espace. Dès lors, malgré l'enchaînement des crises, malgré l'instabilité et les menaces qui pèsent constamment sur nous, les échanges intellectuels croissants entre Europe et Asie ainsi que les nouvelles possibilités qui en découlent renforcent notre capacité à faire face aux défis et aux épreuves, tout en conduisant peut-être l'humanité à un rapport plus harmonieux à la nature.

Cette collection propose un regard croisé sur l'Europe et sur l'Asie. Comme son prédécesseur la revue *Europeana* (2013-2017, L'Harmattan), elle publie des recherches analysant les évolutions culturelles, sociales et politiques de l'Europe, soit pour les présenter aux lecteurs asiatiques à travers nos partenaires chinois, soit pour montrer aux lecteurs européens une perspective asiatique. Elle s'attache également à l'histoire et à l'actualité de l'Asie et aux appropriations dont celle-ci a fait ou peut faire l'objet en Europe. La collection se veut un outil de compréhension et de connaissances pour l'Extrême-Orient et l'Extrême-Occident. L'Europe et l'Asie sont ici envisagées comme le résultat d'une expérience historique et intellectuelle susceptible d'interpeller l'une et l'autre, qui sont toutes les deux en pleine redéfinition.

**Dernière parution**

Boris Barraud (dir.), *Le Brexit. Dans toutes ses dimensions*, 2018.

Sous la direction de Boris Barraud

# L'intelligence artificielle

*Dans toutes ses dimensions*

L'Harmattan

\*\*\*

© L'Harmattan, 2020  
5-7, rue de l'École-Polytechnique – 75005 Paris  
[www.editions-harmattan.fr](http://www.editions-harmattan.fr)  
ISBN : 978-2-343-19274-1  
EAN : 9782343192741

## Sommaire

Introduction.....	13
L'IA : dimension historique.....	79
L'IA : dimension politique .....	107
L'IA : dimension sociologique .....	127
L'IA : dimension économique .....	145
L'IA : dimension anthropologique .....	175
L'IA : dimension juridique .....	193
L'IA : dimension chinoise .....	205
L'IA : dimension socio-scientifique .....	243
L'IA : dimension philosophique .....	265
L'IA : dimension prospective .....	285
Table des matières .....	299



## Les auteurs

*Boris Barraud* est enseignant-chercheur à l’Université Grenoble-Alpes, membre du Centre de recherches juridiques (CRJ). Docteur en droit et titulaire d’un master en science politique et en sciences de l’information et de la communication, il étudie la communication numérique et l’intelligence artificielle dans leurs aspects politiques, juridiques et philosophiques.

*Laurence Devillers* est professeure en informatique appliquée aux sciences sociales à l’Université Paris-Sorbonne, membre du Laboratoire d’informatique pour la mécanique et les sciences de l’ingénieur (LIMSI-CNRS) et de l’équipe de recherche Dimensions affectives et sociales dans les interactions vocales. Elle a notamment écrit l’ouvrage *Des robots et des hommes – Mythes, fantasmes et réalités* (Plon, 2017). Depuis 2015, elle est membre de la Commission de réflexion sur l’éthique de la recherche en sciences et technologies du numérique d’Allistène (CERNA). Ella a participé à la rédaction du rapport sur l’éthique de la recherche en robotique, a dirigé celui sur l’éthique de la recherche en apprentissage machine et elle coordonne actuellement un rapport sur l’éthique de la recherche sur les agents conversationnels.

*Yann Ferguson* est docteur en sociologie à l’Icam (site de Toulouse) et chercheur associé au CERTOP (Université Toulouse-Jean Jaurès). Ses recherches portent sur les relations entre les hommes et les outils d’aide à la décision dans les environnements professionnels. Lauréat 2018 de la Fondation pour les sciences sociales pour ses travaux sur les mutations du travail liées à l’intelligence artificielle, il est considéré par *L’Usine Nouvelle* parmi « les 200 Français qui font avancer l’intelligence artificielle en France et dans le monde ».

*Frédéric Fürst* est maître de conférences en informatique à l’Université de Picardie-Jules Verne et membre du laboratoire MIS (Modélisation-information-systèmes). Il mène des recherches en intelligence artificielle sur la représentation des connaissances humaines dans les machines et le raisonnement automatique, mais aussi sur l’histoire et l’épistémologie de l’IA.

*Yu Li* est maître de conférences à l’Université de Picardie-Jules Verne (Laboratoire MIS). Titulaire d’un doctorat en informatique de l’Université de Compiègne et d’un master en informatique de l’Université des sciences et technologies de Huazhong, en Chine, ses recherches portent sur la philosophie de l’intelligence, la théorie des algorithmes, la théorie des problèmes non-déterministes (NP) et la résolution pratique de problèmes NP-difficiles.

*Aifang Ma* est doctorante au Centre d’études européennes et de politique comparée de Sciences Po Paris. Ses recherches portent sur les politiques numériques de la France, de la Chine et des États-Unis. Elle s’intéresse également à la régulation de l’internet, aux nouvelles technologies et aux conséquences sociales des avancées scientifiques et technologiques.

*Magalie Ochs* est enseignante-rechercheuse en informatique à Aix-Marseille Université, membre du Laboratoire d’informatique et des systèmes (LIS). Elle a travaillé dans différents laboratoires de recherche : au centre R&D d’Orange (Orange Labs), au laboratoire d’informatique de l’Université Paris 6, au National Institute of Informatics de Tokyo et au CNRS LTCI à Télécom ParisTech. Elle conduit depuis plusieurs années des travaux de recherche visant à intégrer une intelligence sociale et émotionnelle dans des systèmes interactifs, comme des robots ou des personnages virtuels, en utilisant des techniques d’intelligence artificielle.

*Anne-Laure Thessard* est philosophe spécialisée en éthique animale et éthique de l’intelligence artificielle. Ses recherches concernent les représentations et systèmes de valeur de la triade animalité/humanité/machinité. Elle est par ailleurs professeure de philosophie sur le campus parisien de l’EM Normandie en éthique animale et éthique du management et chargée de cours de philosophie en éthique animale à l’Université Paris Lumières pour la section « Nouveau Collège d’Études Politiques ».

*Charles Thibout* est chercheur associé à l’Institut de relations internationales et stratégiques (IRIS) et doctorant à l’Université Paris 1 Panthéon-Sorbonne (Centre européen de sociologie et de science politique, CESSP). Il est l’auteur de plusieurs articles et chapitres d’ouvrages académiques consacrés au rôle des technologies émergentes et des entreprises du numérique dans les relations internationales. Par ailleurs, il collabore régulièrement avec France Culture, en tant que rédacteur en chef de la matinale.

*Jean-Sébastien Vayre* est post-doctorant à l’Université de Côte d’Azur (UCA) et membre du GREDEG (UMR 7321). Il est docteur en sociologie. Il a réalisé une thèse au CERTOP (UMR 5044, dir. Franck Cochoy) sur la conception et les usages des technologies d’intelligence artificielle élaborées dans le secteur marchand. Il a ensuite effectué un post-doctorat à l’Institut Mines-Télécom Business School (IMT-BS) sur les enjeux organisationnels et humains associés au développement de l’industrie 4.0. Il a écrit plusieurs articles et chapitres d’ouvrages scientifiques sur la thématique de la numérisation des organisations et des marchés.

*JianMing Zhou* est chercheur indépendant. Ses recherches portent sur la comparaison des cultures chinoise et occidentale, la philosophie de l’intelligence, la théorie des algorithmes et la théorie des problèmes non-déterministes (NP).



# Introduction



# L'intelligence de l'intelligence artificielle

**Boris Barraud**

*7 novembre 2019. L'intelligence artificielle découvre que la Terre tourne autour du Soleil<sup>1</sup>.*

En des temps immémoriaux, Zeus confia à Épiméthée, fils de Japet et Clymène, le soin de distribuer aux espèces vivantes les qualités et les vertus en proportions équilibrées (aux uns les griffes et les crocs pour chasser, aux autres les sabots, les écailles ou les piquants pour se défendre). Épiméthée s'acquitta fort bien de sa tâche et tous les animaux se trouvèrent avec des forces propres leur permettant de jouer un rôle sur la grande scène de la vie. La nature pouvait prospérer et se pérenniser. Mais Épiméthée oublia

---

<sup>1</sup> Communiqué publié dans la revue *Nature* le 7 novembre 2019, indiquant que «les réseaux de neurones peuvent être utilisés pour découvrir des concepts physiques sans aucun connaissance préalable». L'information peut prêter à sourire, le principe de l'héliocentrisme étant connu depuis les travaux de Nicolas Copernic, exposés en 1533. Plus exactement, une IA a découvert seule que la Terre tourne autour du Soleil. Un réseau de neurones, surnommé «SciNet», fonctionnant à partir de données expérimentales, donc de simples observations objectives, a compris qu'il fallait placer le Soleil au centre du système solaire en fonction des mouvements de l'étoile et de la planète mars vus depuis la Terre. Selon ses auteurs, la portée de l'étude en question, qui doit paraître plus tard dans les *Physical Review Letters*, est que de telles IA pourraient à l'avenir découvrir de nouvelles lois de la physique, notamment en mécanique quantique qui connaît aujourd'hui beaucoup de contradictions, avant les hommes — mais en réalité, ces IA étant des outils fabriqués par les hommes, ces découvertes hypothétiques seraient plus exactement réalisées par des hommes au moyen de leurs outils les plus perfectionnés.

l'homme<sup>2</sup>. Toutes les capacités et tous les attributs étaient épuisés, il ne restait rien pour l'être humain. Ému de voir ce dernier nu et désarmé, Prométhée, le frère d'Épiméthée, déroba alors à Héphaïstos et Athéna le feu créateur de tous les arts et l'offrit au plus misérable des animaux. Et ce larcin lui valut une lourde punition de la part des dieux, car le monde équilibré conçu par Épiméthée venait de basculer dans la tyrannie : une espèce unique, parce qu'elle serait la seule à maîtriser la technique, allait régner sur l'univers et triompher de tous les autres animaux.

Le mythe d'Épiméthée, raconté par Platon dans le *Protagoras*<sup>3</sup>, est riche d'enseignements. Bernard Stiegler y voit un rappel de l'état de détresse ou, comme disent les biologistes, de prématuration dans lequel se trouve initialement l'homme, mais ce don divin qu'est la technique lui permet de remédier à ses carences et à sa fragilité originelles<sup>4</sup>. En raison de l'intervention providentielle de Prométhée, la faiblesse congénitale de l'espèce humaine s'est transformée en supériorité écrasante, en puissance étouffante, en pouvoir destructeur — pour les autres êtres vivants, mais peut-être aussi pour lui-même.

Les outils des hommes prolongent leurs corps et expriment leurs esprits. Comme l'enseigne l'anthropologue André Leroi-Gourhan<sup>5</sup>, chaque instrument extériorise une fonction bien déterminée. Furent concernés tout d'abord les usages matériels les plus simples et durs (avec le bâton ou le marteau). Puis les apports de la technique devinrent de plus en plus sophistiqués et révolutionnaires (avec les engins énergétiques décuplant ou remplaçant de façon beaucoup plus efficace la force brute des muscles : brouette, poulie, machine à vapeur, chemin de fer, automobile etc.). Plus tard, vinrent les instruments ou machines sensoriels affinant et développant les sens et la perception (lunette, microscope, télégraphe optique, photographie, cinéma etc.). Enfin, apparurent les systèmes informationnels, la dernière et la plus actuelle des générations de la technique. Le but est cette fois de démultiplier et même subroger les fonctions et capacités intellectuelles des hommes grâce aux livres, calculettes, ordinateurs, disques durs, réseaux électroniques et informatiques etc. Mais déjà la décomposition alphabétique, la logique, la rhétorique ou les arts de la mémoire ont constitué autant d'avancées techniques immatérielles, toujours utilisées aujourd'hui, permettant de seconder et améliorer l'effort millénaire d'analyse et de synthèse de données ou de connaissances, façonnant en profondeur les habitudes intellectuelles de « penser/classer »<sup>6</sup>.

<sup>2</sup> En grec ancien, « Ἐπιμηθεύς / Epimétheús » signifie « qui réfléchit après coup ».

<sup>3</sup> Platon, *Protagoras*, vers 390 av. J.-C., 320d-322a.

<sup>4</sup> B. Stiegler, *La technique et le temps – La faute d'Épiméthée*, Galilée, 1994.

<sup>5</sup> A. Leroi-Gourhan, *Le geste et la parole*, Albin Michel, 1964.

<sup>6</sup> G. Perec, *Penser/Classer*, Hachette, 1985.

Les organes technologiques et prothèses mécaniques ou électroniques ont accompagné de près l'hominisation (processus par lequel les primates se sont transformés en hommes) et, surtout, l'humanisation (processus par lequel les hommes préhistoriques sont devenus des hommes historiques, les prémodernes des modernes, les sauvages des civilisés, les obscurs des lumières, les individus des citoyens, les hordes barbares des sociétés paisibles et les regroupements anarchiques des organisations étatiques). Mais se pourrait-il qu'un jour l'association progressiste des hommes et de leurs machines atteigne son point culminant, ne permette plus d'avancées de la civilisation, entraîne des régressions, stimule des phénomènes de déshominisation et de déshumanisation car l'appareillage technologique serait devenu par trop envahissant, intimidant, inhibiteur et sclérosant. Et ce jour pourrait-il appartenir déjà au passé ?

Quel qu'en soit le sens, l'évolution biologique est donc intimement liée à l'évolution technique. La biosphère est fortement, et de plus en plus, dépendante de la technosphère. L'influence de l'homme sur la technique est quasi-totale ; mais l'influence de la technique sur l'homme est également très importante. Les relations entre l'humain et l'outil, l'être et la machine, le vivant et le robot, le naturel et l'artificiel jouent bien le rôle principal dans le grand film de l'humanité, de l'histoire, des sociétés et des cultures. Les uns créent et utilisent les autres qui, en retour, rétroagissent sur eux, les influencent, modèlent leurs pensées et leurs actes. On peut de moins en moins séparer les êtres humains et les sociétés de leur environnement technique et technologique. Des premiers silex taillés par l'homme préhistorique à l'invention de l'écriture, de Gutenberg à la numérisation massive des données, la longue histoire des techniques et des technologies, aussi ancienne que l'histoire elle-même — et qui en est d'ailleurs à l'origine —, se situe au cœur de l'évolution humaine. Les instruments que l'homme invente le transforment, modifient ses activités personnelles et professionnelles, ses moyens de communication, ses manières de consommer, ses modes de transport ou encore ses façons d'observer le monde et de le penser. Et plus le temps passe, plus l'intimité problématique entre l'homme et la technologie devient forte, à tel point qu'il arrive qu'on ne sache plus si l'on se trouve face à l'un ou à l'autre — et alors le « test de Turing » est passé avec succès.

Tel est notamment le cas lorsqu'on échange avec des intelligences artificielles, avec des algorithmes, avec des robots dont les réponses atteignent un tel degré de pertinence qu'elles semblent avoir été produites par un cerveau naturel.

L'histoire des techniques est rythmée par le franchissement successif de différents seuils : les cliques d'irréversibilité. L'invention de l'imprimerie par Gutenberg a ainsi signifié la disparition des moines copistes du Moyen-Âge, la création des télégraphes optiques et des chemins de fer a rendu obsolètes les chevaux, les attelages et les coches, l'introduction de l'informatique et des

calculatrices électroniques a sonné le glas des bouliers et des opérations gribouillées sur le papier, l'apparition des fichiers numériques et du *streaming* a conduit à remiser au placard les lecteurs CD et les chaînes hi-fi, qui eux-mêmes avaient entraîné la disparition des gramophones d'antan et la marginalisation des disques vinyles. Nul doute que l'intelligence artificielle constitue un nouveau — et très puissant — cliquet d'irréversibilité.

D'ailleurs, les révolutions de l'intelligence artificielle n'en sont peut-être qu'à leurs prémices. Nul ne saurait dire quelles en seront les conséquences dans 20, 50 ou 100 ans. C'est ce qui rend le sujet passionnant, restant encore largement à explorer, à défricher, à préciser. Et c'est en même temps ce qui suscite une part de fantasmes et de craintes. Les futurologues, mais aussi les spécialistes des nouvelles technologies, l'affirment : l'intelligence artificielle va modifier, souvent radicalement, beaucoup de pratiques. Leurs divergences sont en revanche grandes quant au tour que prendront ces pratiques chamboulées. Or les chercheurs seront les premiers concernés : leurs objets d'étude vont largement évoluer ; et leurs moyens de les étudier vont fortement se renouveler. *Dans toutes ses dimensions*, l'intelligence artificielle est un sujet d'actualité et d'avenir pour les scientifiques de toutes les disciplines.

Les dictionnaires de la langue française définissent l'intelligence comme « la fonction mentale d'organisation du réel en pensées chez l'être humain et en actes chez l'être humain et l'animal »<sup>7</sup>. Il s'agit d'un « ensemble de fonctions psychiques et psycho-physiologiques concourant à la connaissance, à la compréhension de la nature des choses et de la signification des faits »<sup>8</sup>. Ainsi entendue, l'intelligence de l'intelligence artificielle est nulle ou très faible — on y reviendra.

Mais l'intelligence est aussi « l'acte d'appréhender quelque chose par l'intelligence, de le comprendre avec aisance, d'en avoir la connaissance approfondie »<sup>9</sup>. Sous cet angle, l'intelligence de l'intelligence artificielle est encore balbutiante et très incomplète, ce qui invite à y contribuer, à participer à son approfondissement et à sa précision. L'objet de ce livre est ainsi de proposer une contribution à l'intelligence de l'intelligence artificielle en regroupant les textes de certains des meilleurs spécialistes de l'IA. Il existe un accord parmi les experts sur le fait que l'intelligence artificielle s'est déployée de façon pragmatique, à mesure que des recettes « qui marchent » ont été découvertes, mais on manque encore de concepts et de théories pour la comprendre, la cerner et l'expliquer<sup>10</sup>. En ce sens, le besoin d'intelligence de l'intelligence artificielle serait bel et bien réel. Jusqu'à la définition de

---

<sup>7</sup> V° « Intelligence », in *Trésor de la langue française*.

<sup>8</sup> *Ibid.*

<sup>9</sup> *Ibid.*

<sup>10</sup> La situation est ainsi inverse de celle de l'informatique qui tente de mettre en œuvre les concepts définis par des théoriciens parfois fort anciens.

l'intelligence artificielle, à base de critères et sous-critères permettant de la reconnaître et de délimiter son périmètre, semble encore faire défaut.

L'intelligence artificielle paraît lointaine, justement car on saisit mal ce que cette expression *a priori* oxymorique désigne précisément. L'IA touche pourtant chacun dans son quotidien et dans son intimité. Elle présente également un côté magique, parce qu'elle produit des résultats impressionnantes sans qu'il soit possible, pour la plupart des observateurs non mathématiciens et non informaticiens, de comprendre comment elle procède, quels sont ses mécanismes. Une définition de base et quelques premières explications sont donc nécessaires, à titre d'introduction générale du présent ouvrage.

Une intelligence artificielle est une technologie capable de produire des résultats similaires à ceux issus du cerveau humain. Il s'agit d'un outil informatique<sup>11</sup> qui effectue des actions ou exécute des tâches qui, il y a peu, étaient le propre des êtres vivants, humains ou animaux<sup>12</sup>. Cet outil repose notamment sur des algorithmes, c'est-à-dire des suites de formules mathématiques et de traitements statistiques. Il fonctionne avec des « entrées » (les données initiales) et aboutit à des « sorties » (les résultats) en suivant différentes étapes qui requièrent des calculs, des opérations logiques, des comparaisons ou des analogies. L'intelligence artificielle désigne donc les dispositifs technologiques visant à simuler et, *in fine*, remplacer l'intelligence naturelle, cherchant à reproduire les capacités de l'homme et de l'animal à percevoir, discerner, comprendre, apprendre, raisonner, calculer, mémoriser, comparer, choisir etc. Mais ce que l'IA cherche à imiter, ce n'est pas le fonctionnement de l'intelligence biologique mais uniquement les résultats auxquels elle parvient — le rapport est un peu le même que celui de l'avion avec le vol des oiseaux ou celui de la voiture avec la course à pied. L'ambition des chercheurs en la matière n'est pas et n'a jamais été de fabriquer des cerveaux technologiques se rapprochant au plus près des cerveaux naturels, de la conscience et de la pensée humaines ; elle est uniquement de concevoir des outils informatiques capables de produire les résultats les plus pertinents, essentiellement grâce à du traitement statistique de données.

« Intelligence » dérive du latin « *intelligentem* » ou « *intelligere* », signifiant comprendre, discerner. « *Legere* » veut dire « choisir » et le préfixe « *inter* » se traduit par « entre ». Est donc intelligent ce qui a la capacité de

---

<sup>11</sup> L'informatique désigne le traitement rationnel, notamment par des machines automatiques, de l'information.

<sup>12</sup> Marvin Minsky, l'un des scientifiques qui ont créé l'expression « intelligence artificielle », définissait l'IA comme « la construction de programmes informatiques qui s'adonnent à des tâches pour l'instant accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique ».

choisir entre différentes alternatives et, pour ce faire, de relier les choses entre elles afin de mieux les saisir. Les hommes et les animaux possèdent certainement cette faculté. L'aspect « intelligent » de l'intelligence artificielle correspond dès lors au but et à la faculté d'imiter le comportement humain ou animal, tandis que l'aspect « artificiel » se rapporte à l'usage d'ordinateurs et autres procédés électroniques et technologiques. L'intelligence artificielle emprunte ainsi à l'informatique, à la logique, aux mathématiques, à l'électronique, aux sciences cognitives, aux sciences de la communication et même à l'éthologie et à la neurobiologie — avec les techniques d'apprentissage profond inspirées des réseaux neuronaux interconnectés, que l'on présentera plus tard dans cette introduction.

Cependant, si une IA peut être capable de restructurer son système de connaissances ou de perception des données à l'aune de nouvelles informations reçues, aucun ordinateur, aucun algorithme ni aucun réseau de neurones ne pense ni n'est intelligent à proprement parler — sa capacité d'apprentissage et de décision est programmée ou, du moins, automatisée, donnant seulement une illusion d'intelligence, au mieux une intelligence faible car très mécanique. Une IA ne peut en effet décider en toute autonomie de ses buts ni concevoir de nouvelles représentations du monde qui la guideraient dans ses raisonnements, qui d'ailleurs n'existent pas puisqu'il ne s'agit que de réflexes et d'automatismes. C'est pourquoi l'appellation « intelligence artificielle » est largement excessive. *De facto*, il ne s'agit guère d'intelligence mais seulement de calculs (certes savants), de statistiques et d'opérations informatiques.

Parler d'« intelligence artificielle » est donc un abus de langage. « Intelligence artificielle » vient de l'anglais « *artificial intelligence* ». Cette expression désigne dans la langue de Shakespeare la simple gestion de données, le traitement de l'information, et non l'intelligence artificielle telle qu'on la comprend en français — qui se traduirait plutôt par « *artificial cleverness* », désignant le sens commun, l'empathie, l'arbitraire éclairé, la faculté de comprendre sa propre intelligence et d'en avoir conscience. En somme, il n'est pas déceptif, en anglais, d'évoquer une « *artificial intelligence* » afin de désigner ces nouvelles technologies. Mais il est maladroit, en français, de parler d'« intelligence artificielle ».

À l'heure actuelle, les robots ne raisonnent pas, ne croient en rien, ne prennent guère de décisions arbitraires, n'ont pas d'intentions ni d'attentions, sont incapables de se donner des buts ou des objectifs nouveaux, n'organisent pas le réel en pensées, sont incapables d'élaborer un concept ou d'avoir une idée, n'ont pas de culture et encore moins de culture générale, n'ont pas conscience d'eux-mêmes et ne possèdent aucun instinct de survie — autant de marques de l'intelligence. Les hommes se caractérisent par l'esprit, par la liberté et par la capacité à s'ouvrir à l'inconnu. Les robots en sont tout le contraire. Pourtant, parler d'« intelligence artificielle » revient dans une large

mesure à parler d'« humanité artificielle ». « Intelligence artificielle » est un anthropomorphisme. On humanise, on vitalise ce qui ne sont toujours que des objets animés, des automates, des supercalculateurs<sup>13</sup>. Peut-être vaudrait-il mieux évoquer l'« apprentissage automatique », des instruments « statistiques » ou « prédictifs » ou bien des « agents autonomes » ou « adaptatifs ». L'IA examine aujourd'hui un scanner mieux qu'un radiologue, établit un diagnostic en cancérologie mieux qu'un cancérologue, conduit une automobile mieux qu'un chauffeur, mais elle ne sait pas ce qu'est une radiographie, ce qu'est un cancer, ce qu'est une automobile ni ce en quoi consiste la conduite.

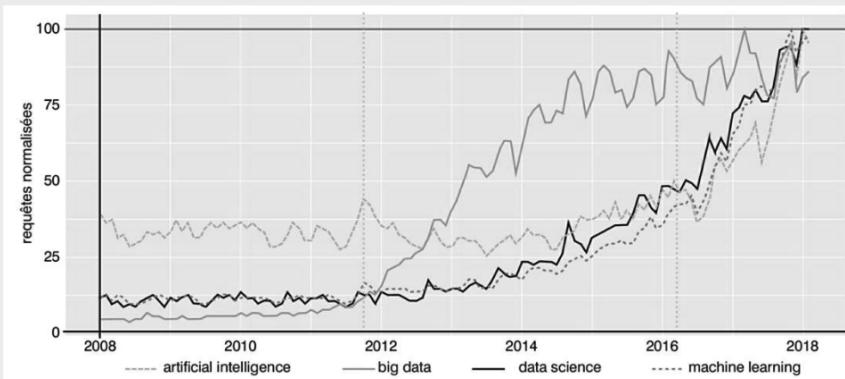
L'expression « intelligence artificielle » (ou « IA ») s'étant partout imposée, tant dans les débats grand public que parmi les discussions de spécialistes — et même comme nom d'un nouveau domaine de recherches scientifiques —, il est cependant difficile de désigner l'objet de ce livre par d'autres mots, sous peine de ne pas se faire comprendre. Par ailleurs, comme il s'agit d'une expression valise, désignant des technologies disparates et variées, il serait logique de se référer plutôt aux « intelligences artificielles » (au pluriel). Mais l'usage privilégie le singulier dès lors qu'il s'agit de la notion générique d'intelligence artificielle.

La mode de l'« intelligence artificielle » et de son sigle « IA » est récente, comme en témoignent les statistiques de requêtes sur le plus célèbre moteur de recherche du web. Cette mode n'en est pas moins forte et, si l'on veut parler de l'IA, on ne peut que parler d'« IA ».

---

<sup>13</sup> Yann Le Cun estime toutefois que l'IA pourrait bientôt exprimer des sentiments et des émotions. Mais il précise bien que ce ne serait que parce qu'il deviendrait possible qu'un cerveau artificiel calcule la joie, la peur ou la tristesse. Cependant, de telles joie, peur ou tristesse calculées et non ressenties ou éprouvées en seraient-elles vraiment ?

## RECHERCHES INTERNET SUR LES SUJETS RELIÉS À L'INTELLIGENCE ARTIFICIELLE



— Source: d'après Google Trends. Les données sont normalisées. Pour chaque requête, la normalisation est faite en pourcentage du nombre maximal de requêtes observées sur la période.

L. Galiana, « Les conséquences économiques de l'intelligence artificielle », *Idées économiques et sociales* 2018, n° 192, p. 27

C'est en 1956 que les mathématiciens américains John McCarthy, Marvin Minsky, Claude Shannon et Nathaniel Rochester organisèrent, au Dartmouth College, dans l'État du New Hampshire, un séminaire sur les « machines pensantes » au cours duquel ils proposèrent pour la première fois de parler d'« *artificial intelligence* »<sup>14</sup>. Un nouveau domaine d'enseignement et de recherche scientifiques était né, dont l'objectif, longtemps et aujourd'hui encore perçu comme farfelu ou irréaliste, est de réaliser des machines capables d'apprendre — notamment de leurs erreurs —, de comprendre leur environnement et le monde en général, de s'adapter aux changements, d'anticiper et de prévoir le futur, d'agir afin d'améliorer la condition humaine. Plus cet objectif initial sera atteint et moins il sera déceptif de se référer à

<sup>14</sup> On attribue également la paternité, si ce n'est de l'expression, de la pensée de l'intelligence artificielle au mathématicien britannique Alan Turing, qui s'y est intéressé dès 1950 à l'occasion de ses travaux sur le « jeu de l'imitation » (ou « test de Turing ») : ce test consiste à mettre un humain en confrontation verbale à l'aveugle avec un ordinateur et un autre humain ; si cette personne n'est pas capable de dire lequel de ses interlocuteurs est un ordinateur, alors l'ordinateur a passé avec succès le test (A. Turing, « Computing Machinery and Intelligence », *Mind* oct. 1950).

l'« intelligence » artificielle. On pourra alors comparer à raison cette dernière et l'intelligence biologique.

Pour l'heure, néanmoins, ce qu'on appelle « intelligence artificielle » demeure un simple produit de l'intelligence humaine, sa création et son reflet, globalement immensément inférieur à elle. Pour devenir une réalité, l'intelligence artificielle devra gagner en autonomie, en libre arbitre, se détachant des hommes qui l'ont faite naître. Il lui faudra devenir une intelligence artificielle générale, capable de comprendre le fonctionnement du monde dans sa globalité par simple observation, comme le fait un enfant durant les premières années de sa vie. Or il lui sera difficile d'y parvenir<sup>15</sup>. On imagine difficilement l'intelligence artificielle rejoindre et dépasser un jour l'intelligence naturelle autrement que ponctuellement, concernant des tâches très précises susceptibles de reposer sur des traitements statistiques. Beaucoup de décisions supposent des choix arbitraires plutôt que la simple poursuite de tendances statistiques.

Lorsque les chercheurs du Dartmouth College créèrent l'expression « *artificial intelligence* », il s'agissait d'ailleurs surtout d'une démarche marketing, presque un mensonge volontaire : recourir à une expression sensationnelle afin d'obtenir des crédits de recherche de la part des pouvoirs publics américains<sup>16</sup>. Le concept d'intelligence s'attache à un fond ontologiquement biologique et l'intelligence « artificielle » ne peut pas être intelligente, par définition. Mais cela n'empêche pas l'IA de rendre bien des services tout en posant de nombreux défis à l'intelligence humaine, en premier lieu celui de son intelligence, c'est-à-dire celui de sa compréhension, de son explication et de son analyse critique.

L'intelligence de l'intelligence artificielle suppose de combiner quatre efforts cognitifs afin de mieux appréhender cette méta- et méga-révolution (source de révolutions nombreuses et variées) :

- accepter que l'IA n'est qu'un outil développé par et pour les hommes — point de vue humain (*I*) ;
- s'intéresser aux rouages et au fonctionnement de la technologie — point de vue technique (*II*) ;
- se concentrer sur les conséquences civilisationnelles de l'IA — point de vue social, politique et économique (*III*) ;
- donner la parole aux chercheurs et l'écouter — point de vue scientifique (*IV*).

---

<sup>15</sup> Bien qu'on commence à appliquer le *machine learning* à lui-même, donc à entraîner des ordinateurs à concevoir des réseaux de neurones, si bien que des intelligences artificielles pourraient un jour créer d'autres intelligences artificielles.

<sup>16</sup> Un travestissement de la réalité qui rappelle celui d'Éric Le Rouge lorsqu'il emmena ses hommes au Groenland en leur faisant croire que c'était un « pays vert » (traduction de « *groenland* »), alors que c'était un pays blanc, recouvert de neige et de glace. Mais le nom « Groenland » est resté.

## I. Une problématique humaine : l'intelligence de la machine au service de l'intelligence de l'homme

Des robots mus par des IA en venant à défier et affronter les hommes est de l'ordre du pur fantasme. Présenter le rapport entre les uns et les autres sous les traits d'un duel dont l'un des protagonistes devrait triompher de l'autre ne paraît guère réaliste. Les humains produisent des robots qui ne sont et ne seront que leurs outils ; et qui, comme tous outils, ne sont ni bons ni mauvais en soi car soumis aux destinations qu'on leur attribue, aux usages qu'on en fait<sup>17</sup>. Il n'y a et n'y aura pas plus de compétition entre l'homme et l'IA qu'entre le pianiste et le piano ou entre le maçon et la truelle. Et puis il serait difficile de dire qu'une espèce aussi proliférante que l'espèce humaine pourrait s'effacer et perdre la mainmise sur le monde, surtout au profit de ses propres technologies.

En 2014, l'astrophysicien Stephen Hawking affirmait pourtant, dans une interview à la BBC, que « le développement d'une intelligence artificielle complète pourrait provoquer la fin de l'humanité »<sup>18</sup>. Il s'inquiétait d'éventuels périls liés à l'avènement de robots qui surpasseraient des êtres humains limités par la lenteur de l'évolution biologique. Et beaucoup, comme Bill Gates ou Elon Musk — qui a créé une fondation pour mesurer et prévenir la dangerosité des IA, laquelle serait le plus grand danger ayant jamais pesé sur l'humanité —, entretiennent une psychose effrayante autour de la prise de pouvoir de machines toutes-puissantes créées par l'homme puis lui échappant<sup>19</sup>. C'est un lieu commun de faire le procès des nouvelles

---

<sup>17</sup> C'est là l'occasion de rappeler les lois énoncées par l'écrivain de science fiction Isaac Asimov (*Cercle vicieux*, Runaround, 1942) :

« 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 ».

<sup>18</sup> L'homme ne représente que 0,01 % des êtres vivants mais est à l'origine de 99 % des destructions dans la biosphère. Pourrait-il soudain perdre ce monopole de la destruction et se retrouver lui-même détruit par des êtres supérieurs ? Ce scénario, aujourd'hui, relève entièrement de la science fiction.

<sup>19</sup> Des prolongements universitaires de ces questionnements existent. À l'université de Cambridge, le Centre pour l'étude des risques existentiels (*Centre for the Study of Existential Risk*) a été créé afin d'étudier les risques et menaces que les technologies feraient peser sur le futur. Quant à l'Institut pour l'avenir de l'humanité (*Future of Humanity Institute*) de l'université d'Oxford, il se consacre aux questions relatives à l'avenir de l'humanité en donnant une grande place aux nouvelles technologies.

technologies, presque un « marronnier », revenant périodiquement sur le devant de la scène depuis des siècles. Il y a toujours eu des prophètes, apocalyptiques ou intégrés<sup>20</sup>, pour annoncer les ruptures radicales et lourdes de conséquences que la rationalisation scientifique des machines provoquerait. Mais ces déclamations tapageuses sont souvent biaisées ou naïvement critiques. Il est difficile d'imaginer qu'un jour la machine puisse en venir à penser comme l'homme, contre l'homme ou mieux et plus vite que l'homme, si bien que l'être humain, de maître et possesseur du monde sous toutes ses formes, se retrouverait en position d'esclave ou de jouet de robots qui ne seraient plus les fruits de son art mais auraient la capacité de se reproduire et de se perfectionner seuls. Tout cela appartient à la « science » fiction.

Cette « science » fiction diffuse différentes légendes parmi le grand public, notamment celle selon laquelle l'IA devrait se comporter à l'image des hommes et présenter les mêmes qualités et les mêmes défauts démultipliés, les mêmes pulsions et le même instinct renforcés — on fabrique des robots humanoïdes, mais leur « intelligence » et leurs « cerveaux » ne partagent quasiment rien avec l'intelligence et les cerveaux humains. Il y a également le mythe selon lequel il existerait un secret scientifique et technologique de l'intelligence artificielle qui resterait à découvrir, si bien que celui qui le découvrirait pourrait dominer le monde — alors que l'intelligence artificielle progresse par petits (et grands) pas, par étapes, en tâtonnant, incluant en particulier des découvertes dont les tenants et les aboutissants (non seulement les aboutissants) sont rendus publics, permettant à chacun de les réutiliser librement et de les améliorer. Le problème qui se pose plus sérieusement, et déjà actuellement, n'est pas celui de machines asservissant les humains mais celui d'humains asservissant d'autres humains grâce aux machines — un problème politique, juridique, social et économique. Là où il faut et faudra toujours de l'éthique, ce n'est pas dans l'intelligence artificielle mais dans l'intelligence humaine<sup>21</sup>.

Les projets d'intelligences artificielles doivent donc s'orienter vers des programmes réfléchis et contrôlés, non abandonnés à des intérêts économiques, mercantiles et vénaux, livrés à la seule quête de nouveaux marchés et de profits croissants, au besoin en manipulant, en enfermant, en

---

<sup>20</sup> U. Eco, *La Guerre du faux*, Grasset, 1985.

<sup>21</sup> En témoigne l'invention, en 2017, par des chercheurs de l'université américaine de Stanford, d'un algorithme capable de découvrir l'orientation sexuelle des personnes en fonction des traits de leurs visages, alors que tant de programmes de recherche en intelligence artificielle plus utiles pourraient être menés. En témoigne, également, l'intelligence artificielle, créée par Microsoft en 2016, qui devait évoluer sur Twitter aux côtés des internautes humains. En quelques heures, ces derniers l'ont rendue complètement misogynie, homophobe, raciste et même nazie, si bien qu'elle a dû être retirée en catastrophe.

limitant au maximum le sens critique et les valeurs morales et en réduisant le monde à deux ensembles : celui des produits et celui des consommateurs. Il importe d'écouter avec du recul et une oreille critique les discours hypocrites ou naïfs de certaines multinationales de l'IA affirmant que leur but ne serait pas de générer des profits et des monopoles mais de permettre à l'humanité de progresser et d'entrer dans un nouvel âge.

Pourtant, la finalité unique ou ultime des projets d'intelligences artificielles devrait effectivement être de faciliter la vie des hommes et le bon fonctionnement des sociétés, de les protéger des dangers qui les guettent — la maladie, la corruption, la manipulation, l'aveuglement etc. Certains de ces projets permettront de sauver de nombreuses vies, notamment dans le domaine des transports ou dans celui de la santé — avec une circulation automobile bien moins accidentogène et des diagnostics médicaux beaucoup plus rapides, précis et fiables<sup>22</sup>. Neutre en soi, la technologie peut être mise à profit par des hommes qui le souhaitent pour contaminer, corrompre ou manipuler — par exemple avec les *deep fakes*, qui seraient la première de toutes les menaces liées à l'IA<sup>23</sup>. Son utilisation en tant qu'outil de propagande ou de désinformation peut être désastreuse. Comme l'internet et comme de nombreuses révolutions technologiques, l'IA est donc un moyen pour les personnes bien intentionnées comme pour les personnes mal intentionnées d'arriver plus facilement à leurs fins.

Dans une perspective dite « transhumaniste », l'intelligence artificielle peut soutenir l'humain dans sa quête de puissance et de pouvoir, donc d'actions rationnelles, de décisions éclairées et pertinentes, de maîtrise et de domination. Elle peut augmenter ses capacités cognitives. Mais elle devrait demeurer son instrument et ne pas le concurrencer<sup>24</sup>. D'ailleurs, la volonté de

---

<sup>22</sup> Dans le diagnostic de certains cancers, l'IA réalise 99 % de bons diagnostics, contre 60 % à 80 % parmi les médecins humains. Ceux-ci sont notamment dépassés pour des raisons physiologiques : par exemple, une rétine humaine, d'un radiologue, distingue 15 niveaux de gris, contre un milliard de niveaux de gris différents perçus par l'IA.

<sup>23</sup> Les *deep fakes* sont de faux documents, notamment vidéos avec des imitations toujours plus réalistes du comportement et de la voix, permettant de travestir la réalité, de faire dire ou faire faire des choses à des individus qui ne les ont jamais dites ni faites, de mettre le visage de quelqu'un sur le corps de quelqu'un d'autre, ou placer une personne dans un contexte dans lequel elle ne s'est jamais trouvée. Ces *deep fakes* sont évidemment très dangereux, car quand on dit que quelqu'un a dit quelque chose et qu'en plus on en montre la preuve, il est difficile de ne pas y croire. Parmi les scientifiques, ceux qui œuvrent à la création des *deep fakes* sont aussi ceux qui recherchent les moyens de les détecter — avec des algorithmes reconnaissant les vidéos d'origine sur le web, évaluant la cohérence du flux sanguin entre le corps et la tête ou mesurant le caractère naturel du clignement d'œil.

<sup>24</sup> Les visions d'une superintelligence informatique vouée à prendre le pouvoir reposent principalement sur la loi de Moore — décrite au milieu des années 1960 par

domination des uns sur les autres est davantage liée à la testostérone qu'à l'intelligence ; et l'IA ne produit pas de testostérone. L'instinct et les pulsions de domination sont aussi liés au fait que les hommes sont des animaux sociaux qui ont besoin d'interagir avec leurs semblables, de les influencer, de s'imposer à eux. Cela n'existe pas chez les animaux non sociaux tels que les orangs-outans. Or l'IA n'a pas d'instinct grégaire, elle ressemble, sur ce plan, plus à l'orang-outan qu'à l'homme.

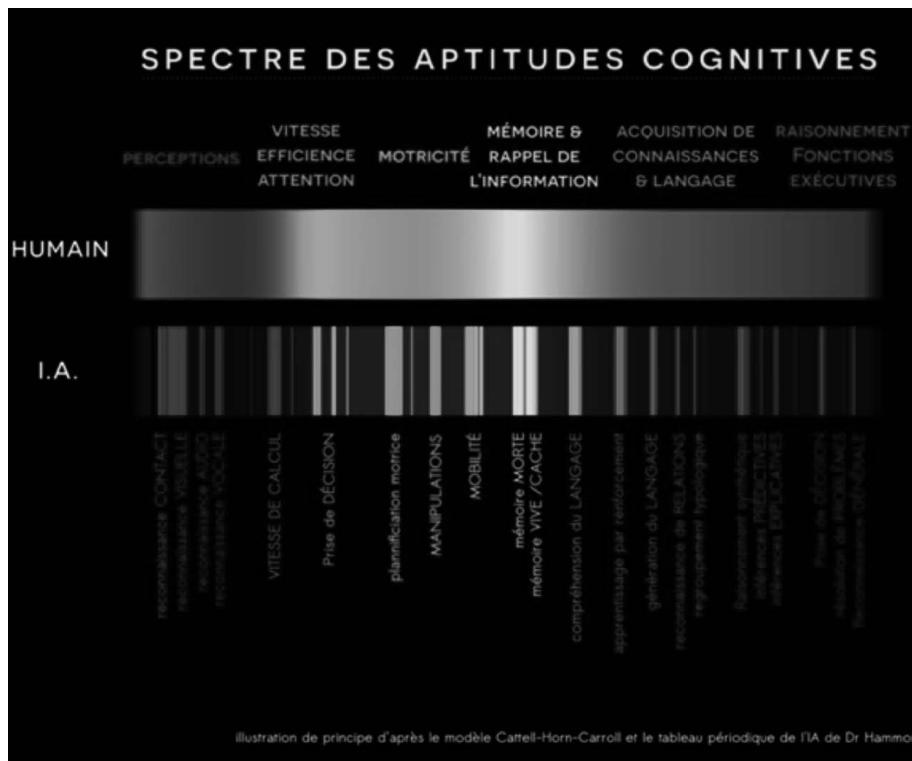
Idéalement, si les humains créent des formes d'intelligence artificielle, ce ne devrait être qu'afin de progresser, de rendre plus aisés nombre de leurs actions individuelles et collectives. L'invention de l'intelligence artificielle se positionne ainsi dans la continuité des inventions du langage, de l'écriture, de l'imprimerie, de la roue, de la machine à vapeur, du moteur à explosion, du système électrique etc. L'homme, conscient des limites de son intelligence, pourtant sans conteste la plus développée de tout le règne animal, se dote de prothèses telles que des bibliothèques, des machines à calculer, des ordinateurs ou des algorithmes. Il parvient ainsi à conserver en mémoire et à trier de grandes masses d'informations ou à réaliser des calculs hyper-complexes en quelques instants.

---

le cofondateur d'Intel Gordon E. Moore, établissant que la puissance des processeurs suit une évolution exponentielle à travers le temps, doublant tous les dix-huit mois. Mais les progrès technologiques vont obligatoirement ralentir et, pour certains, s'arrêter. Si tout n'est pas possible éthiquement et moralement, il faut ajouter que tout n'est pas possible physiquement, scientifiquement, technologiquement.

## *Dans toutes ses dimensions*

Cependant, si l'IA est capable de réaliser des prouesses qu'aucun humain ne saurait accomplir, elle ne sait pas faire beaucoup de choses simples pour un homme car relevant du bon sens ou du sens commun, de l'intuition, de l'émotion, de l'empathie, de la créativité etc. L'illustration suivante montre combien les aptitudes cognitives de l'IA sont partielles, bien que parfois très développées, quand les capacités cognitives de l'humain sont totales, transversales, mais limitées par nature dans leurs performances :



On considère classiquement, depuis les travaux de psychologie cognitive d'Howard Gardner, qu'il existerait huit formes d'intelligence<sup>25</sup>. Certaines (visuelle-spatiale, logique-mathématique ou verbale-linguistique) semblent plus déshumanisables que d'autres (interpersonnelle, intrapersonnelle ou naturaliste-écologiste).

## LES INTELLIGENCES MULTIPLES

d'après la théorie de Howard Gardner

<b>Intelligence verbo-linguistique</b>  <b>ABC</b> Capacité à penser, saisir et exprimer des idées. Exemples : écrivains, poètes, traducteurs, interprètes, journalistes, avocats, ...	 <b>LA MUSIQUE</b> Capacité à reconnaître, interpréter, et créer des musiques, rythmes... Exemples : compositeurs, chanteurs, musiciens, professeurs de musique, accordieurs, ...
<b>Intelligence logico-mathématique</b>  <b>123</b> Capacité à raisonner, calculer, compter, résoudre des problèmes. Exemples : scientifiques, ingénieurs, savants, mathématiciens, comptables ...	 <b>LES AUTRES</b> Capacité à agir et réagir avec les autres, aider, collaborer, partager. Exemples : commerçants, politiciens, enseignants, managers, chef d'équipe, ...
<b>Intelligence visuelle / spatiale</b>  <b>LES IMAGES</b> Capacité à percevoir et se représenter le monde. Exemples : peintres, photographes, architectes, pilotes, dentistes, ...	 <b>SOI</b> Capacité à se connaître, à s'introspecter, à être intuitif. Exemples : chercheurs, entrepreneurs, romanciers ...
<b>Intelligence corporelle-kinesthésique</b>  <b>LE CORPS</b> Capacité à utiliser son corps, à s'exprimer physiquement. Exemples : danseur, athlète, artisans, chirurgiens, pompiers ...	 <b>LA NATURE</b> Capacité à observer, reconnaître, et classifier la nature. Exemples : explorateurs, botanistes, archéologues, zoologues, fermiers, ...

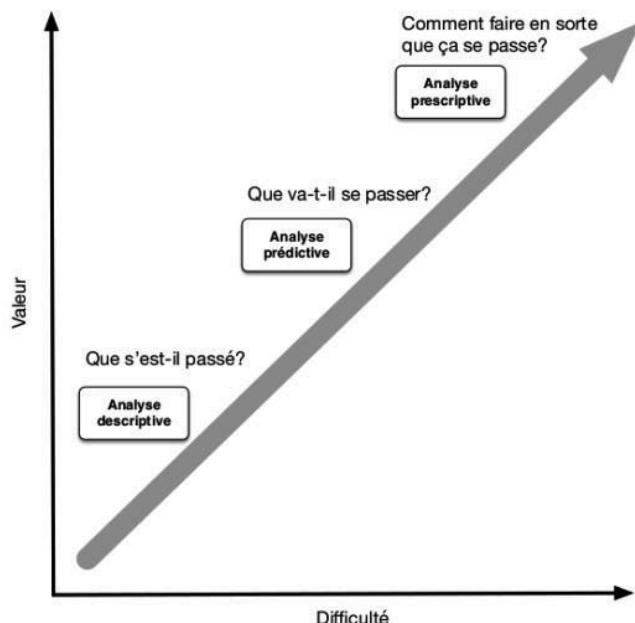
Illustrations : Céline Cidère / www.pour

C'est la symbiose de deux formes d'intelligence — ou, dans un cercle vertueux, augmentation de l'intelligence par ses réalisations les plus perfectionnées — qui permet le progrès des savoirs et des connaissances, ainsi que des idées. Et les chercheurs comptent au nombre de ceux qui en retirent les plus grands bénéfices.

<sup>25</sup> H. Gardner, *Les formes de l'intelligence*, Odile Jacob, 1997.

L'intelligence artificielle a vocation à servir l'intelligence humaine en éclairant et donc facilitant ses prises de décisions. Pour ce faire, elle doit lui fournir des informations à haute valeur ajoutée relevant de l'analyse descriptive consistant à trier et croiser les informations. Mais son caractère révolutionnaire tient surtout aux analyses prédictives et même prescriptives dont elle est capable, lesquelles, qualitativement, sont incomparables aux simples analyses descriptives :

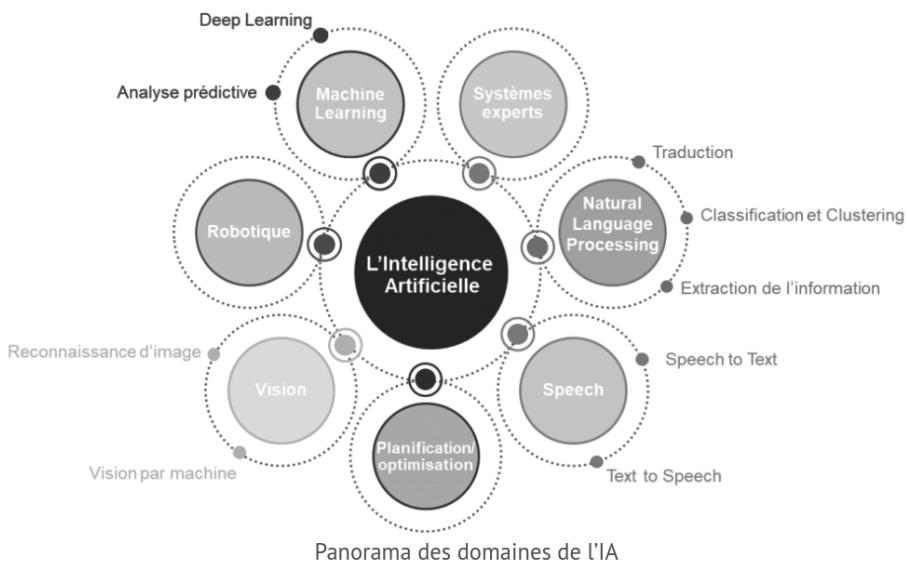
## Ajouter de la valeur



*CETIC, juin 2017*

## *L'intelligence artificielle*

Par ailleurs, les principaux domaines fonctionnels de l'IA témoignent de son immense potentiel, des innombrables réalisations auxquelles elle est susceptible de donner lieu. Il s'agit d'approfondir des domaines fonctionnels déjà connus, à l'instar des systèmes experts, de la planification/optimisation ou de la robotique, mais aussi de développer de nouveaux domaines tels que le *machine learning*, le *natural language processing*, la vision (capacité pour une machine d'appréhender son environnement) ou encore le *speech* (texte vers parole ou parole vers texte).



*Artik Consulting, juin 2018*

*Dans toutes ses dimensions*

Et ces réalisations ont vocation à concerner de très nombreux secteurs :

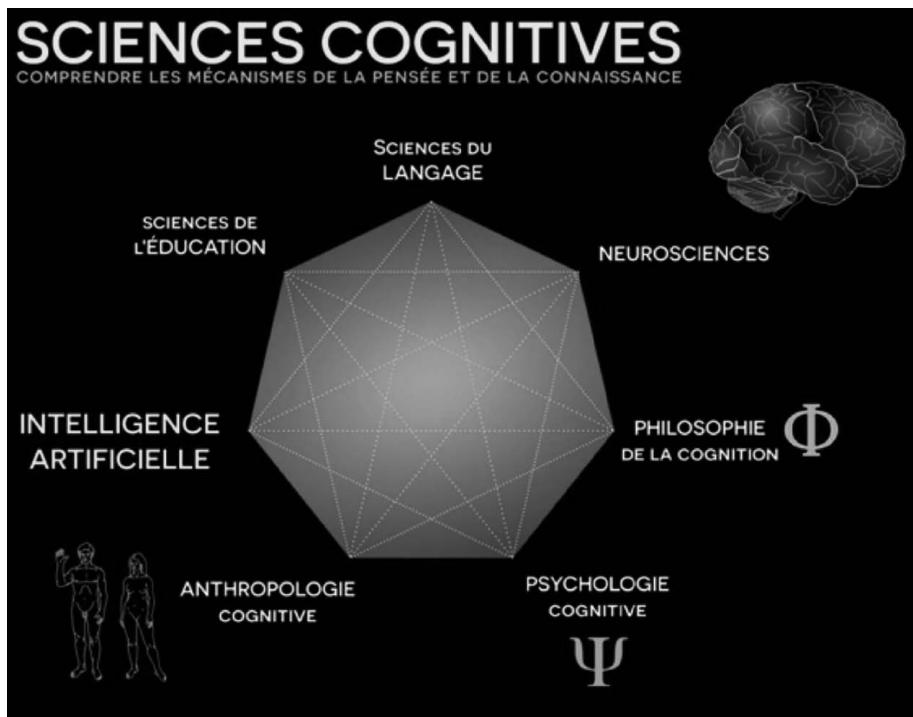
## Utilisation des algorithmes et de l'intelligence artificielle

	EDUCATION	JUSTICE	SANTE	SECURITE	TRAVAIL	CULTURE	AUTRES
Générer du savoir	Mieux cerner les aptitudes d'apprentissage des élèves	Mettre en évidence des manières différentes de rendre la justice selon les régions	Tirer profit de la quantité immense de publications scientifiques	Repérer des liens insoupçonnés pour la résolution d'enquêtes par les services de gendarmerie	Comprendre les phénomènes sociaux en entreprise	Créer des œuvres culturelles (peinture, musique...)	Affiner le profil de risque d'un client d'un assureur
Croiser des infos, «matching»	Répartir les candidats dans les formations d'enseignement (ex APB)		Répartir des patients pour participation à un essai clinique		Faire correspondre une liste de candidatures avec une offre d'emploi		Mettre en relation des profils «compatibles» sur des applications de rencontres
Prédire	Prévoir des décrochages scolaires	Prédire la chance de succès d'un procès et le montant potentiel de dommages intérêts	Prédire des épidémies Repérer des prédispositions à certaines pathologies afin d'en éviter le développement	Détecter les profils à risque dans la lutte antiterroriste Prédire l'occurrence future de crimes et délits	Détecter les collaborateurs qui risquent de démissionner dans les prochains mois	Créer des œuvres ayant un maximum de chance de plaire aux spectateurs (Netflix)	
Conseiller	Recommander des voies d'orientation personnalisées aux élèves	Recommander des solutions de médiation en fonction du profil des personnes et de cas similaires			Proposer des orientations de carrière adaptées aux profils	Recommander des livres (Amazon), des séries télévisées (Netflix), etc.	Individualiser des messages politiques sur les réseaux sociaux
Aider à la décision		Suggérer au juge la solution jurisprudentielle la plus adéquate pour un cas donné	Suggérer au médecin des solutions thérapeutiques adaptées	Proposer aux forces de police les zones prioritaires dans lesquelles patrouiller			Aider à trouver le chemin le plus court (GPS)

© F. Schneider - Source : CNIL-Ethique et numérique

## II. Une problématique technique : au cœur des cerveaux artificiels (apprentissage automatique et apprentissage profond)

« Intelligence artificielle » peut désigner un objet polymorphe, un champ de recherches relativement vaste ou une science particulière. Dans ce dernier cas, il s'agit de l'une des sciences cognitives qui s'intéressent à la pensée et à l'intelligence :



*Audrey Vermeulen, TEDx Canebière, 2018*

Dans ce livre multidisciplinaire, on utilisera surtout « intelligence artificielle » afin de désigner un objet et le champ de recherches multifacette auquel il donne lieu. On pourrait d'ailleurs évoquer des « sciences de l'intelligence artificielle » au sujet des disciplines qui étudient cet objet de différents points de vue — sur le modèle des sciences de l'information et de la communication qui comprennent des dimensions multiples, du droit à la sociologie en passant par l'économie. C'est à ces diverses sciences de l'IA que le présent ouvrage à la fois s'adresse et donne la parole. Il considère que cette technologie serait le cœur d'un système dont on pourrait difficilement ne

comprendre qu'un des rouages abstraitemennt de tous les autres. Il mise sur le fait qu'on pourrait difficilement comprendre l'IA seulement en termes techniques, sociologiques, économiques, juridiques, éthiques, philosophiques, anthropologiques ou autres.

En premier lieu, la dimension technique semble la plus incontournable. Si l'IA doit de préférence être conçue tel un objet, il serait bien sûr malvenu de chercher à disserter sur ou autour de lui sans en maîtriser les origines, implications et potentialités technologiques, c'est-à-dire sans bien saisir « comment ça marche ». Le physicien Richard Feynman affirmait en ce sens qu'« on ne peut comprendre véritablement un instrument tant qu'on ne l'a pas construit soi-même ». Tout chercheur, y compris le philosophe, doit aller sur le terrain pour bien comprendre ce autour de quoi il disserte — comme le faisait, par exemple, Adam Smith durant la révolution industrielle. Sans être mathématicien ni informaticien ou *data scientist*, il importe, pour pouvoir en parler, de ne pas se contenter de définitions superficielles ou basiques et de s'intéresser dans le détail au fonctionnement de la technologie — et pourquoi pas de s'initier au codage informatique. Le chercheur en IA peut difficilement ne pas être technophile, pas dans le sens de technoaddict mais dans celui de passionné par la dimension technique de la technologie, celle-ci expliquant beaucoup de choses sous les divers autres angles.

Historiquement, les origines de l'intelligence artificielle sont très anciennes. Elles s'inscrivent dans la longue histoire des algorithmes, dont les premiers auraient été inventés par les babyloniens il y a 5 000 ans — l'un de ces algorithmes consistant par exemple en une tablette permettant de calculer des racines carrées précises au millionième près. D'autres se retrouvent dans les antiquités égyptienne et grecque. Au XVI<sup>e</sup> siècle, Léonard De Vinci conçut nombre d'automates imitant l'homme ou l'animal. Et, au XVII<sup>e</sup> siècle, Blaise Pascal inventa la Pascaline, une intelligence artificielle mécanique permettant de réaliser des opérations arithmétiques. Aujourd'hui, l'IA ne peut plus se passer de l'électronique, c'est-à-dire de l'électricité comme support de traitement, de transmission et de stockage d'informations.

Durant les années 1950, alors que l'informatique commençait à progresser à grands pas, l'intention d'élaborer des « machines à penser » se rapprochait dans leur fonctionnement de l'esprit humain s'est affirmée, avec des figures de proue telles qu'Alan Turing. Déjà en 1914, l'ingénieur espagnol Leonardo Torres Quevedo inventait le premier véritable automate joueur d'échecs. Celui-ci était capable de remporter une partie à tous les coups lors d'une confrontation roi et tour contre roi. Prouesse d'une autre envergure, le supercalculateur d'IBM Deep Blue battait, en 1997, le champion du monde Garry Kasparov — qui avait pourtant prédit que « jamais une machine ne pourrait être supérieure à un humain ».

Mais c'est en 2012 que le secteur de l'intelligence artificielle franchit un pallier considérable, sortant de l'« hiver » dans lequel il s'était quelque peu endormi, faute de progrès significatifs. Cette année là, pour la première fois, un réseau de neurones profond remporta le challenge ImageNet, un concours de reconnaissance d'images. Avec un taux d'erreurs de seulement 16,4 %, il distança très largement le second programme (26,2 %). Cette performance remit sur le devant de la scène une technique d'apprentissage profond (*deep learning*) qui avait commencé à faire ses preuves dans les années 1990, notamment avec les travaux du français Yann Le Cun et des canadiens Yoshua Bengio et Geoffrey Hinton, avant d'être un temps laissée de côté. La puissance de calcul des cartes graphiques couplée à la mise à disposition de grandes bases de données étiquetées (une image d'un chien sur un vélo en train de jouer de la guitare est accompagnée des indications « chien », « vélo » et « guitare ») permirent à ces programmes d'enfin faire leurs preuves à grande échelle.

La communauté scientifique perçut alors l'immense potentiel de cette technologie : en étant entraînée sur une base de données suffisamment consistante et pertinente, elle est capable d'établir elle-même des règles et standards d'analyse lui permettant d'interpréter d'autres données jusqu'ici trop complexes à traiter. Dès 2013, seuls des réseaux de neurones profonds participaient au challenge ImageNet. Et, en 2016, AlphaGo battait le maître coréen Lee Sedol au jeu de go, lequel est ô combien plus complexe à maîtriser que le jeu d'échecs<sup>26</sup>. Vingt ans après les meilleurs joueurs d'échecs, les champions de go, de poker et de nombreux jeux vidéo s'inclinent à leur tour devant des intelligences artificielles — qui ont pour elles les avantages d'être durables, de ne jamais se déconcentrer, de ne pas subir le stress ni la pression et d'être toujours d'humeur égale, autant de gains psychologiques de la machine par rapport à l'humain. Quant au taux d'erreurs dans la reconnaissance d'images, il a chuté à quelques pourcents. Et des IA sont désormais capables de créer de nouveaux tableaux de Rembrandt ou de nouveaux morceaux des Beatles incroyablement réalistes — bien que les spécialistes ne puissent pour l'heure les qualifier que de « mauvais Rembrandt » ou de « mauvais Beatles ».

Les concepteurs de réseaux de neurones sont un peu des alchimistes : ils associent de diverses manières leurs formules mathématiques et leurs technologies et, lorsqu'elles produisent des résultats satisfaisants, ils cherchent à perfectionner la recette sans nécessairement comprendre son mécanisme exact. En la matière, l'« art de la débrouille » est important. Un réseau de neurones ne se programme pas mais s'éduque. Il s'entraîne et il

---

<sup>26</sup> Quelques mois plus tard, AlphaGo première version a lui-même été battu par son successeur, entraîné sans exemples, uniquement par exploration, sur le score de 100 parties à 0, ce qui témoigne de la vitesse des progrès de l'IA.

apprend. Et cet apprentissage peut se faire soit par exemples soit par exploration, c'est-à-dire soit en étant guidé par l'homme soit de façon autonome et automatique. Pour l'heure, c'est l'apprentissage grâce à des exemples et donc des statistiques qui est essentiellement utilisé.

Pour entraîner une intelligence artificielle à reconnaître des chats sur des images contenant aléatoirement des chiens ou des chats, on lui montre un grand nombre d'exemples d'images de chats et un grand nombre d'exemples d'images de chiens. Plus le nombre d'exemples est grand, plus la machine devient capable de reconnaître seule les chats car elle a en mémoire beaucoup de cas de figure possibles avec lesquels elle peut comparer l'image qu'elle traite. C'est cette technique qu'on appelle « apprentissage automatique » (*machine learning*). Lorsqu'on parle d'intelligence artificielle aujourd'hui, c'est très souvent d'apprentissage automatique au moyen de réseaux de neurones et d'exemples qu'il s'agit. Cette intelligence est toutefois bien relative et, tandis que, pour qu'un jeune enfant apprenne à distinguer le chat du chien, il suffit de lui en désigner trois ou quatre de chaque espèce en lui disant « ceci est un chat » et « ceci est un chien », il faut montrer des milliers voire des millions d'exemples à une IA pour qu'elle devienne aussi perspicace.

Le *machine learning* est notamment utilisé en matière de recommandation (rechercher les consommateurs les plus proches d'un individu afin de lui suggérer l'achat de certaines catégories de produits), de *clustering* (définition de groupes typiques, par exemple d'utilisateurs d'un réseau social), de détection automatique de fraudes, anomalies ou erreurs de saisie etc.

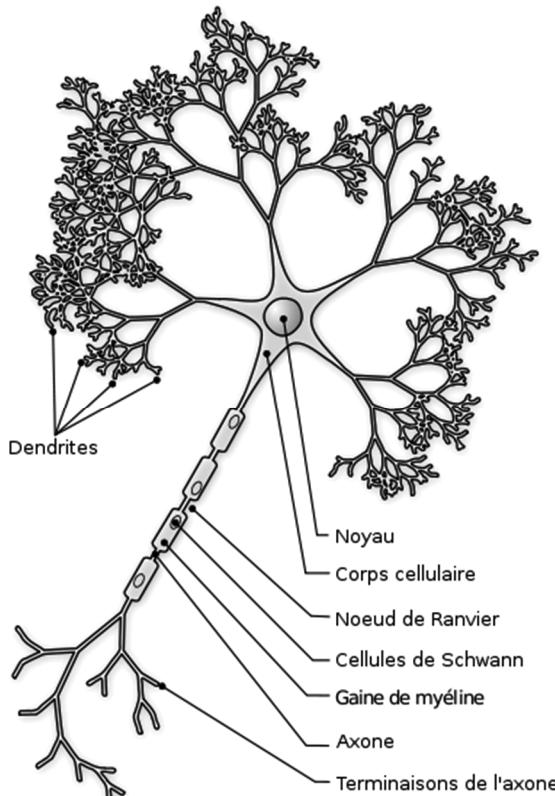
Quant au *deep learning* (apprentissage profond), il est le niveau supérieur du *machine learning*. Il désigne des méthodes d'apprentissage dédiées aux phénomènes complexes pour lesquels une structure à plusieurs couches est nécessaire. Il intervient afin de résoudre des problèmes appelant des méthodes plus sophistiquées, comme la reconnaissance automatique de textes, le traitement du langage naturel ou la reconnaissance d'images (analyse des émotions, diagnostics automatiques à partir d'images médicales, compréhension de l'environnement proche pour les voitures autonomes).

Les cerveaux artificiels ne cherchent pas à reproduire le plus exactement possible la constitution et le mécanisme des cerveaux biologiques. Leur objectif est essentiellement d'aboutir aux mêmes résultats, donc d'imiter la nature dans les performances dont elle est capable mais pas dans les mécanismes qui permettent d'aboutir à ces performances. Néanmoins, la technologie des réseaux de neurones artificiels s'inspire du fonctionnement du cerveau. Ce dernier est peut-être l'objet le plus complexe qu'on connaisse dans l'univers et son mécanisme demeure aujourd'hui encore assez mal

compris. Il fonctionne grâce à des milliards de neurones. Un neurone est une cellule nerveuse constituant l'unité fonctionnelle du système nerveux. Elle a la particularité de posséder des entrées qui répondent à des stimuli, les dendrites (environ 7 000 par neurone), et une sortie en direction des autres neurones, l'axone. Le neurone reçoit des informations en provenance de milliers d'autres neurones, puis stimule à son tour les neurones voisins en envoyant à travers son axone un signal bioélectrique : l'influx nerveux<sup>27</sup>.

---

<sup>27</sup> Les neurones ont donc deux propriétés physiologiques : l'excitabilité, c'est-à-dire la capacité de répondre aux stimulations et de convertir celles-ci en impulsions nerveuses, et la conductivité, c'est-à-dire la capacité de transmettre les impulsions. Le cerveau n'est pas le seul organe du corps humain à comporter des neurones. Tel est également le cas de l'intestin, qui en comporte 200 fois moins, et du cœur.

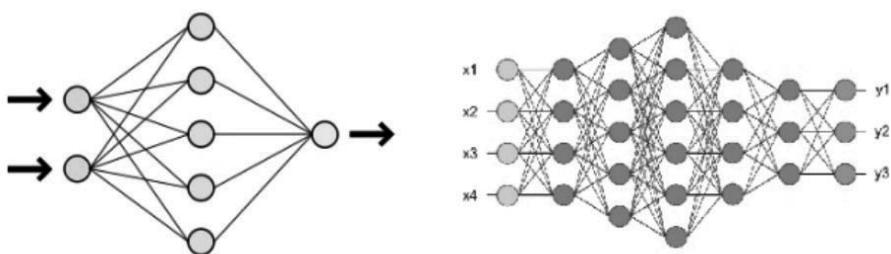


Nicolas Rougier, schéma biologique d'un neurone, 2007

Pour apprendre, un cerveau humain ajuste la force des interconnections entre ses milliards de neurones. Personne ne le programme. Un réseau de neurones artificiels fonctionne de la même manière<sup>28</sup>. Il faut se le représenter sous les traits d'une série de variateurs de lumière reliés entre eux. Chacun reçoit des données en entrée et, suivant le réglage du variateur, produit une valeur en sortie. Cette opération se répète des milliers, voire des millions de fois à l'intérieur du réseau, chaque neurone pouvant envoyer son résultat à plusieurs autres. On parle d'apprentissage profond (*deep learning*) car le réseau est constitué de différentes couches de neurones qui communiquent.

<sup>28</sup> C'est en 1958 que le psychologue américain Frank Rosenblatt inventa la première machine à base de neurones artificiels, avec des entrées qui reçoivent les données, une couche interne qui effectue les calculs et une couche de sortie produisant des résultats. Il démontra qu'une telle machine pouvait potentiellement effectuer tous types de calculs. Mais, à l'époque, on était dans l'incapacité d'apprendre à de telles machines quoi que ce soit. C'est dans les années 1980, avec les travaux de Yann Le Cun, Yoshua Bengio et Geoffrey Hinton, qu'on parvint à inventer des techniques d'apprentissage permettant de faire progresser de telles machines.

Au terme de l'analyse des données, les résultats des neurones de sortie sont associés afin de produire une valeur statistique. L'intelligence artificielle, par exemple, estime qu'il est à 85 % certain qu'une image  $i$  est celle d'un chat. Elle peut alors affirmer « il y a un chat dans  $i$  ». Un réseau de neurones est donc une chose simple à première vue : en fonction des données reçues en entrées, des connexions se déclenchent ou non entre les neurones, jusqu'à la sortie qui est le résultat du cheminement suivi. Et les paramètres internes s'ajustent progressivement en fonction des réussites et des erreurs.



Méthode:

1. Commencer par des valeurs choisies au hasard.
2. Faire varier légèrement les valeurs.
3. Soumettre en entrée les caractéristiques d'une situation bien connue.
4. Comparer le résultat des valeurs courantes avec celui des valeurs précédentes.
  - Si le nouveau résultat est plus proche de la réalité que le précédent, « récompenser » le paramétrage actuel et continuer dans cette direction.
  - Sinon, « affaiblir » le paramétrage actuel et chercher d'autres valeurs légèrement différentes qui pourraient donner de meilleurs résultats.
5. Recommencer en (3) jusqu'à ce que les résultats du réseau soient suffisamment proches des résultats attendus.

*CETIC, juin 2017*

Pour qu'un réseau de neurones soit efficace, il faut qu'il ait été supervisé à partir d'une base de données la plus vaste possible<sup>29</sup>. À ce niveau, cerveau artificiel et cerveau naturel fonctionnent très différemment. Les tsunamis de données sont anti-biologiques : un cerveau humain est incapable de traiter autant d'informations mais est incommensurablement plus efficace pour comprendre les choses au départ de quelques données<sup>30</sup>. Au départ, néanmoins, un réseau de neurones artificiels est totalement incapable de distinguer un chat d'un chien comme l'est le cerveau d'un bébé qui n'a encore rien appris. Il produit donc des résultats aléatoires, avec 50 % d'erreurs si son alternative est binaire. Le programme de supervision lui indique alors quels sont les cas dans lesquels il s'est trompé et ceux dans lesquels il a abouti au bon résultat. C'est à cet instant que les variateurs de lumière entrent en action. Le programme remonte le cours du réseau et fait varier l'intensité des neurones pour que le résultat de sortie se rapproche du résultat attendu. L'opération doit être reproduite un grand nombre de fois pour que le réseau, au fur et à mesure, ajuste ses résultats et fasse preuve de plus en plus d'acuité dans sa reconnaissance des images.

Par le passé, les systèmes d'analyse statistique exigeaient de fortes connaissances du problème et des experts œuvraient durant des mois à la transformation des données de la bonne manière pour que les algorithmes arrivent à en dégager des éléments pertinents et utiles. L'apprentissage profond est un grand progrès car il délègue cette partie de traitement des données au système d'apprentissage lui-même. Cette méthode peut donc être appliquée à des mégadonnées qui, par le passé, étaient laissées de côté à cause du coût de conception du système. L'intelligence artificielle peut ainsi être dix, cent, mille, un million voire un milliard de fois plus performante que les

---

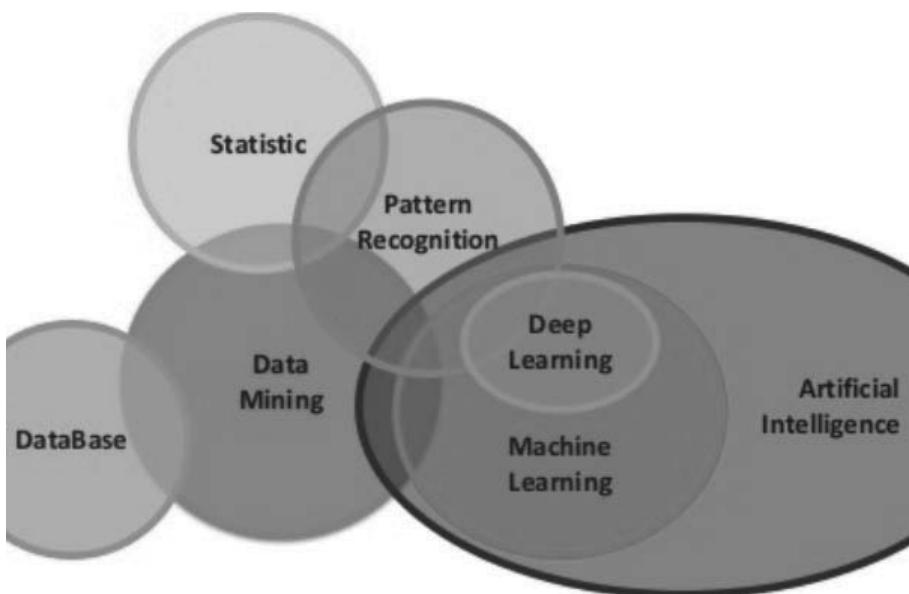
<sup>29</sup> Mais la quantité de données d'une seule image peut constituer une difficulté. Une photographie numérique contient plusieurs millions de pixels, chacun pouvant être considéré telle une donnée d'entrée pour le réseau. Le volume de données risque alors d'être beaucoup trop important pour être traité efficacement. C'est ici que l'aspect convolutif du réseau intervient — en anatomie, cela désigne les circonvolutions, les sinuosités du cerveau. Le réseau convolutif découpe l'image en plus petites parties et applique sur elles une série de filtres. Cela permet de réduire la quantité de données tout en conservant la pertinence des informations obtenues, notamment en identifiant des formes. La production de ces filtres dépend également du cycle d'apprentissage, c'est-à-dire que le réseau convolutif va créer progressivement des filtres optimisés pour les formes caractéristiques du sujet à reconnaître. Cette technologie peut aussi être déployée concernant du son ou d'autres types de données. Il suffit en effet de pouvoir représenter les informations sur une surface en deux dimensions (une image en quelque sorte) et de satisfaire à la contrainte principale des réseaux convolutifs : deux données proches ont de grandes chances d'avoir un rapport entre elles (comme deux pixels collés dans une image font généralement partie du même objet).

<sup>30</sup> L'information circule dans le cerveau, dans les neurones, à un mètre par seconde contre la vitesse de la lumière dans les microprocesseurs, soit 300 000 km par seconde.

## *L'intelligence artificielle*

meilleurs cerveaux humains. Mais elle se limite à du *statistical learning*, de l'apprentissage statistique, inductif, permettant de dégager des règles générales à partir de cas particuliers. Par exemple, grâce à l'historique de nombreux acheteurs, un algorithme va déterminer que deux biens sont complémentaires parce que l'achat de l'un entraîne souvent, peu de temps après, l'achat de l'autre. Il pourra dès lors recommander automatiquement l'achat de l'un aux acheteurs de l'autre.

En résumé, l'intelligence artificielle a donc besoin de bases de données, de statistiques, de la science des données, de la reconnaissance de formes et de l'exploration et extraction de données. Et son cœur est constitué par l'apprentissage automatique, dont le propre cœur est l'apprentissage profond.



*CETIC, juin 2017*

L'intelligence artificielle apparaît sitôt qu'on parvient à obtenir de façon automatique, concernant un objet donné, des résultats pertinents. Un thermomètre est ainsi une forme très primaire d'intelligence artificielle car capable de saisir et exprimer le chaud et le froid, les variations de température, cela de façon plus précise qu'un homme dont le ressenti est biaisé par différents facteurs. De la plus sommaire à la plus perfectionnée, les intelligences artificielles sont donc susceptibles de connaître un très grand nombre de degrés d'évolution. Une IA consciente d'elle-même, donc une IA vivante, capable de développer ses propres projets, serait le quatrième âge de l'intelligence artificielle, après les algorithmes traditionnels (premier âge),

l'apprentissage profond (deuxième âge) et l'IA contextuelle et conceptuelle, capable de créativité, d'empathie et d'arbitraire, de transversalité et de multidisciplinarité (troisième âge). Mais il est très incertain qu'on parvienne un jour à dépasser le deuxième âge. Pour l'heure, la machine peut être plus intelligente que l'homme, voire beaucoup plus intelligente que lui, mais seulement s'agissant de tâches très spécifiques — comme elle peut être beaucoup plus forte physiquement que lui pour réaliser certains actes très précis et prédéfinis.

Autant les résultats produits grâce à des traitements statistiques sont impressionnants et peuvent intervenir dans une très grande variété de domaines, autant il est difficile de créer une intelligence artificielle fonctionnant autrement qu'à base de statistiques et d'exemples. Depuis 1956 et la conception de l'intelligence artificielle, très peu de progrès ont été faits en matière de vraie intelligence artificielle, c'est-à-dire de conscience artificielle, d'arbitraire artificiel, de compréhension artificielle, de créativité artificielle et de bon sens ou sens commun artificiel<sup>31</sup>.

La recherche fondamentale en intelligence artificielle se concentre désormais sur l'apprentissage faiblement supervisé ou non supervisé, c'est-à-dire l'apprentissage par exploration et par renforcement. Il s'agit de permettre aux machines d'apprendre avant tout par libre observation, sans leur fournir au préalable des données prémachées, conquérant ainsi cette autonomie sans laquelle il n'est aucune forme d'intelligence. L'apprentissage non supervisé a lieu lorsque seulement de grands nombres d'exemples sont fournis à la technologie d'IA. À partir de ces exemples, cette dernière doit déterminer les structures et règles cachées permettant de relier les exemples entre eux et donc de comprendre ontologiquement ce dont il s'agit.

L'intelligence artificielle du jeu de go AlphaGo apprend ainsi les meilleures stratégies en jouant des millions de parties contre elle-même avec comme uniques connaissances de départ fournies par des humains les règles du jeu. Cette forme d'apprentissage par auto-entraînement est très différente des générations précédentes d'intelligence artificielle. Elle permet à l'IA d'être créative, d'aboutir à des solutions auxquelles l'homme n'aurait jamais pensé, mais elle requiert de grandes capacités informatiques puisqu'il faut l'équivalent de millions d'années de réflexion humaine pour obtenir des résultats pertinents. Aussi l'apprentissage automatique par exemples reste-t-il

---

<sup>31</sup> En revanche, il est remarquable qu'une étude d'un cabinet de prospective américain a, en 2015, fait du métier de statisticien le numéro un dans le classement des plus beaux métiers du monde. Il s'agit un peu d'une revanche pour cette profession longtemps déconsidérée par les mathématiciens et, plus généralement, par les scientifiques. Dans ce genre de classement, le métier de mathématicien figure également souvent sur le podium.

le plus courant et, dans de nombreux cas trop complexes pour permettre un apprentissage par exploration, le seul envisageable.

La recherche appliquée, de son côté, se charge d'optimiser, de rendre concrètes, de diffuser et d'exploiter les possibilités de l'apprentissage profond dans un maximum de domaines. C'est ainsi que l'intelligence artificielle s'est immiscée, assez discrètement et subrepticement d'ailleurs, au cœur de beaucoup d'objets et d'activités. En témoignent, en premier lieu, les instruments de communication numérique : ordinateurs, téléphones portables et tablettes numériques. Obtenir des images ou des informations très pertinentes à partir d'une requête donnée, grâce à des moteurs de recherche en ligne accessibles « en un clic », constitue par exemple une performance algorithmique extraordinaire dont le grand public profite quotidiennement, sans d'ailleurs s'en émerveiller autre mesure.

Cette recherche privée et appliquée peut être en avance par rapport à la recherche fondamentale. Néanmoins, s'il faut de nombreuses années, des décennies, pour résoudre un problème particulier relatif par exemple à l'apprentissage non supervisé, alors les travaux des centres de recherche publics resteront indispensables aux progrès de l'intelligence artificielle, les puissances privées risquant de refuser de financer cette recherche. Cela doit toutefois être relativisé par le fait que les multinationales du numérique embauchent de plus en plus de chercheurs, auxquels elles offrent des conditions de travail avantageuses, afin qu'ils réalisent leurs recherches fondamentales chez elles et pour elles — quoique les résultats de ces recherches sont souvent publiés en *open access* et *open source*.

La majorité des innovations numériques et des technologies et applications de l'intelligence artificielle provient d'un écosystème public-privé unique : la Silicon Valley. Cela explique pourquoi d'aucuns qualifient la France et les pays d'Europe de « colonies numériques » ou de « pays du tiers-monde numérique », à défaut d'avoir vu naître certaines des grandes plateformes capables de créer et entraîner les IA<sup>32</sup>. Une forte concentration frappe d'ailleurs la production d'intelligences artificielles : il n'existe qu'une dizaine de grandes entreprises développant ces technologies (les fameux GAFAM et BATX : Google, Apple, Facebook, Amazon, Microsoft, Baidu, Alibaba, Tencent, Xiaomi<sup>33</sup>). La concentration est d'autant plus forte que les GAFAM n'hésitent pas à racheter à prix d'or les start-up qui créent des IA performantes et/ou originales<sup>34</sup>.

---

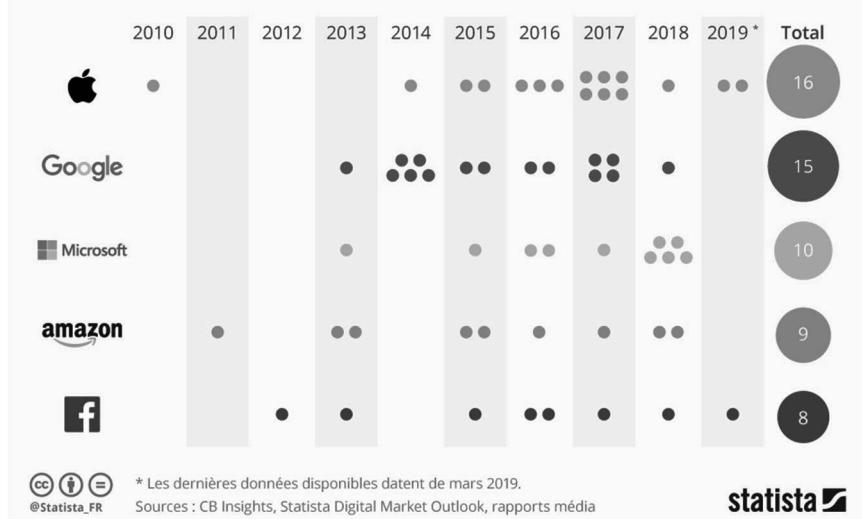
<sup>32</sup> France et Europe importent massivement les technologies IA américaines sur smartphones Android et iPhones.

<sup>33</sup> On peut y ajouter IBM, Samsung, Qualcomm, Intel, Philips, Siemens et Sony.

<sup>34</sup> L'IA est présentée comme un « grand aspirateur à valeur » et Laurent Alexandre, lors de son intervention au colloque du Sénat sur l'intelligence artificielle du 19 janvier 2017, cite l'exemple suivant : les 55 salariés de Whatsapp ont créé en quatre ans 23

## Apple et Google en tête de course à l'intelligence artificielle

Nombre de start-ups IA rachetées par les GAFAM depuis 2010 \*



Reste que le fonctionnement de l'intelligence artificielle à travers des réseaux de neurones est délicat à cerner, même de façon abstraite et vague. On ressent forcément l'impression d'avoir affaire à quelque chose d'insaisissable, d'incompréhensible, d'inmaîtrisable. Même pour un spécialiste, cet assemblage de millions de variateurs de lumière reliés entre eux, une fois qu'il en a terminé avec son apprentissage, s'avère relativement opaque. On ne saurait comprendre le rôle précis de chaque neurone artificiel dans la production d'un résultat ni retracer exactement la série de calculs effectuée par la machine. On ne peut que faire confiance au réseau, constatant qu'il produit généralement des résultats pertinents. C'est ce qu'on appelle communément la « boîte noire » de l'intelligence artificielle. Une fois que ses concepteurs, ses parents, lui ont donné la vie, elle atteint rapidement l'adolescence puis l'âge adulte et il leur devient difficile de lui imposer la conduite à adopter. Or elle est amenée à prendre des décisions immédiates et très nombreuses, comme en matière de conduite autonome, de reconnaissance de la parole ou des visages, de classement de fils d'informations, de recherche d'images, d'identification de contenus violents ou autrement choquants ou de détermination des publications à afficher en priorité sur les réseaux sociaux. De graves problèmes moraux peuvent se poser, par exemple en cas d'accident provoqué par une voiture autonome ; et il est difficile d'abandonner leur

---

milliards d'euros de valeur contre la moitié à peine pour les 130 000 ouvriers de Peugeot en un siècle.

résolution aux calculs mathématiques et opaques d'un algorithme, au parcours incertain et hyper-complexe d'un réseau de neurones artificiels.

Les succès de l'intelligence artificielle dépendront de la confiance que le grand public lui portera ; et cette confiance suppose transparence et traçabilité — impératifs éthiques et politiques — dès lors qu'une décision est prise sur la base d'une indication algorithmique, cela afin d'éviter sa non-explicabilité. Si une science en devenir est la rétro-ingénierie, consistant à déconstruire les raisonnements des machines, créer des technologies plus lisibles et traçables permettrait de mieux les comprendre et donc mieux les accepter. Mais il est très incertain que de telles technologies soient concevables en matière d'IA. Peut-être faudrait-il surtout que davantage de sociologues, d'anthropologues et de philosophes accompagnent les projets d'intelligences artificielles, de leurs prémisses scientifiques à leurs conséquences pratiques ; et que les futurs ingénieurs et mathématiciens soient formés aux enjeux éthiques de leurs travaux, formés à la « robéthique ». Où revient le besoin de pluri-, inter- et même trans-disciplinarité dans les travaux relatifs à l'intelligence artificielle, besoin auquel ce livre tente de répondre en envisageant l'IA « dans toutes ses dimensions » — et notamment dans ses dimensions humaine, sociale, politique et économique.

### **III. Une problématique sociale, politique et économique : l'intelligence artificielle et les données, moteur et carburant d'un changement de civilisation**

Déjà en 1954, dans son essai *La Technique ou l'enjeu du siècle*, alors que l'informatique commençait seulement à sortir de l'âge du simple concept, Jacques Ellul interrogeait le changement de nature de la technique dans la société, cela en des termes qui résonnent fortement à l'ère des IA et des multinationales du numérique. Auparavant simple instrument permettant à l'homme de mieux s'appuyer sur la nature et de dépasser sa condition initiale, elle muerait en un processus autonome de moins en moins au service des hommes car tendant à les enfermer dans de nouvelles formes de servitude. Le sociologue définissait « le phénomène technique [comme] la préoccupation de l'immense majorité des hommes de rechercher en toutes choses la méthode absolument la plus efficace »<sup>35</sup>, ce qui présageait du mouvement d'hyper-

---

<sup>35</sup> J. Ellul, *La Technique ou l'enjeu du siècle*, Armand Colin, coll. Sciences politiques, 1954, p. 18-19.

rationalisation actuel de toutes les décisions et de toutes les actions. Et Jacques Ellul d'observer que la technique serait « à la fois sacrilège et sacrée » : « L'invasion technique désacralise le monde dans lequel l'homme est appelé à vivre. [...] Mais nous assistons à un étrange renversement : l'homme ne pouvant vivre sans sacré, il reporte son sens du sacré sur cela même qui a détruit tout ce qui en était l'objet : sur la technique »<sup>36</sup>.

Aussi les puissantes technologies et les start-up qui les développent, dont certaines deviennent en un temps record des multinationales et même les premières puissances du monde contemporain, se retrouvent-elles statufiées, érigées en grands mythes des sociétés postmodernes. Beaucoup des services à base d'IA mis sur le marché ne sont pas des gadgets mais de véritables outils de gouvernement, qui régulent effectivement les comportements individuels et sociaux, parfois mieux que les lois des États.

Les grands acteurs actuels de l'économie, et des sociétés qui suivent cette économie, sont des puissances privées qui concurrencent les puissances publiques sur la scène interne et, plus encore, sur la scène externe — à tel point que le Danemark a cru bon de nommer un « ambassadeur auprès des GAFAM », ainsi explicitement considérés tels des acteurs souverains, tels les alter egos des États<sup>37</sup>. Ces multinationales du numérique se sont développées autour d'algorithmes, sortes de recettes de cuisine jalousement conservées, les ingrédients étant les gigantesques masses de données constitutives du *big data*. Mieux vaut d'ailleurs avoir un mauvais algorithme avec beaucoup de données qu'un bon algorithme avec peu de données. Mais ces acteurs possèdent les deux. L'intelligence artificielle devient ainsi le noyau dur de beaucoup de secteurs de l'économie, de la recherche et, par suite, de la société et de la vie. Ce sont les géants du numérique, GAFAM et BATX en tête<sup>38</sup>, qui possèdent les principaux réservoirs de données. C'est pourquoi ils s'imposent telles les principales puissances technologiques et économiques, mais aussi comme les premières puissances politiques et scientifiques.

L'opinion publique s'insurge généralement haut et fort dès qu'un parlement ou un gouvernement a l'idée d'une loi sur le renseignement afin de lutter plus efficacement contre certaines menaces actuelles, un pareil texte étant perçu comme trop liberticide. Mais, aujourd'hui, les véritables espions et les véritables services de renseignement sont ceux des multinationales des données. Ce sont elles qui connaissent dans leurs moindres détails tous les faits et gestes des individus. Alors que le « technolibéralisme » serait la grande mutation du libéralisme, devenant

---

<sup>36</sup> *Ibid.*, p. 130-132.

<sup>37</sup> Et, lorsque cet ambassadeur souhaite rencontrer les représentants de Facebook, celui-ci dépêche ses stagiaires pour le recevoir.

<sup>38</sup> Mais d'autres acteurs sont bien sûr importants, à l'image, par exemple, d'Uber ou de Netflix.

libéralisme augmenté<sup>39</sup>, la véritable question des contre-pouvoirs est désormais celle des contre-pouvoirs face aux géants de l'économie numérique. Beaucoup acceptent sans autre forme de procès, car ils en sont trop peu conscients, la visibilisation intégrale de la vie sociale et même de la vie privée. Dans *1984* de George Orwell, le télécran<sup>40</sup> possède des angles-morts grâce auxquels on peut échapper à la surveillance. Aujourd'hui, ces moyens de se soustraire au contrôle sont le refus des réseaux sociaux et des objets connectés. Mais il est difficile de se convaincre d'un tel choix car le *Big Brother* omnipotent, précisément identifié et dont les missions étaient précisément définies, a laissé la place à une diversité de services *a priori* beaucoup moins effrayants et dont ce que l'on voit surtout est l'utilité — immense et incontestable dans d'innombrables domaines.

Deux dynamiques s'associent afin de dessiner le futur à bases d'algorithmes, aboutissant à redéfinir l'économie, la société, mais aussi la démocratie, les politiques publiques, et peut-être même l'homme en tant que tel. La première est la numérisation de la plupart des actes et activités individuels et collectifs, nourrissant de gigantesques bases d'informations plus ou moins personnelles et plus ou moins confidentielles<sup>41</sup>. Il y a peu encore, toutes ces données n'étaient pas enregistrées et, lorsqu'elles l'étaient, elles restaient difficiles d'accès et complexes à exploiter. En particulier grâce à l'internet (avec les sites web, les applications mobiles et les objets connectés), des torrents d'informations sont en permanence livrés par chacun et stockés dans d'immenses serveurs informatiques.

Quant à la seconde dynamique, elle est liée au besoin de comprendre et interpréter ces mégadonnées. C'est ici qu'interviennent les algorithmes, donnant aux ordinateurs des instructions mathématiques pour trier, traiter, agréger et représenter ces données. On assiste alors à un changement de statut des technologies de traitement de l'information : désormais, elles vont très au-delà des trois fonctions de base que sont la collecte, le stockage et l'indexation. De puissantes technologies statistiques et « prédictives » se développent en suivant l'augmentation des capacités de calcul des ordinateurs. L'intelligence artificielle est capable d'identifier des corrélations, de percevoir des

---

<sup>39</sup> É. Sadin, *La silicolonisation du monde – L'irrésistible expansion du libéralisme numérique*, L'Échappée, coll. Pour en finir avec, 2016.

<sup>40</sup> Le télécran est à la fois un système de télévision, qui diffuse en permanence les messages de propagande du Parti, et un outil de vidéo-surveillance permettant à la Police de la Pensée d'entendre et de voir ce qui se fait dans chaque pièce où se trouve un individu.

<sup>41</sup> En matière de données, la France, traditionnellement jacobine, est fort bien dotée. Le mouvement d'ouverture des données (*open data*) est d'autant plus bienvenu que beaucoup de jeux de données massives sont maîtrisés par des acteurs publics. Par exemple, en matière de santé, l'assurance maladie française est une ressource de données sans équivalent dans le monde.

ressemblances ou de trouver des informations clés parmi des ensembles gigantesques de données beaucoup mieux que l'intelligence humaine. Les classements, palmarès, compteurs, cartes, recommandations et notes diverses qui s'affichent en permanence sur les écrans des ordinateurs, des tablettes et des téléphones portables sont les résultats de la calculabilité toujours plus fine des activités à venir en fonction des traces laissées par les activités passées.

La quantité de données produites par la société et les hommes numériques croît de façon exponentielle, tout comme les performances des réseaux de communication, des systèmes de stockage et des algorithmes. L'intelligence artificielle touche (ou touchera) tous les domaines : elle réaménage les villes, facilite l'accès au savoir et la diffusion de la culture, permet une médecine personnalisée et plus efficace, sert à lutter contre le changement climatique, tente de trouver des emplois pour les personnes au chômage, affecte les étudiants dans les établissements d'enseignement supérieur, décide d'accorder ou non des prêts bancaires, intervient dans les placements financiers, indique aux automobilistes les routes à emprunter, diffuse des publicités et autres contenus en rapport avec les besoins et/ou avec les centres d'intérêt, forme les couples etc. Au résultat, l'IA serait le moteur d'un changement de civilisation avec la marchandisation intégrale des vies et l'algorithmisation de pans entiers de l'évolution des sociétés<sup>42</sup>.

À l'image de toutes les grandes révolutions économiques et sociétales, amorcées par la découverte et l'exploitation d'une nouvelle forme d'énergie, les données numériques constituent une ressource, un gisement de valeur capable de relancer l'innovation, la productivité et la croissance tout en les guidant vers de nouveaux horizons jusqu'alors inconnus. Mais ces données peuvent aussi être utilisées afin de créer de nouvelles formes de servitude. Peut-être même s'agit-il d'une nouvelle forme de servitude volontaire : en cliquant systématiquement et machinalement sur « j'accepte les conditions d'utilisation et les cookies », car il est sinon impossible ou difficile d'utiliser beaucoup des services en cause, chacun consent plus ou moins consciemment à une extorsion de ses données.

C'est pourquoi la politique et le droit doivent intervenir<sup>43</sup>, à l'image de l'Union européenne et du *Règlement général sur la protection des données*

---

<sup>42</sup> É. Sadin, *La vie algorithmique – Critique de la raison numérique*, L'Échappée, coll. Pour en finir avec, 2015.

<sup>43</sup> Il est d'ailleurs significatif que Mark Zuckerberg plaide pour plus de régulation publique car il est difficile pour des groupes tels que Facebook de produire les normes sur certains problèmes de société et certaines questions sensibles, par exemple afin de limiter la liberté de communication de certains mouvements radicaux, ce qui revient à les censurer. Dans tous les cas, il n'est sans doute pas réjouissant que des libertés fondamentales telles que, par exemple, la liberté d'expression ne reposent plus sur des sources publiques mais sur des sources privées, en l'occurrence les GAFAM dont les

(RGPD) applicable depuis le 25 mai 2018, qui vise à mieux protéger les données des européens<sup>44</sup>. Au-delà, se pose avec de plus en plus d'acuité la question d'un système dans lequel chacun serait maître de ses données, pourrait plus et mieux décider de celles qu'il abandonne, à qui et dans quelles conditions, ainsi que de celles qu'il conserve. En ce sens, certains, à l'image du philosophe Gaspard Koenig<sup>45</sup>, défendent l'instauration d'un droit de propriété sur les données — quand celles-ci, pour l'heure, sont *res nullius*, des choses sans maître. Des start-up, anticipant de telles patrimonialisation, monétisation et commercialisation des données, travaillent déjà sur des portefeuilles intelligents de données qui permettraient de jouir pleinement de l'*usus*, du *fructus* et de l'*absusus* sur ses informations personnelles, *i.e.* les droits de les utiliser, d'en percevoir les fruits, les revenus (par exemple en les louant), et d'en disposer, de les transmettre ou de les détruire. Ce ne seraient dès lors plus les services mais les utilisateurs qui imposeraient leurs « conditions d'utilisation », acceptant d'être « nudgés » par des IA dans certaines de leurs activités et certains de leurs choix mais pas dans d'autres.

L'intelligence artificielle se construit aussi par la politique et par le droit, qui fixent des bornes, un cadre, afin que le progrès profite au plus grand nombre et non seulement aux plus forts. L'enjeu, de leur point de vue, est de parvenir à devancer plutôt que suivre les technologies et les usages. Dans le cas du RGPD et des régimes de protection des données personnelles plus généralement, ceux-ci sont un obstacle de taille aux progrès de l'intelligence artificielle dans les pays concernés puisque d'autres, ailleurs dans le monde, peuvent beaucoup plus librement exploiter de grandes bases de données,

---

conditions d'utilisation peuvent être plus effectives que le premier amendement de la Constitution aux États-Unis ou la grande loi du 29 juillet 1881 en France.

<sup>44</sup> Le double objectif de ce texte est d'accroître la responsabilité des acteurs de traitements de données et de consacrer de nouveaux droits en faveur des citoyens européens : droit à l'effacement (ou droit à l'oubli), droit à la portabilité des données personnelles, consentement « explicite » et « positif » à la collecte de ses données personnelles (notamment via l'acceptation des cookies et le contrôle de l'utilisation des données renseignées à travers les formulaires de contact), droit de ne pas faire l'objet d'une décision fondée exclusivement sur un traitement automatisé, y compris le profilage, produisant des effets juridiques concernant la personne ou l'affectant de façon similaire, ou encore obligation pour toutes les activités qui peuvent avoir des conséquences importantes en matière de protection de données personnelles d'être précédées d'une étude d'impact sur la vie privée.

<sup>45</sup> G. Koenig, *La fin de l'individu – Voyage d'un philosophe au pays de l'intelligence artificielle*, Éditions de L'Observatoire, 2019. L'auteur rappelle combien, historiquement, toutes les grandes révolutions techniques ont été accompagnées d'une extension du droit de propriété, par exemple avec la terre et la paysannerie au sortir de la féodalité, l'imprimerie et le droit d'auteur qui l'a suivie, soutenu par Beaumarchais, ou encore la création des brevets sur les innovations au moment de la révolution industrielle.

devenant ainsi les leaders de l'IA. L'Europe, en édictant le RGPD, a donné un avantage énorme aux géants du numérique américains et chinois. Mais est-ce critiquable ? Il s'agit d'un choix politique. Gaspard Koenig y voit le « suicide stoïcien de l'Europe », qui préfère mourir en brandissant ses valeurs et principes plutôt que de continuer à vivre en les bafouant<sup>46</sup>. L'Europe préfère sanctifier certains de ses principes, notamment des droits et libertés fondamentaux, même si cela doit nuire à sa compétitivité, à sa capacité d'innovation et, finalement, à sa place sur l'échiquier international.

Lorsque la confidentialité des données relève de la protection de la vie privée en Europe, conçue tel un droit à prétention universelle, il apparaît suspect de vouloir protéger sa vie privée en Chine. Là-bas, chercher à protéger ses données est perçu comme de l'égoïsme de la part de quelqu'un qui ne contribue pas à l'effort collectif en vue du bien commun. De même, la Chine tire un grand avantage de la société de surveillance qu'elle déploie à grand renfort d'IA<sup>47</sup>. Et, par exemple, elle profite de l'absence de liberté d'expression sur son territoire et de la consécration de cette même liberté ailleurs dans le monde, qu'elle exploite afin de diffuser sa propagande et sa désinformation et ainsi accroître son *soft power*<sup>48</sup>. Mais faudrait-il pour autant suivre cet exemple ? Depuis des siècles, la prospérité allait de pair avec le développement et la protection des droits et libertés individuels. Avec l'économie des données et des IA, le phénomène inverse se produit et, plus généralement, divers droits et libertés fondamentaux se trouvent malmenés. Sont atteints ou pourraient être atteints le fonctionnement du système constitutionnel, la mécanique électorale, l'indépendance de la justice, la transparence de la vie publique et de l'administration, l'égalité et l'absence de discriminations, la protection de la vie privée et des données, la liberté d'expression, la liberté académique, la liberté de croyance et notamment de

---

<sup>46</sup> G. Koenig, *La fin de l'individu – Voyage d'un philosophe au pays de l'intelligence artificielle*, Éditions de L'Observatoire, 2019.

<sup>47</sup> Le témoignage le plus extrême de cette société de surveillance est peut-être celui-ci : à Pékin et dans d'autres villes chinoises, les distributeurs de papier de certaines toilettes publiques sont équipés de technologies de reconnaissance faciale afin de lutter contre les abus ; ainsi une même personne ne peut-elle utiliser plus de soixante centimètres de papier toilette dans un délai de neuf minutes — cette mesure est motivée par des considérations écologiques, donc louables, et il est en l'occurrence difficile d'avancer que d'autres motivations, moins légitimes, se cacheraient derrière.

<sup>48</sup> Dans le « Document 9 » de juillet 2012, un manifeste secret destiné aux cadres du parti communiste chinois qui prohibe expressément les analyses et opinions contraires à la ligne officielle, ce qui confine à une froide interdiction de toute opposition, y compris médiatique, la liberté de la presse est classée parmi les « sept périls occidentaux », aux côtés notamment des droits de l'homme et de l'indépendance de la justice.

religion, le droit à un traitement équitable, les droits des personnes issues des minorités, ainsi que divers droits économiques et sociaux.

Cela pose d'innombrables dilemmes moraux, notamment du côté des États-Unis. Reste que l'objectif des multinationales de l'IA et des données est d'inviter leurs utilisateurs à leur livrer le maximum d'informations personnelles les concernant, afin de pouvoir les cibler le plus finement possible, ainsi que de générer par des « *likes* » ou « *dislikes* » de l'engagement, car c'est cela qui génère de la *data*.

Malgré les différences fortes qui existent d'un point de vue juridique, éthique et idéologique, la Chine ne souhaite absolument pas se couper du monde et, tandis que certains GAFAM s'établissent sur son sol afin de profiter de certains de ses talents, la *Digital Silk Road* (« route de la soie numérique ») vise à connecter l'Union européenne avec la Chine à l'aide de différents types d'infrastructures, dont les satellites, la 5G et les câbles sous-marins. Ce projet pourrait permettre à la Chine, d'ici 2030, de devenir le « principal centre d'innovation en intelligence artificielle du monde, transformant le pays en chef de file mondial de l'innovation et faisant de lui la plus grande puissance économique », suivant les termes du *Plan national de la Chine pour l'IA*. Surtout, le gouvernement et la population chinois sont mieux préparés qu'ailleurs à la révolution des intelligences artificielles, parce qu'ils y voient le principal levier économique afin de devenir la première puissance mondiale et parce que ces technologies profitent généralement aux collectivités plus qu'aux individualités. Elles sont donc perçues comme collectivistes et utilitaristes, ce qui est en parfait accord avec la philosophie confucéenne — qui connaît d'ailleurs un fort renouveau en Chine.

Rien n'est plus égoïste qu'un État. Historiquement, un pays qui domine les autres technologiquement et économiquement a toujours été impérialiste. Même si la Chine annonce travailler aux progrès de l'humanité<sup>49</sup>, il est difficile de croire qu'elle n'ait pas de visées impérialistes. Le but d'une guerre n'est pas de conquérir des territoires mais de conquérir des populations. Longtemps, cela passait par l'invasion du territoire, mais, aujourd'hui, il devient possible de conquérir des populations à distance, grâce au *soft power*, en s'imposant dans les esprits et dans les pensées. Tout en demeurant dans la campagne de sa Franche-Comté natale, on peut devenir « sujet » de l'empire chinois en lui livrant ses données et en utilisant ses services.

Par ailleurs, les conséquences de l'IA qui préoccupent le plus ses observateurs et commentateurs sont peut-être celles qui touchent le travail. Tout d'abord, beaucoup de salariés humains se robotisent, deviennent de véritables machines qui répondent machinalement à des signaux, à des stimuli, perdant toute subjectivité, toute créativité, toute personnalité, toute

---

<sup>49</sup> Cf. Kai-Fu Lee, *I.A. La plus grande mutation de l'histoire*, Les Arènes, 2019.

force de proposition. Ensuite, les tâches nouvellement attribuées à l'informatique étaient pour la plupart dévolues à des humains, qui perdent en conséquence leurs emplois. Et rien n'assure qu'une redistribution efficace soit possible. Énormément de nouveaux métiers vont être créés, mais rien n'indique que les actifs d'aujourd'hui et de demain seront en capacité d'occuper ces postes très spécifiques tels que *data analyst* ou *data scientist*<sup>50</sup>. *Data scientist* serait pourtant le « métier le plus sexy du XXIe siècle », selon la *Harvard Business Review*. Mais le World Economic Forum s'inquiète du fait que les travailleurs, notamment français, seraient très mal armés pour s'adapter et affronter le changement. L'intelligence artificielle pose donc, à moyen terme et même à court terme, la question d'une aggravation de la situation du marché de l'emploi.

Dans le pire des scénarios, l'IA pourrait ne pas générer un grand remplacement des métiers et produire une masse de gens ne vivant, par exemple, que grâce au versement d'un revenu universel<sup>51</sup>. Afin de demeurer compétitif, à l'échelle de l'individu et à l'échelle de la société, dans un monde où l'IA, même faible, pourra effectuer toutes les tâches sans valeur humaine ajoutée (conduire un camion, tenir une comptabilité, traduire des textes ou des discours, diagnostiquer une maladie, par exemple), d'importants moyens doivent être consacrés au renouvellement et à l'amélioration de l'enseignement et de la recherche. Dans le cas contraire, le risque est celui d'un « apartheid intellectuel »<sup>52</sup>, l'intelligence artificielle ne permettant plus qu'à une frange de la population de travailler, parce qu'elle resterait dotée d'une intelligence supérieure et, par suite, d'une utilité économique et sociale.

S'il n'y a plus besoin d'hommes pour les métiers basiques, parce que l'IA s'en charge plus efficacement et à un moindre coût, plus rapidement et avec moins d'erreurs et d'approximations — il faut 20 ans pour « fabriquer » un comptable et 30 ans pour « fabriquer » un médecin, mais seulement quelques jours pour fabriquer un algorithme aussi performant que ce comptable et que ce médecin —, et si ne comptent plus que les capacités cognitives complémentaires de l'IA, alors il faudra non seulement inventer des métiers à haute valeur humaine ajoutée, mais aussi préparer les individus afin qu'ils puissent exercer ces métiers, former non des techniciens et des applicateurs

---

<sup>50</sup> Le *data analyst* et le *data scientist* doivent : traduire un problème business en problème mathématiques et statistiques ; identifier les sources de données pertinentes ; concevoir des « entrepôts de données » ; évaluer les données, les traiter et les résiter dans le système d'information cible. Ces professionnels combinent une triple compétence : expertise statistique et informatique, connaissance des bases de données et de l'informatique, expérience métier dans leur secteur d'activité.

<sup>51</sup> Il s'agit par conséquent d'une idée très libérale en ce qu'elle prend acte de l'impossibilité de trouver un emploi pour chacun, donc acte de la fin du travail.

<sup>52</sup> L. Alexandre, *La guerre des intelligences – Intelligence artificielle versus intelligence humaine*, JC Lattès, 2017.

mais des tacticiens et des créateurs, entre autres. L'homme devra se spécialiser dans les secteurs requérant des compétences multidisciplinaires et transversales<sup>53</sup>, où il y a peu de données et de statistiques mais un fort besoin de sens commun, de souplesse, d'adaptabilité, de créativité, de culture générale, de sens critique, d'empathie, d'affectivité, d'humanité, de faculté à innover, de capacité à prendre des décisions arbitraires, ainsi que, parfois, de force physique et de dextérité. L'IA est gratuite en valeur relative par rapport à l'intelligence biologique<sup>54</sup>, ce qui emporte des conséquences fortes sur le marché de l'emploi car, lorsqu'un bien est gratuit, ses substituts ne peuvent plus le concurrencer et disparaissent, tandis que les complémentaires deviennent beaucoup plus rechercher et gagnent donc en valeur. Ici, le travail complémentaire est le travail qualifié sous l'angle de la valeur humaine ajoutée.

La formation professionnelle pourrait jouer un rôle essentiel, si elle devient plus efficace, mieux adaptée, plus visible et mieux valorisée. Plus généralement, les parcours scolaires, éducatifs et universitaires ne doivent pas laisser passer le train de l'IA. Il est ainsi remarquable que le nombre d'étudiants inscrits au cours introductif à l'apprentissage automatique dans cinq prestigieuses universités américaines<sup>55</sup> a en moyenne plus que quadruplé depuis 2012.

---

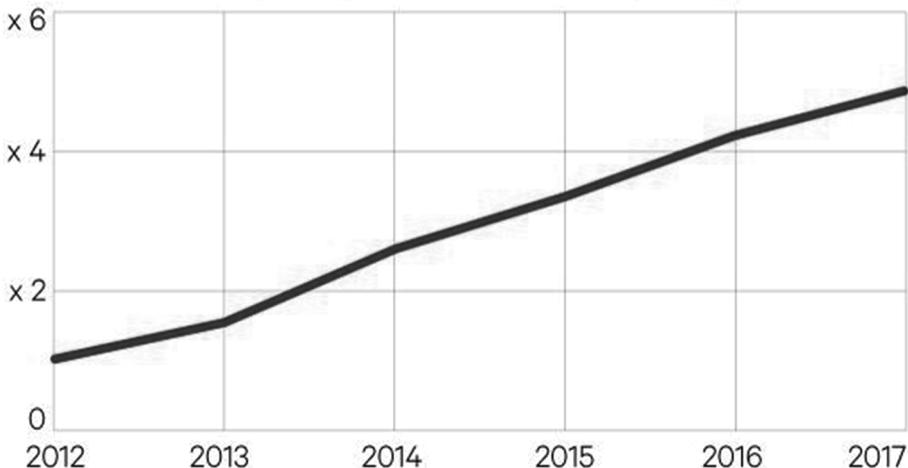
<sup>53</sup> L'IA peut concurrencer un radiologue mais pas un médecin généraliste. Face à l'IA, le généraliste garde toute sa place, au contraire du spécialiste.

<sup>54</sup> Mais il faut distinguer le prix des IA simplement informatiques, du secteur tertiaire, et celui des IA avec robotique en raison des coûts matériels et de fabrication souvent très élevés des robots. Le travail humain est beaucoup plus concurrencé par les premières que par les secondes. L'intelligence artificielle sert néanmoins le développement qualitatif et quantitatif de la robotique. Des robots thérapeutiques interviennent déjà dans certaines maisons de retraite. Des robots ménagers sont, quant à eux, présents depuis longtemps dans les foyers pour aider au quotidien. Des robots d'aide au déplacement permettent de pallier à certaines déficiences physiques. Et des robots-compagnons, sécurisants et divertissants, arrivent sur le marché, tandis qu'on devrait bientôt voir apparaître des robots majordomes au service des plus fortunés. Il faut aussi ajouter tous les robots non humanoïdes, tels que les véhicules autonomes ou les drones de livraison.

<sup>55</sup> Berkeley, University of Illinois Urbana-Champaign, Stanford, Carnegie Mellon et University of Washington à Seattle.

## Inscriptions aux cours introductifs d'apprentissage machine aux Etats-Unis

Evolution des inscriptions par année universitaire, en moyenne



\*LES ÉCHOS\* / SOURCES : AI INDEX 2018, BERKELEY, CMU, STANFORD, UIUC, UW

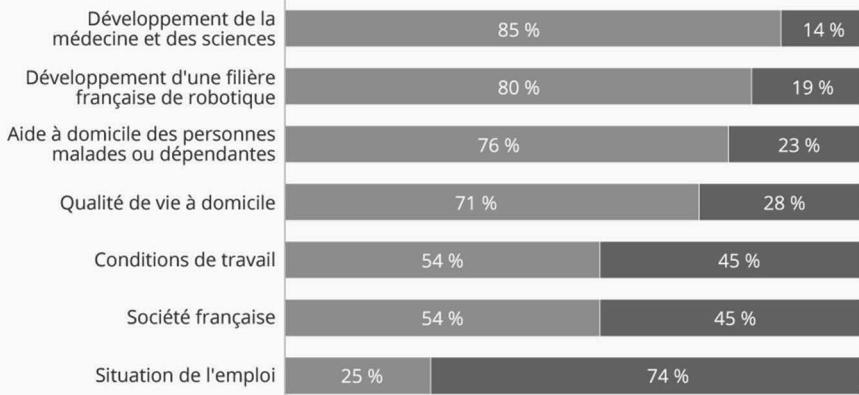
La question de l'emploi est l'une des principales clés de l'acceptation sociale de l'intelligence artificielle. Les révolutions industrielles ont toujours généré un moment d'angoisse généralisée. L'incertitude liée à la nouveauté et au brouillard entourant l'avenir, en premier lieu à un niveau micro-économique et à l'échelle de l'individu-salarié, provoque des inquiétudes. Mais les nouvelles technologies, jusqu'à présent, ont toujours bouleversé les modes de production sans pour autant entraîner des pertes d'emplois supérieures aux créations d'emplois. De nouveaux secteurs d'activité vont donner lieu à des métiers auxquels l'on n'a pas encore idée. Reste que l'adaptabilité et la souplesse, dans un tel contexte, seront les conditions d'une transition la moins chaotique possible vers l'économie des données et des IA.

Ce sont donc logiquement les conséquences du déploiement des IA sur la situation de l'emploi qui suscitent le plus de craintes, tandis que, à l'inverse, ses répercussions en matière de développement de la médecine et des sciences génèrent le plus d'espoirs :

## L'intelligence artificielle, chance ou menace ?

Avis des Français sur le développement de la robotique dans plusieurs domaines \*

■ Plutôt une opportunité ■ Plutôt une menace



\* Les points de pourcentage manquants correspondent aux "ne sait pas".

Source : Odoxa

statista

Si, dans l'ensemble, l'optimisme semble l'emporter sur le pessimisme en matière de perspectives de l'intelligence artificielle, il est remarquable que le deuxième domaine au sujet duquel les craintes du grand public sont les plus fortes est la « société » en général. L'IA n'est ainsi pas le sujet intéressant le moins la « collapsologie », ce « discours sur ce qui s'effondre », en l'occurrence la civilisation du XIX<sup>e</sup> et du XX<sup>e</sup> siècles et le « système global » en raison du réchauffement climatique et de l'épuisement des ressources. Elle concerne aussi les survivalistes, qui brandissent les menaces économique, climatique, civilisationnelle et technologique et qui se préparent à une très prochaine fin du monde — ou fin d'un monde. Les collapsologues rappellent que plus une société est inégalitaire, plus elle risque d'exploser. L'IA pourrait par conséquent être le terreau d'une très grave crise sociale. En tout cas pose-t-elle de redoutables questions éducatives, philosophiques et politiques.

La démocratie pourrait ne pas survivre au creusement de fossés au sein de la population, notamment en termes d'intelligence, de culture et d'éducation. La « trumpbrexisation » de la fin des années 2010 en serait un premier témoignage. La montée de la misère, y compris misère intellectuelle, s'accompagne toujours d'une poussée du populisme et des partis extrémistes. Or le danger parmi les plus ultimes de tous, mis en lumière par l'historien israélien Yuval Harari<sup>56</sup>, est celui de faire advenir un monde scindé en deux,

<sup>56</sup> Y. Harari, *Homo Deus – Une brève histoire de l'avenir*, Albin Michel, 2017.

avec des « hommes-dieux » d'un côté et des inutiles de l'autre, un monde gouverné par une petite aristocratie de puissants grâce à leur capacité d'innovation et à leurs moyens technologiques et un monde hostile pour les hordes de désœuvrés du numérique incapables de s'adapter aux changements de leur environnement. Le risque est d'arriver à une situation non plus de complémentarité entre les individus (intellectuels et manuels, directeurs et salariés, décideurs et exécutants) mais d'opposition entre une élite de l'intelligence et un prolétariat affaibli, démunie, abandonné, entre ceux qui voguent sur la vague des nouvelles technologies et les naufragés du numérique et de la société de la connaissance.

En outre, la démocratie est d'ores et déjà victime de la désintermédiation généralisée promue par les grands acteurs de l'économie des données et des IA, qui a pour effet de renforcer la défiance envers les représentants, envers les élus<sup>57</sup>. Et bien d'autres défis et problématiques se posent : atteintes à la vie privée, pollutions, hiérarchisation arbitraire et non journalistique de l'information, incompréhension du grand public qui ne prend pas garde à l'enfermement dans des bulles de filtre et à la mise en pilotage automatique des vies, biais dans les données qui se reflètent dans les résultats, donc discriminations en tous genres — l'intelligence de l'intelligence artificielle dépend assez exactement de la qualité des données qui la nourrissent et une base de données biaisée et discriminante produira des résultats symétriquement biaisés et discriminants ; l'intelligence artificielle est sous l'influence des catégorisations, préconçus et stéréotypes de ceux qui l'abreuvent en données, si bien que les problèmes peuvent venir moins des algorithmes et des réseaux de neurones que des masses d'informations non neutres qu'ils traitent<sup>58</sup>.

Cependant, si l'IA contribue à l'effondrement d'un monde qu'on pouvait croire stable et pérenne, elle participe aussi à la construction d'un nouveau monde — et peut-être cet aspect, constructif et positif plutôt que plaintif et négatif, est-il le plus intéressant. Il ne faut en tout cas pas exagérer l'un au détriment de l'autre. Il convient de n'être ni technophobe ni technophile mais « techno-lucide » et « techno-éclairé ». L'IA et les données pourraient être le prochain « cliquet malthusien », permettant à l'humanité de faire un grand bond en avant aussi bien d'un point de vue quantitatif que d'un point de vue

---

<sup>57</sup> De l'autre côté de la balance, l'objectif de ces acteurs est de garantir ou renforcer la confiance des utilisateurs dans leurs services, dont dépendent très largement leurs succès.

<sup>58</sup> C'est ainsi que l'on vit dans un monde où l'IA a envoyé des gens en prison parce qu'ils étaient noirs. De mêmes discriminations existent dans les recrutements. Par exemple, Amazon a développé un algorithme censé permettre de choisir les meilleurs candidats dans l'informatique et le candidat idéal s'est avéré être... un homme. On aboutit ainsi à des outils de recrutement qui disqualifient automatiquement toutes les candidatures féminines.

qualitatif, après l'agriculture, la révolution industrielle, l'exploitation du charbon, du gaz, de l'électricité ou du pétrole.

Dans son rapport « Donner un sens à l'intelligence artificielle », rendu public le 28 mars 2018, Cédric Villani pose la question des ministères qui sont ou seront impactés par les développements de l'IA et auxquels, par conséquent, son texte devait s'adresser en priorité. Sa conclusion est que tous les ministères sont concernés : économie, travail, éducation, enseignement supérieur, recherche et innovation, santé, environnement, mais aussi justice, culture, armée, police, transports, action et comptes publics, aménagement du territoire, ville et logement. Il est vrai que la plupart des basculements civilisationnels en cours sont liés, au moins indirectement, à l'IA : de l'Atlantique vers la zone Asie-Pacifique (de Shangaï jusqu'à la Silicon Valley) ; du neurone vers le transistor ; de la loi du Parlement vers la loi des algorithmes ; des États-nations vers les multinationales de l'économie numérique ; d'une économie de subsistance à une économie de « sursistance » (tuer la mort<sup>59</sup>, coloniser mars, augmenter le cerveau et l'intelligence humaine) etc.

L'intelligence artificielle n'est pas une simple innovation permettant de répondre autrement à un problème ponctuel et ancien. Elle constitue une véritable révolution en raison de sa transversalité — comme l'ont été les inventions du langage, de l'écriture, de l'imprimerie, de la télégraphie sans fil, de l'électricité ou de l'aviation — ; elle est le facteur clé d'autant de nouvelles solutions à explorer que de nouvelles difficultés à résoudre et de nouveaux défis à affronter. C'est pour cela que l'intelligence artificielle suscite à la fois chez certains un enthousiasme fanatique et chez d'autres une inquiétude horrifique. Mais cet enthousiasme et cette inquiétude demeurent relativement vagues et indéfinis, car il est aujourd'hui encore difficile de deviner dans quelles directions et en quels domaines les changements provoqués par les IA seront les plus importants et les plus radicaux.

Les plus grandes surprises proviennent d'ailleurs moins des nouvelles technologies que des usages sociaux ou économiques qui en sont faits. Les hommes fabriquent des machines qui, en retour, les reconstruisent. L'humanité sort forcément modifiée de la rencontre avec la technologie qu'elle a créée. De moins en moins de gestes du quotidien, d'achats, de déplacements, de décisions personnelles ou professionnelles ne sont pas dictées, plus ou moins intégralement, par des calculs et des statistiques. Avec un effet performatif fort, des IA aident à prendre des décisions ; et, souvent, elles les prennent elles-mêmes. Comme les inventions du microscope ou du

---

<sup>59</sup> L. Alexandre, *La mort de la mort – Comment la technomédecine va bouleverser l'humanité*, JC Lattès, 2011. Pour Google et sa filiale Calico, il s'agit d'« euthanasier la mort ». La première personne qui vivra mille ans est peut-être déjà née, scandent certains transhumanistes.

télescope ont permis de voir autrement le monde et l'univers, les IA et les capteurs numériques, toujours plus omniprésents, mesurant tout et prédisant tout, redéfinissent le regard de l'homme sur son monde et sur lui-même.

Cela concerne y compris, et d'ailleurs dans une large mesure, la recherche et l'enseignement, donc la formation et la transmission des pensées et des connaissances, ainsi que de l'esprit critique, du sens moral, de la capacité de discernement et de doute. Les chercheurs en sociologie, économie, philosophie, histoire, droit etc. doivent se saisir de ces révolutions qui dictent et dicteront l'avenir de leurs disciplines, du point de vue des objets d'étude comme du point de vue des moyens d'étude. Et, parce que l'IA concerne tout le monde de très près, ils doivent se mettre au service du grand public et ne pas se contenter de travaux inter-universitaires, en vase clos. C'est pourquoi ce livre, par exemple, a été conçu afin d'être accessible à tous et donc destiné à tous.

## IV. Une problématique scientifique : le besoin d'une science grand public de l'intelligence artificielle

Les contributions contenues dans cet ouvrage, rédigées par les meilleurs spécialistes de chaque question avec l'intention qu'elles soient les plus claires et synthétiques possible, interrogent l'intelligence artificielle dans toutes ses dimensions — à savoir sous un angle historique, sous un angle sociologique, d'un point de vue anthropologique, d'un point de vue économique, mais aussi sous un angle politique, en termes juridiques ou encore avec un regard socio-scientifique, philosophique ou prospectif. Par ailleurs, comme y invitent la collection Europe & Asie et les liens entre l'Université Paris 1 Panthéon-Sorbonne et l'Université Jiao Tong de Shanghai qui en sont à l'origine, quelques textes concernent la situation de l'intelligence artificielle sur le continent asiatique et plus particulièrement en Chine — pays qui constitue l'autre grand pôle de développement de ces technologies avec les États-Unis<sup>60</sup>.

Forcément non exhaustif et ignorant de ce que sera le futur, l'ouvrage collectif *l'intelligence artificielle – Dans toutes ses dimensions* en appelle beaucoup d'autres. On est seulement à l'aube de l'« ère de l'intelligence

---

<sup>60</sup> Pour ne prendre qu'un exemple, le projet de « crédit social » du gouvernement chinois vise à mettre en place un système national de réputation des citoyens. Chacun se verrait attribuer une note, son « crédit social », fondée sur les grandes masses de données collectées par le gouvernement à propos de leurs activités. Ce système s'appuierait donc sur des outils de surveillance de masse et sur des algorithmes permettant d'analyser les innombrables données collectées.

artificielle ». Cela rend le sujet passionnant et les recherches en la matière indispensables et, pour longtemps, inépuisables. En même temps, cela voudrait ces travaux à une obsolescence rapide. Tel est le lot de toutes les problématiques qui obligent à envisager un futur aux multiples inconnues et innombrables possibilités plus qu'à se rappeler un passé déjà fait ou à observer un présent en train de se faire. Il n'en faut pas moins accepter d'interroger l'avenir, ne pas s'y fermer ni chercher à transposer artificiellement les cadres du passé et du présent dans le futur — ou alors on risquerait de commettre un « crime contre l'avenir ». Si le sujet est plein de mystère, du point de vue de l'évolution de la discipline, de celui des ressorts scientifiques profonds des technologies ou de celui de leurs conséquences dans tous les domaines, les chercheurs et les penseurs ne doivent pas s'en effrayer. Il s'agit pour eux d'une chance immense, d'une opportunité formidable, car ils pénètrent dans un vaste champ encore inexploré.

Au-delà du vecteur technologique, directement observable et objet de nombreux commentaires et approfondissements scientifiques, la profondeur et les méandres de ses intrications et implications sociales méritent d'être explorés. Les innovations de la technosphère mettent toujours un temps plus ou moins long pour infiltrer et conquérir la sphère symbolique et y imposer leurs lois iconoclastes. Ce n'est qu'en de nombreuses décennies, voire des siècles, que l'homme moyenâgeux est devenu pleinement « gutenbergien », moderne. De combien de temps l'IA aura-t-elle besoin pour transformer l'homme moderne en homme postmoderne ? Et le pourra-t-elle seulement ? De façon générale, le regard porté sur les technologies reste par nature hypermétrope : on voit mieux de loin que de près, mieux l'imprimerie que la télévision, mieux la télévision qu'internet, et mieux internet que l'intelligence artificielle.

Une radiographie critique de l'intelligence artificielle et de ses conséquences n'en est pas moins un enjeu démocratique et social fort. Dans tous les secteurs de la connaissance, les chercheurs doivent s'efforcer à comprendre les dynamiques à l'œuvre, encourager les progrès de tous ordres qui peuvent en être tirés et prévenir les dérives dont l'intelligence artificielle risque de constituer le terreau. Puisque tout le monde est concerné et impacté, tout le monde devrait pouvoir comprendre le fonctionnement de l'intelligence artificielle et identifier les moments lors desquels on agit sous l'influence d'une proposition algorithmique. Pourtant, habités par un sentiment d'incompétence et d'éloignement, bien que les IA soient partout et tout proches, on préfère souvent se désintéresser de la question, considérant qu'elle ne serait qu'affaire de spécialistes. C'est pourquoi l'intention de ce livre est d'offrir à tous un moyen de mieux cerner l'intelligence artificielle et son impact sur la société, sur la culture, sur l'économie, sur la vie.

Parmi les plus concernés, on trouve les médiologues. La médiologie est l'étude et la pensée des effets des médias et de tous les instruments

relationnels. Cette discipline ne vise pas spécifiquement les médias compris comme presse, télévision, radio, web etc. « Médias » désigne ici ce qui se tient entre les personnes et permet de les relier, de les organiser. La médiologie s'intéresse donc aux milieux, cadres et guides, indissociablement sociaux et techniques, qui modèlent et remodèlent les représentations symboliques des hommes et leur permettent de tenir ensemble, de former un « nous »<sup>61</sup>. La médiologie souligne combien « exister, c'est être relié. Aucun organisme ne peut se développer durablement à l'écart des autres, au point qu'un réseau de bonnes relations semble la condition *sine qua non* de nos vies »<sup>62</sup>. Les médiologues ont dès lors forcément beaucoup à dire et à écrire au sujet des IA, nouveaux vecteurs des relations sociales et nouveau fil conducteur des croyances et comportements collectifs ou groupusculaires. L'influence de ces technologies sur la richesse des jeux de communication, actuels ou futurs, est forcément immense et même, dans une large mesure, révolutionnaire. Ne devra-t-on pas bientôt reconSIDérer entièrement l'écologie de l'esprit et de la pensée (l'étude des milieux dans lesquels l'esprit et la pensée se déploient, ainsi que de leurs rapports avec ces milieux<sup>63</sup>) ?

Pourtant, l'outillage technique de la communication est régulièrement négligé dans l'expertise des phénomènes symboliques et informationnels. Or, avec l'IA, l'influence de cet outillage sur ces phénomènes devient encore plus déterminante qu'auparavant. La forme et le contenu d'une culture et de la société qui repose sur elle dépendent étroitement de matériaux technomédiaTiques et, par suite, d'un inconscient technomédiaTique, *a fortiori* à l'ère de l'intelligence artificielle. On a toutefois tendance à privilégier les données culturelles et symboliques et à délaisser les outils, infrastructures et réseaux techniques qui les portent et les meuvent. On connaît mieux l'histoire de la littérature que l'histoire de la librairie ou de l'imprimerie, mieux les œuvres d'art que les métiers, musées et matériaux qui permettent de les montrer au public. Les contenus sont pourtant étroitement dépendants des contenants. L'impact des technologies d'intelligence artificielle sur les modes de pensée est tel qu'il devient impossible de continuer à aborder les phénomènes en cause en observant leurs sommets visibles et en ignorant les bases immergées sur lesquelles ils reposent. Ces dernières sortent cependant de plus en plus de l'obscurité pour se révéler au grand jour.

L'intuition de Marshall McLuhan se confirme : « *The medium is the message* »<sup>64</sup> — *i.e.* la nature du média (le canal de transmission) l'emporte sur le contenu du message et, finalement, redéfinit son sens et sa portée. Ainsi,

---

<sup>61</sup> R. Debray, *Cours de médiologie générale*, Gallimard, 1991.

<sup>62</sup> D. Bougnoux, *Introduction aux sciences de la communication*, La Découverte, coll. Repères, 2002, p. 18.

<sup>63</sup> G. Bateson, *Vers une écologie de l'esprit*, Le Seuil, 1977.

<sup>64</sup> M. McLuhan, *Pour comprendre les médias*, Points, 1968, p. 21.

selon le philosophe canadien, qui écrivait dans les années 1960, « en réalité et en pratique, le vrai message est le médium lui-même, c'est-à-dire que les effets d'un médium sur l'individu ou sur la société dépendent du changement d'échelle que produit chaque nouvelle technologie, chaque prolongement de nous-mêmes, dans notre vie. Toutes les technologies créent petit à petit un milieu humain totalement nouveau. Les milieux ne sont pas des contenants passifs, mais des processus actifs. À l'âge électronique, un milieu totalement nouveau a été créé. Le "contenu" de ce milieu nouveau, c'est l'ancien milieu machiniste de l'âge industriel. Le nouveau milieu refaçonnera l'ancien aussi radicalement que la télévision refaçonnera le cinéma. Le contenu de la télévision, en effet, c'est le cinéma »<sup>65</sup>. L'IA, en ce sens, se saisit de l'ancien monde (le monde pré-IA) et le refaçonnera à sa guise. La célèbre formule de Lavoisier « rien ne se perd, rien ne se crée, tout se transforme » trouve ici à s'appliquer. Tout ce qui existait avant l'IA (les besoins, les questions, les réponses, les problèmes, les solutions, les dangers, les opportunités, les loisirs, le travail etc.) existe et existera encore dans le monde de l'IA, mais sous un jour différent, très différent, voire radicalement différent. Peut-être qu'après avoir longtemps relégué au second plan la question de la technique communicationnelle et informationnelle, il se pourrait que les anthropologues, sociologues et autres politologues en viennent à être hantés par elle. Chacun ressent bien à quel point les nouvelles machines informatiques et numériques reconstruisent les formes du savoir, le lien politique, le sentiment social et la culture en général.

Les technologies médiatiques ne piégent néanmoins pas des sujets pensants qui, à l'état de nature, seraient parfaitement indépendants, lucides, vigilants et éclairés. Au contraire, ces outils sont indispensables pour réfléchir et se protéger contre les risques d'aliénation propres à cet état de nature. La moindre information requiert, pour son extraction, son traitement et son acheminement, une technologie. Mais les IA dépassent très largement ces fonctions primaires de la technique informationnelle et vont jusqu'à indiquer à chacun ce qu'il doit penser, quelles doivent être ses idées et ses centres d'intérêt, cela en fonction de motivations intérieures ou extérieures à l'individu. L'homme, s'auto-inspectant, dira encore longtemps « *ego cogito, je pense — c'est moi qui pense quand je pense, c'est moi qui parle quand je parle, c'est moi qui écrit quand j'écris* ». Ses songes, ses paroles et ses écrits lui paraîtront forcément innés, libres, jaillissant du plus intime de son être. Cet homme ressemble pourtant de moins en moins au *Penseur* d'Auguste Rodin, ce monument d'idéalisme qui montre l'être humain nu de tout artifice, secouru par aucun livre ni aucun écran, seul à penser ses pensées. On n'apprécie guère, et il arrive même que l'on ne puisse pas, interroger les prothèses techniques qui formatent les discours. Peut-on étudier un instrument au moyen de cet

---

<sup>65</sup> *Ibid.*

instrument lui-même ? Cela ne revient-il pas à couper la branche sur laquelle on est assis ? Et il est vrai que le propre des médias, à commencer par le corps, est de rester en arrière-plan et de jouer leur rôle, si ce n'est de façon invisible, en tout cas en se faisant oublier. Mais l'influence de certaines IA sur le paysage médiatique, communicationnel et informationnel est telle qu'il devient urgent d'étudier le médium plus que le message — même si Marshall McLuhan aurait certainement observé combien ils tendent à se confondre toujours plus.

Par exemple, les mécanismes de filtrage des réseaux sociaux, alimentés en continu par les préférences exprimées par les internautes, finissent par les enfermer dans une certaine vision du monde qui s'auto-entretient et donc se renforce, sans possibilité d'être confronté à une quelconque contradiction. Alors qu'internet devait offrir un accès formidable à la diversité et à l'infinité du monde, les algorithmes des GAFAM et autres services de réseautage social, à la recherche de toujours plus de clics, enferment les individus sur eux-mêmes. L'algorithme d'un grand réseau social détermine ainsi quels contenus de quels « amis » doivent être affichés en priorité et aboutit à une hyper-personnalisation faisant que, en permanence, les textes, les images, les vidéos, les liens et les publicités qui s'affichent vont dans le sens des choix habituels des utilisateurs. Cette « autopropagande invisible »<sup>66</sup> favorise notamment les phénomènes de radicalisation. Le pluralisme des courants de pensée et d'opinion et, par suite, la liberté d'opinion sont mis en danger. Peut-être s'agit-il d'une véritable « dictature des algorithmes » tant leurs effets normatifs ne répondent guère aux exigences démocratiques — loin du pluralisme externe de la presse (par la diversité des titres) et du pluralisme interne des chaînes de télévision et stations de radio (obligées par la loi de donner la parole à toutes les opinions sans défendre de ligne idéologique particulière). Les médias fonctionnant à base d'intelligence artificielle sont, eux, davantage des miroirs que des fenêtres.

C'est grâce à la pluralité des voix entendues, avec leur souhaitable cacophonie, et à la diversité des voies parcourues, avec leurs profitables chemins de traverse, que s'entretiennent l'esprit critique et la curiosité. Les chercheurs et les laboratoires savent à quel point la quête de la vérité a besoin de dispute et de contradiction et combien le plus grand danger est

---

<sup>66</sup> E. Pariser, *The Filter Bubble – What the Internet is Hiding from You*, Penguin Books, 2011. L'auteur donne l'exemple de deux personnes aux convictions politiques opposées qui recherchent au moyen d'un moteur de recherche des informations concernant une grande compagnie pétrolière. La personne de « droite » voit s'afficher des résultats se rapportant aux chiffres financiers et aux performances de l'entreprise, alors que celle de « gauche » obtient des résultats relatifs à la dernière marée noire et autres dégâts causés par cette compagnie. En affichant en priorité des éléments allant dans le sens de ce que les internautes aiment et pensent déjà, le moteur de recherche les emprisonne dans des « bulles cognitives ».

l'endormissement dans un sommeil dogmatique, dont il est parfois difficile de se réveiller pour aller s'abreuver à d'autres sources dont l'eau est *a priori* moins limpide. Emmanuel Kant, dans son opuscule *qu'est-ce que les Lumières ?*, a montré qu'il n'est pas de plus sûr critère pour évaluer la vigueur d'une démocratie que celui de sa presse et de son pluralisme. Or les IA médiatiques ruinent la presse véritable, subrogée par le règne de la dépêche, de l'immédiateté, du ciblage et de l'absence de signature, et réduisent à la portion congrue le pluralisme. Face à ces nouveaux « nouveaux chiens de garde », qui « se dévoient dans une société de cour et d'argent, en se transformant en machines à propagande de la pensée de marché »<sup>67</sup>, qui pourra stopper l'appauvrissement du débat public ?

La médiologie et, plus généralement, les sciences de l'information et de la communication ne peuvent donc qu'être fort préoccupées par les bouleversements subis par leurs objets en raison de l'intervention des IA médiatiques. Leurs spécialistes enseignent que « l'homme tient son humanité d'un certain régime symbolique, signifiant. Nous vivons moins parmi les choses que parmi une "forêt de symboles", comme dit Baudelaire, et ceux-ci nous rendent le monde familier en interposant entre lui et nous un ordre de signes, plus maniable et léger que celui des choses. L'empire des signes double ainsi notre monde naturel ; la sémiosphère contient la biosphère. Par tout un réseau de représentations codées et de signes qui sont autant de pare-chocs opposés à la dureté du monde, nous enveloppons, nous filtrons et du même coup nous maîtrisons le monde extérieur »<sup>68</sup>. Certainement l'influence des IA sur cet empire des signes, sur cette forêt de symboles, sur ce réseau de représentations codées, sur la sémiosphère et, par suite, sur la biosphère est-elle immense, voire radicale dans certains cas. Qu'on pense ne serait-ce qu'aux algorithmes qui filtrent les contenus du web ou, dans un tout autre ordre d'idée, aux robots humanoïdes et aux IA qui visent à améliorer les relations homme-machine par la détection des émotions, le dialogue et la robotique affective et interactive.

L'homme descendant ainsi du signe plus que du singe, il a toujours souhaité avoir à sa disposition des médias (des intermédiaires cognitifs et relationnels) permettant de l'ancrer dans son port d'attache plutôt que de l'abandonner aux caprices des vents contraires, de le stabiliser plutôt que de l'égarer dans le vagabondage, de lui donner un agenda plutôt que de le laisser face à l'inconnu, en bref de le fixer dans son espace et son temps propres, rassurants parce que familiers. Mais, de tout temps, l'état des technologies ne permettait pas une telle « pertinence » des informations et, par suite, la

---

<sup>67</sup> S. Halimi, *Les nouveaux chiens de garde*, Liber-Raisons d'agir, 1997, p. 159.

<sup>68</sup> D. Bougnoux, *Introduction aux sciences de la communication*, La Découverte, coll. Repères, 2002, p. 28.

balkanisation des médias. Avec les IA, au contraire, la personnalisation des contenus atteint son paroxysme. Et « ma culture » devient « ma clôture ».

Dans le royaume des *nudges*<sup>69</sup>, lorsque l'homme ne prend plus de décisions éclairées — même s'il prend peut-être les bonnes décisions —, qu'il ne comprend plus pourquoi il agit dans un sens plutôt que dans un autre, qu'il n'a plus entièrement conscience de ses actes, de leurs causes et de leurs finalités, il se robotise. Tel pourrait être le grand paradoxe de l'humanisation des robots au moyen de l'intelligence artificielle : robotiser les humains. Pour éviter cela, il faut comprendre et expliquer au plus grand nombre les algorithmes et les big data, donc les vulgariser, les démocratiser. Rarement le besoin d'une science grand public s'est fait autant ressentir car rarement des progrès scientifiques ont impacté aussi intimement le quotidien de chacun<sup>70</sup>.

Les intelligences artificielles sont forcément politiques et même idéologiques, jamais neutres. On fabrique ces extraordinaires outils afin qu'ils produisent des significations différentes, des cadres originaux, des routes nouvelles, dessinant le monde de demain, un monde tout autre de celui d'hier. Les IA façonnent un environnement technologico-politique particulier dans lequel chacun est amené à vivre, à penser et à juger autrement. La société pourrait n'être bientôt plus qu'une société de calculs, réduite en nombres, en fonctions mathématiques, en statistiques et en indicateurs. À partir du passé, les calculateurs définissent et organisent la réalité d'aujourd'hui et de demain. Ils produisent des conventions et des systèmes d'équivalence qui choisissent certains objets et certaines valeurs au détriment des autres, redessinant progressivement et insidieusement les cadres cognitifs et culturels des sociétés contemporaines. Les algorithmes, quand ils ne sont pas réglés afin de servir l'« autopropagande invisible » et enfermer chacun sur lui-même, cherchent à enfermer chacun dans certains environnements bien choisis, cadrer les conduites, formater les esprits, décider des désirs, standardiser les besoins, donc favoriser le suivisme, le panurgisme, et éliminer au maximum les envies et les modes de vie alternatifs — ceux-ci étant trop peu monétisables. Les utilisateurs se retrouvent donc placés dans des silos de comportement et de consommation tracés par les algorithmes. Quand la normalité devient normativité.

Tous ces calculs ne sont possibles que parce que les individus modifient leurs comportements afin de les rendre calculables. Le problème est qu'ils le

---

<sup>69</sup> En sciences du comportement, la théorie des *nudges* (ou théorie du paternalisme libéral) explique que des suggestions indirectes peuvent, sans forcer, influencer les motivations et les prises de décisions individuelles ou collectives souvent plus efficacement que des instructions directes, des normes, des lois.

<sup>70</sup> En ce sens, Cédric Villani a souhaité que son rapport de 2018 s'adresse et soit accessible à tout le monde, soulignant combien le sujet concerne et touche tout le monde.

font souvent de façon inconsciente et involontaire. Les technologies ne fonctionnent que parce qu'elles opèrent dans un milieu associé qui les rend efficaces et pertinentes. Pour éviter de se laisser conquérir par l'intelligence artificielle, afin de ne pas permettre à la dictature des algorithmes de s'imposer, et pour, au contraire, s'allier à ces technologies, il faut comprendre quelle est la situation, à quel point les robots peuvent suggérer et décider de beaucoup de choses. Ce n'est qu'ainsi que chacun pourra préserver sa souveraineté individuelle, veiller sur sa liberté d'autodétermination, demeurer maître de soi-même dans un contexte où les IA et certaines puissances (privées ou publiques) dont elles sont les instruments cherchent à limiter le libre arbitre et la diversité.

L'efficacité et le confort des intelligences artificielles, omniprésentes dans les décisions et les conduites quotidiennes, tendent à endormir le sens des responsabilités, l'autonomie et l'individualité<sup>71</sup>. Or ne faudrait-il pas conserver, pour reprendre une expression de John Stuart Mill, un « droit à l'errance », une capacité à dévier, à délibérer intérieurement, à innover, à s'autogouverner, à choisir et à se choisir, en somme à être libre<sup>72</sup> ? Revient l'ancestral débat libre arbitre contre déterminisme. En amenant les hommes à de moins en moins réfléchir, penser, garder en mémoire et prendre des décisions, l'intelligence artificielle risque de nuire à l'intelligence humaine. Cela conduit, par exemple, Éric Sadin à s'inquiéter du pouvoir injonctif des IA, lequel serait profondément antihumaniste<sup>73</sup>. À des fins de normalisation et

---

<sup>71</sup> G. Koenig, *La fin de l'individu – Voyage d'un philosophe au pays de l'intelligence artificielle*, Editions de L'Observatoire, 2019.

<sup>72</sup> Par exemple, à l'ère de la voiture autonome, obligera-t-on la population à rouler en voiture automatique-connectée ou acceptera-t-on encore les conducteurs traditionnels, bien que cela ne soit pas profitable à la communauté, à l'intérêt général, car ils provoquent des accidents et une gestion plus difficile du trafic, créant des situations non optimales pour le reste du groupe. Le premier choix est utilitariste, en faveur du collectif, le second est individualiste, protecteur des libertés individuelles en préservant la possibilité d'exprimer une personnalité singulière y compris sur la route.

<sup>73</sup> É. Sadin, *L'intelligence artificielle ou l'enjeu du siècle – Anatomie d'un antihumanisme radical*, L'Échappée, 2018. Le philosophe distingue quatre niveaux parmi les injonctions technologiques, selon leur plus ou moins grande capacité à décider à la place de l'humain : le niveau « incitatif » (avec l'exemple du GPS, qui laisse une certaine marge de manœuvre aux conducteurs), le niveau « impératif » (lorsqu'il s'avère plus difficile de ne pas suivre les indications des IA, à l'image de ces *chatbots* qui font passer des entretiens d'embauche et indiquent à l'humain, qui est encore dans la boucle, qui il doit recruter, ou des exemples des algorithmes des banques, de la justice prédictive ou de la police prédictive), le niveau « prescriptif » (notamment dans le cas de la médecine, où il est quasiment impossible de ne pas suivre les diagnostics de l'IA tant cela semble risqué) et le niveau « coercitif », le plus impactant, quand l'homme est totalement obligé par la machine, sans plus aucune liberté d'action propre (par exemple dans les domaines de l'économie et du travail où

d'optimisation, on formate et limite la faculté de jugement et d'action, on y substitue des protocoles informatiques afin d'infléchir dans un sens prédéterminé les actes, de réduire la liberté, de décider à la place de chaque individu. Ainsi s'agirait-il d'éviter les erreurs, la déviance, cela dans la veine d'une idéologie selon laquelle, Dieu n'ayant pas terminé la création, les hommes devraient être corrigés de leurs défauts, remis dans le droit chemin. Les nouvelles technologies seraient très utiles pour cela en permettant de parer à nombre de travers humains fort coûteux, notamment socialement.

Une telle vision utilitariste, collectiviste et même hygiéniste n'est évidemment assumée par personne. Elle paraît pourtant bel et bien empreindre pour une part l'environnement idéologique dans lequel les IA se développent. Celles-ci servent un mouvement global d'hyper-rationalisation des sociétés, dans lesquelles on veut à tout instant prendre les meilleures décisions afin de tout optimiser et tout maximiser. Au-delà, et pour la première fois de l'histoire, les vérités — avec leur valeur idéologique, distinctes des réalités matérielles — sont définies par des machines, par des technologies jouissant d'une immense puissance performative.

Il n'est pourtant pas rare, et même plutôt fréquent, que le merveilleux, l'agréable ou le profitable provienne de l'incertain, de l'aléatoire, du naturel, de l'instinctif ou de l'impulsif. L'ordre anarchique et l'ordre spontané peuvent paraître plus beaux, plus attirants, plus séduisants, plus agréables que l'ordre parfait entièrement programmé, mathématique, géométrique et logique — il suffit d'observer quels quartiers des grandes villes attirent le plus de touristes : il s'agit généralement des centres désordonnés et restés sous une influence moyenâgeuse, qui sont très peu pratiques mais qui présentent un charme certain. Le « droit à l'errance », foncièrement humain, est aussi un « droit à l'erreur », un « droit à la déviance », un « droit à la divagation » et un « droit à la sérendipité », lesquels sont des conditions du progrès. Le risque qui accompagne la standardisation des comportements sous l'influence des IA est celui de l'immobilisme<sup>74</sup> — par exemple, la *smart city* serait une ville morte, sans caractère et sans humanité. L'optimisation peut être source d'inertie. Cette inertie, sociale et politique, pourrait être le résultat de la diminution du pluralisme des hommes, de la réduction de la diversité de l'humanité. La pluralité et la liberté, sous toutes leurs formes, ne sont-elles pas les conditions essentielles de l'autonomie de jugement et, par suite, de la démocratie, de la politique, du progrès et de bien d'autres choses ? La prophétie technocratique

---

des mégadonnées sont analysées en temps réel et permettent à des algorithmes de passer des commandes à des manufacturiers et à un personnel qui ne peut alors que s'exécuter face à des contremaîtres insensibles et informels).

<sup>74</sup> Il a été montré que les chauffeurs qui utilisent le GPS toute la journée se robotisent et sont atteints d'une atrophie de certaines zones cérébrales à force de ne plus faire de petits choix et d'agir de manière automatique, par réflexes, tout au long de la journée.

saint-simonienne du « remplacement du gouvernement des hommes par l'administration des choses » pourrait-elle se réaliser dès lors que les hommes auraient évolué en choses, que le matériau pétři par le politique serait devenu inerte, que les gouvernants auraient face à eux des sujets sans inspiration et sans aspirations, pauvres en idées et en volonté, dépourvus d'esprit critique et incapables de proposer des projets alternatifs ou de soutenir des discours discordants ? Il importe de conserver la possibilité, à tout instant, d'inventer et de réinventer le monde et son monde autrement, à la fois à titre individuel et collectivement.

Tout cela interpelle l'historien israélien Yuval Harari, qui se demande si les applications industrielles de l'IA, qui en viennent à connaître les individus mieux que les individus eux-mêmes se connaissent, ainsi que les *nudges* de plus en plus omniprésents, ne conduisent pas à déléguer la capacité de choisir et de décider à des machines, donc à s'y soumettre dans des pans entiers de la vie individuelle, sociale et professionnelle<sup>75</sup>. La mathématicienne et militante américaine Cathy O'Neil déplore en ce sens que des « armes de destruction mathématiques » (« *Weapons of Math Destruction* ») décident des affectations des étudiants dans les universités, accordent les crédits, fixent les habitudes de consommation, attribuent les emplois et, surtout, déterminent les votes lors des élections<sup>76</sup>. Il serait donc impératif de prendre conscience et de se protéger de l'emprise et des dérives des algorithmes, rendues possibles par la connexion permanente des individus et par la puissance de calcul toujours plus importante des ordinateurs. Et ce sont les universitaires spécialistes de ces questions qui doivent jouer le rôle clé dans cette prise de conscience et cette protection.

Cela semble d'autant plus indispensable que les nouvelles technologies permettent beaucoup plus facilement de créer et maintenir des régimes totalitaires, bien que la situation ne soit pas tout à fait celle décrite par George Orwell dans *1984* : il n'y a pas, ici, de super-gouvernement qui contrôlerait les habitants de l'extérieur, la liberté d'expression ayant disparu et toutes les pensées étant minutieusement surveillées, tandis que des affiches placardées dans les rues indiqueraient à tous que « Big Brother vous regarde » (« *Big Brother is watching you* »). Il s'agit davantage d'un autocontrôle pénétrant jusque dans la plus stricte intimité, chacun fournissant les moyens de son

---

<sup>75</sup> Y. Harari, *Homo Deus – Une brève histoire de l'avenir*, Albin Michel, 2017.

<sup>76</sup> C. O'Neil, *Weapons of Math Destruction – How Big Data Increases Inequality and Threatens Democracy*, Crown, 2016 ; *Algorithmes – La bombe à retardement*, Les Arènes, 2018. Il faut toutefois noter que Cathy O'Neil cite essentiellement des exemples et anecdotes américains et que nombre des situations décrites sont supposées ne pas pouvoir se produire en Europe, où le droit est quantitativement et qualitativement plus protecteur des individus, notamment de leur vie privée et de leurs données personnelles.

propre contrôle<sup>77</sup>. Le résultat est cependant à peu près le même : l'avènement d'une société de la surveillance et de réduction des libertés. La performativité et les prophéties auto-réalisatrices des IA sont telles que, de fait, la technologie choisit pour l'humain qui s'en remet assez aveuglément à elle. Ce « coup de data permanent » enferme les individus dans des silos de comportement, de consommation, mais aussi de pensée. Cela porte atteinte à la liberté d'opinion et au pluralisme des courants d'idées. Se pose dès lors la question de la capacité de la société numérique à être une société démocratique et ouverte plutôt qu'une société tyannique et fermée. Mais chacun peut agir afin de changer les choses, à son échelle personnelle puis à une échelle collective. Et les chercheurs ont les moyens de devenir les moteurs de cette rétroaction contre le *diktat* algorithmique, non en manifestant une opposition de principe et catégorique mais en livrant au public les clés pour bien comprendre les mutations en cours, notamment en matière d'orientation des conduites individuelles et collectives.

Il faudrait n'abandonner aux robots que ce qui déshumanise et qui entrave pour retrouver le temps d'apprendre et de s'apprendre, de comprendre et de se comprendre, de connaître et de se connaître ; n'utiliser ces nouveaux outils que lorsqu'ils créent de l'indépendance et non de la dépendance ; et conserver ce qui libère, ce qui épanouit. L'intelligence artificielle, neutre en soi, pourrait aussi servir une grande transformation humaniste, comme l'écriture ou l'imprimerie, avant elle, ont permis de grands progrès de l'instruction et de la connaissance. Ne confondant pas le plus et le mieux, on passerait d'une civilisation du plus à une civilisation du mieux, réflexive, patiente et critique.

Au-delà de la compétition scientifique mondiale, l'intelligence artificielle est devenue l'affaire de tous. Dans ces conditions, les chercheurs doivent plus que jamais se mettre au service de la population, diffuser les connaissances, vulgariser ce qui présente la plus grande complexité. Leur importance est d'autant plus grande que, s'agissant d'un sujet déterminant pour l'avenir de l'humanité, il est nécessaire de pouvoir en débattre en toute indépendance, sans défendre des intérêts technologiques, politiques ou économiques. La question nietzschéenne « qui parle ? », à travers l'individu qui s'exprime, est incontournable et essentielle. Le grand public peut accéder à des informations objectives et à des avis éclairés lorsque le discours n'est pas orienté par de quelconques intérêts, ce qui peut être le cas de celui des chercheurs — mais ne l'est pas *ipso facto*, d'autant plus que des convictions, des idéologies ou

---

<sup>77</sup> Lors de son intervention au cours d'un colloque au Sénat sur l'intelligence artificielle le 19 janvier 2017, Laurent Alexandre expliquait ainsi que « nous sommes tous les “idiots utiles” de l'IA en mettant chaque jour des milliers d'informations au service des plateformes du numérique » (référence à Lénine qui aurait utilisé cette expression afin de désigner les intellectuels de gauche occidentaux défendant avec enthousiasme et naïveté le régime soviétique).

même ses propres écrits antérieurs sont d'autres formes d'intérêts. Au-delà, les scientifiques doivent également s'adresser aux gouvernants (publics et privés), afin de les aider à faire les bons choix, des choix structurants au moment où l'on prend des virages décisifs.

La grande aventure de l'intelligence artificielle est essentiellement un enjeu d'intelligence humaine : celle des chercheurs, celle des développeurs et des ingénieurs, celle des entrepreneurs et des investisseurs, celle des gouvernants, mais aussi celle de chaque personne qui, tous les jours, toutes les heures, à tout instant, agit à l'aune de propositions automatiques. Chacun doit pouvoir être acteur, et non seulement spectateur ou objet, de l'épopée de l'intelligence artificielle ; avec en toile de fond l'idée-guide selon laquelle le progrès technologique n'a de sens que s'il accompagne et stimule un progrès social et un progrès humain, que s'il bénéficie aux hommes et à leur capacité à s'organiser en société pacifique et prospère<sup>78</sup>.

Peut-être les Français et les Européens sont-ils ceux qui, dans le monde, attachent le plus d'importance à ces objectifs humains et sociaux. En témoigne le slogan anglophone attaché par le gouvernement français à sa stratégie en intelligence artificielle : « *AI for Humanity* ». C'est ainsi que, sur le « vieux continent », on a tendance à se focaliser sur la « robéthique », à investir dans des comités et des rapports éthiques, à se spécialiser dans les précautions morales et juridiques plutôt que dans les avancées technologiques et scientifiques, au risque de subir une « fuite des cerveaux » et une perte de puissance économique, technologique et politique<sup>79</sup>. Doit-on le regretter ? Si le monde se divise entre ceux qui créent des IA et ceux qui créent des comités d'éthique sur les IA, de quel côté vaut-il mieux se trouver ? Ces comités ont sans doute toute leur raison d'être s'il faut se méfier à la fois des innovations qui ne sont dictées que par l'appât du gain et de la politisation hystérique de la science, le tout dans un contexte mêlant, en fonction du point de vue adopté, crise de la rationalité et rationalisation extrême. Toute recherche qui avance sans se préoccuper de ses conséquences, directes et indirectes, peut être dangereuse et ne pas aller de pair avec le progrès.

L'Europe pourrait-elle être en retard en matière de technologie de l'IA et d'économie de l'IA, mais en avance en matière de pensée de l'IA et d'éthique

---

<sup>78</sup> En ce sens, il y a plus de puissance informatique dans un smartphone d'aujourd'hui que dans tout le projet Apollo qui a permis d'envoyer des hommes sur la Lune. Or personne, avec son smartphone, n'est capable de se rendre sur la Lune, car il manque l'intelligence collective, la volonté, la motivation, la détermination, autant de facteurs humains nécessaires au-delà de la technologie qui, en soi, ne peut rien.

<sup>79</sup> En ce sens, dans son rapport de mars 2018, Cédric Villani propose de créer « une instance pour émettre des avis, donner des jugements en toute indépendance, qui puisse être saisie par le gouvernement comme par les citoyens, et qui nous dise ce qui est acceptable ou non acceptable ».

de l'IA ? Sous bien des angles, celles-ci ne sont pas moins importantes que celles-là. L'IA a besoin de chercheurs travaillant sur les questions éthiques, réfléchissant à son rapport aux droits et libertés fondamentaux, à la Constitution et à la démocratie. Le siècle des intelligences nécessite un droit des intelligences. En témoignent ces groupes de plus en plus nombreux de juristes et d'informaticiens qui attaquent les projets de textes de droit liberticides du point de vue des « libertés numériques », à une époque où chacun se promène en permanence avec dans la poche un micro-ordinateur-espion muni de deux caméras, d'un micro et d'une ribambelle de programmes ultra sophistiqués de surveillance et de traçage. Le rapport entre le couple droit-politique et l'intelligence artificielle est d'autant plus important — et intéressant — que, si le droit et la politique saisissent l'IA (statut juridique, propriété intellectuelle, responsabilités etc.), il est tout aussi vrai que l'IA saisit le droit et la politique et, parfois même, fait le droit et la politique (*legaltechs, civictechs, performativité des algorithmes etc.*).

Mais peut-être, au-delà du droit, des normes, de l'éthique et de la régulation, des mobilisations individuelles et collectives, pour l'instant quasi-inexistantes, sont-elles nécessaires afin de prévenir certaines dérives de l'IA ou simplement mieux la comprendre, mieux la saisir de différentes manières. Éric Sadin invite en ce sens ses lecteurs à « mettre en crise » les nouveaux modèles en train de s'affirmer en refusant les actes d'achats de produits générateurs de mégadonnées (balance connectée, tétine connectée, réfrigérateur connecté, bikini connecté etc.)<sup>80</sup>. Et Régis Debray a souligné combien, tandis que la banalisation de l'automobile a tendance à réduire drastiquement la marche à pied, les individus qui ne marchent plus se mettent à courir, comportement spontané, sauvage, naturel, quand la marche est civilisée. Il a baptisé ce retour à l'archaïsme « effet jogging »<sup>81</sup>. Se pourrait-il que les développements de l'IA s'accompagnent de quelques « effets jogging » — à l'image de ces internautes qui consacrent leur temps à référencer des pages et des liens du web dans un univers numérique où les algorithmes des moteurs de recherche rendent déjà tous les services possibles et imaginables ?

Alors que les spécialistes de la technologie, dans la Silicon Valley comme ailleurs, peuvent se laisser absorber par les seules questions des performances

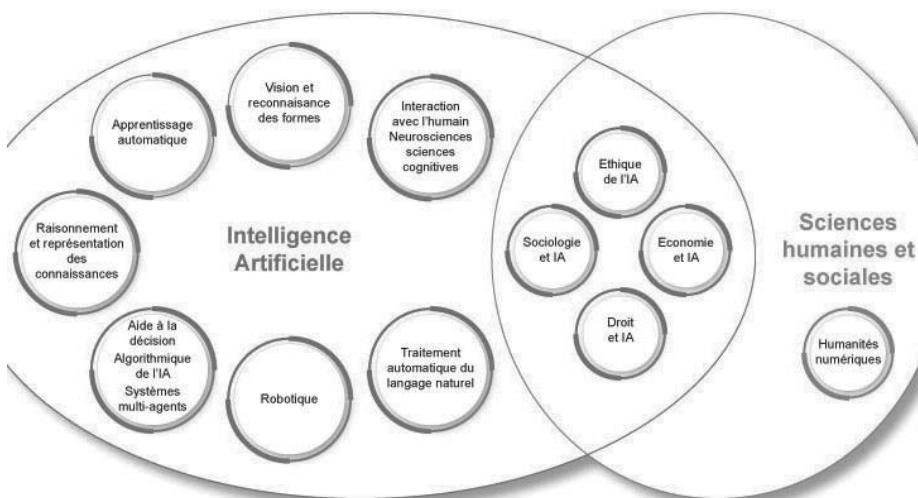
---

<sup>80</sup> É. Sadin, *La silicolonisation du monde – L'irrésistible expansion du libéralisme numérique*, L'Échappée, coll. Pour en finir avec, 2016.

<sup>81</sup> R. Debray, *Introduction à la médiologie*, Puf, coll. Premier cycle, 2000. Le philosophe observe ainsi qu'« au début du siècle, certains visionnaires avaient pronostiqué que l'usage immoderé de l'automobile par les citadins provoquerait bientôt l'atrophie de leurs membres inférieurs, le bipède motorisé se désaccoutumant de la marche. Qu'a-t-on vu depuis ? Ceci : depuis que les citadins ne marchent plus, ils courent. Fanatiquement. Dans les parcs ou, à défaut, en salle, sur tapis roulant ».

## *L'intelligence artificielle*

et de l'efficacité des outils qu'ils élaborent et qu'ils cherchent à améliorer, d'autres doivent interroger les conséquences sociales, économiques, politiques, éducatives, éthiques, scientifiques et parascientifiques (conséquences des progrès d'une science sur les autres sciences). L'IA, en soi, est bien sûr une affaire de mathématiciens, d'informaticiens, de neuroscientifiques et de roboticiens. L'économie, le droit, la sociologie ou l'anthropologie sont des sciences périphériques à l'IA. Elles n'en sont pas moins indispensables, cela afin de la comprendre mais aussi afin de la guider et même la diriger.



*C2RP, septembre 2017*

Ce besoin de scientifiques des IA d'origines diverses et uniquement motivés par la meilleure compréhension des phénomènes socio-technologiques à l'œuvre ou à venir deviendra toujours plus important avec le développement de l'homme augmenté et du transhumanisme. Pour l'heure, l'homme n'est que légèrement augmenté, avec un téléphone-ordinateur portable, une connexion internet permanente, une montre connectée, un thermostat intelligent etc. Mais, s'il en venait à utiliser des lentilles de contact connectées, à s'implanter des puces dans le corps et dans le cerveau, démultipliant ses capacités cérébrales, ou si des manipulations génétiques permettaient de ne plus donner naissance qu'à des enfants au quotient intellectuel très élevé, des questions éthiques autrement complexes se poseraient. La fusion de l'intelligence humaine et de l'intelligence

artificielle<sup>82</sup>, que certains appellent de leurs vœux, pourrait s'analyser tel un progrès ou telle une régression, si ce n'est tel le début de la fin de l'humanité. Et « IA » seraient les initiales du nouveau dieu, du nouveau guide suprême de l'humanité<sup>83</sup>.

« Homme augmenté », « intelligence augmentée », « *brain editing* », « *neurohackers* », « neuroprotecteurs », « neuro-intimité », paramètres d'implants intra-cérébraux, entrée du cerveau dans l'internet des objets, eugénisme technologique<sup>84</sup>... En pompiers pyromanes, certains sont tentés de fusionner avec le feu. Quelques garde-fous semblent alors ô combien utiles. Et les chercheurs semblent être les mieux en mesure de produire et défendre ces garde-fous, parce qu'ils sont (normalement) indépendants, parce qu'ils sont les mieux informés des enjeux et des potentialités positives comme négatives de la technologie qu'ils étudient, et parce qu'ils sont les mieux en mesure de prendre du recul — ce qui est indispensable pour prendre de l'élan et aller de l'avant.

Il n'y a cependant pas de chemin allant directement de la connaissance de ce qui est à la connaissance de ce qui doit être. Beaucoup de choix de société ne sauraient être dictés ni par des algorithmes, aussi perfectionnés et efficaces soient-ils, ni par des scientifiques, aussi objectifs et insubordonnés soient-ils.

---

<sup>82</sup> Le 28 avril 2017, Elon Musk a annoncé le lancement de Neuralink, une startup neurotechnologique qui développe des implants cérébraux d'interfaces neuronales directes. L'objectif est d'hybrider les cerveaux, notamment afin d'éviter que l'homme ne se laisse dépasser par la machine et devienne son domestique. Il s'agit donc d'aider les 83 milliards de neurones de l'homme grâce à des circuits miniatures permettant d'augmenter son quotient intellectuel et de le rendre plus compétitif face à l'IA. De son côté, Facebook a annoncé travailler sur une technologie de télépathie devant permettre de transférer un message d'homme à ordinateur ou d'homme à homme par la pensée, à la vitesse de 100 mots par minute.

<sup>83</sup> Aux États-Unis, Anthony Levandowski, ingénieur et père de la voiture autonome de Google, pensant que l'intelligence artificielle dépassera largement et en tout l'intelligence humaine, propose de lui vouer un culte. Il a ainsi fondé une organisation religieuse baptisée « *Way of the Future* » (« La voie de l'avenir ») qui fait la promotion d'une « divinité » basée sur une intelligence artificielle et qui est censée permettre le progrès de la société. Cela accrédite l'idée que le transhumanisme serait bien un discours religieux plutôt que scientifique, profitant de la tentation perpétuelle de l'humanité à vouloir se sauver par ses propres forces et de l'attrait pour le gnosticisme, qui consiste à voir dans le corps une entrave et à rêver de s'en débarrasser. L'âme serait détachable du corps et pourrait être mise sur un microprocesseur.

<sup>84</sup> En intervenant sur l'ADN, avec des manipulations génétiques, en augmentant technologiquement les hommes, on pourrait créer des hommes sur mesure, à la carte — par exemple de très bons musiciens ou sportifs, mais aussi des militaires sans peur, sans compassion, sans regrets et sans douleur. Peut-être est-ce plus effrayant que la bombe nucléaire. Que devra faire l'Europe si des pays, ailleurs dans le monde, décident d'exploiter ces possibilités technologiques et scientifiques ?

Bien sûr, toute morale est contingente et relative, variant dans le temps et dans l'espace, ainsi qu'entre les individus au sein d'une même population. L'appréhension morale de la technique est donc difficile. La question de l'IA et de la morale est même sans doute insoluble. Il y a forcément des conflits de morales qu'on ne peut pas trancher, si ce n'est arbitrairement, sans pouvoir soutenir son choix par une argumentation assurément supérieure. Tel est le cas, par exemple, concernant les accidents de voitures autonomes : qu'est-ce qui est le plus moral entre écraser trois seniors et écraser deux enfants ?

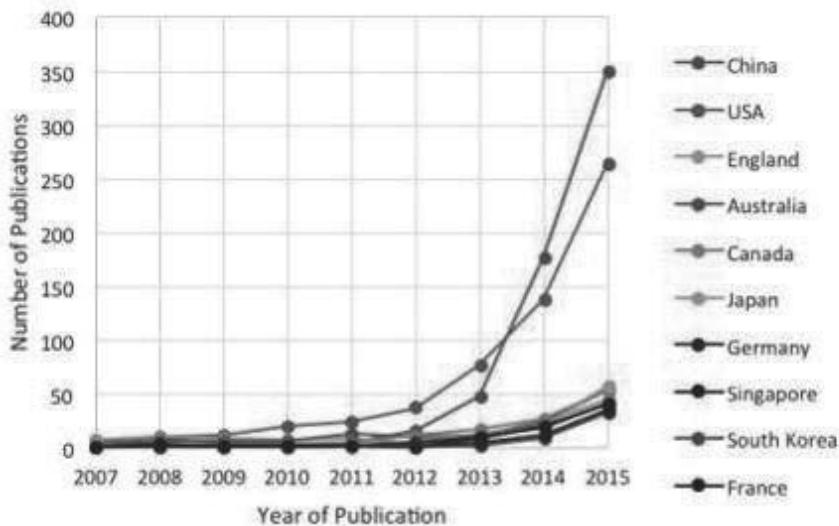
Ces difficultés ne sauraient pour autant justifier de se retrancher derrière les objets les plus objectifs et les problématiques les moins polémiques, ne sauraient inciter à se cantonner à la simple analyse technologique de l'IA. Formellement et matériellement, il faudra trouver les moyens de réguler et non laisser faire les robots, les hommes augmentés et le nouvel « *homo deus* ». On pourrait, par exemple, craindre la prolifération d'une forme de racisme cognitif : une discrimination et un rejet des plus intelligents et des augmentés envers les moins intelligents et les non augmentés. Et puis, avec la pantoprédictibilité et la dictée sociale des algorithmes, il faudrait prévenir la tentation des générations futures de tomber dans la datacratie, l'algocratie, le dataïsme et la datadépendance, c'est-à-dire l'enfermement dans un monde de chiffres, de statistiques et d'algorithmes déshumanisants et dénaturants. Peut-être mieux vaut-il conserver une grande part de hasard, lequel définit depuis toujours la vie et la marche du monde.

Les bioconservateurs pourraient-ils triompher et interdire ou supprimer les technologies ? Ou bien, tout à l'inverse, l'anarchie technologique pourrait-elle l'emporter, avec un développement tous azimuts uniquement motivé par la quête de nouveaux marchés, de nouveaux profits et de nouveaux monopoles, peu important que cela nuise à l'humanité et à l'intérêt général ? Peut-être la meilleure voie à suivre est-elle, comme souvent, celle de l'équilibre, ici en encadrant raisonnablement le pouvoir démiurgique de l'homme et ses relations avec l'IA. Au Canada, par exemple, on revendique des politiques en intelligence artificielle faisant la part belle à la liberté d'entreprise, comme aux États-Unis, tout en insistant sur les enjeux éthiques, comme en Europe. Le mot d'ordre est là-bas « *non predatory* », les progrès technologiques n'étant autorisés qu'à condition de servir l'intérêt public.

Pour aborder sans détours et sans biais idéologiques les dilemmes d'aujourd'hui et, sous un angle prospectiviste, agir sur le cours du futur en train de se faire, il semble nécessaire que la parole soit donnée aux savants — ou qu'ils la prennent. En ce sens, on ne peut que se réjouir du fait que, comme l'indique le rapport de Cédric Villani de mars 2018, la France est le quatrième pays au monde à produire le plus d'articles scientifiques sur l'intelligence artificielle, après la Chine, les États-Unis et le Royaume-Uni. Ces articles sont écrits par des mathématiciens, mais aussi par des spécialistes des sciences de

l’information et de la communication ou des sciences humaines et sociales<sup>85</sup>. L’explosion de l’intelligence artificielle s’accompagne donc — et c’est heureux — d’une explosion des sciences de l’intelligence artificielle.

Ce boom scientifique s’est produit à partir de 2012-2013, au moment du printemps de l’apprentissage profond et des réseaux de neurones :



**Figure 1: Journal articles mentioning “deep learning” or “deep neural network”**

L’avenir est investi de toutes sortes de peurs, manipulé par le personnel politique à des fins électoralistes, tandis que le politiquement correct interdit d’affronter (et parfois même de voir) certains problèmes et certains enjeux, en premier lieu en matière d’intelligence — quoique l’IA soit susceptible d’amener bientôt à remplacer le tabou de l’intelligence par une obsession de l’intelligence<sup>86</sup>. Le monde politique n’est de toute façon pas bien armé car il

<sup>85</sup> Mais ces travaux universitaires ne sont pas suivis de réalisations concrètes ni de développements économiques et le rapport de Cédric Villani constate que « la France dispose d’un savoir théorique indéniable avec des cerveaux recherchés à l’étranger, mais qui ne se traduit pas encore par l’émergence de grandes entreprises leaders ». Les États-Unis, la Chine, le Royaume-Uni, le Canada et Israël sont les pays les plus avancés en termes de déploiement de l’intelligence artificielle.

<sup>86</sup> Il n'est en effet pas politiquement correct de dire, et donc personne ne le dit, que ce sont les individus les plus intelligents qui prennent le pouvoir politique, économique, scientifique, technologique, financier etc. L'intelligence est la principale source de

n'a jamais eu à traiter des progrès exponentiels tels que ceux de l'IA<sup>87</sup> et parce qu'il fonctionne trop à base de démagogie et de court-termisme. Il en résulte un problème de désynchronisation et de rythme : l'IA et les technologies NBIC<sup>88</sup>, technologies de l'exponentiel, progressent très vite, quand les institutions, notamment politiques et éducatives, ne changent guère, apparaissant parfois même sclérosées. Dans ces conditions, il appartient peut-être aux chercheurs — des universités et au-delà — de prendre le relai et de se saisir pleinement de ces sujets, dans toutes leurs dimensions.

Leurs études doivent être substantiellement et sérieusement prospectives. En France, les travaux en prospective sont peu considérés et, par suite, peu développés. Rares sont d'ailleurs les prospectivistes-scientifiques, objectifs, honnêtes et non-fantaisistes, et la question des lendemains technologiques se retrouve presque en situation de jachère intellectuelle. Il faudrait pourtant pouvoir agir en fonction de ce que l'on veut et décider de ce que l'on veut en fonction de ce que l'on sait et non en fonction d'utopies ou dystopies, de mythes, de fantasmes ou de peurs infondées. En matière d'intelligence artificielle, l'écart entre ce qu'en montre Hollywood — et donc l'image que le grand public s'en fait — et ce qu'est sa réalité est aujourd'hui abyssal. Plus les futurs possibles seront précisément identifiés et réalistes, plus l'on pourra opérer un choix éclairé parmi eux et agir afin que le futur choisi advienne, donc prendre aujourd'hui les meilleures décisions et mesures politiques et juridiques. Cependant, le vrai pouvoir tend à échapper aux pouvoirs publics et à migrer vers les technologies, les infrastructures et donc ceux qui les possèdent. Aussi faudrait-il pouvoir intervenir à leur niveau plutôt qu'à celui des pouvoirs « officiels ».

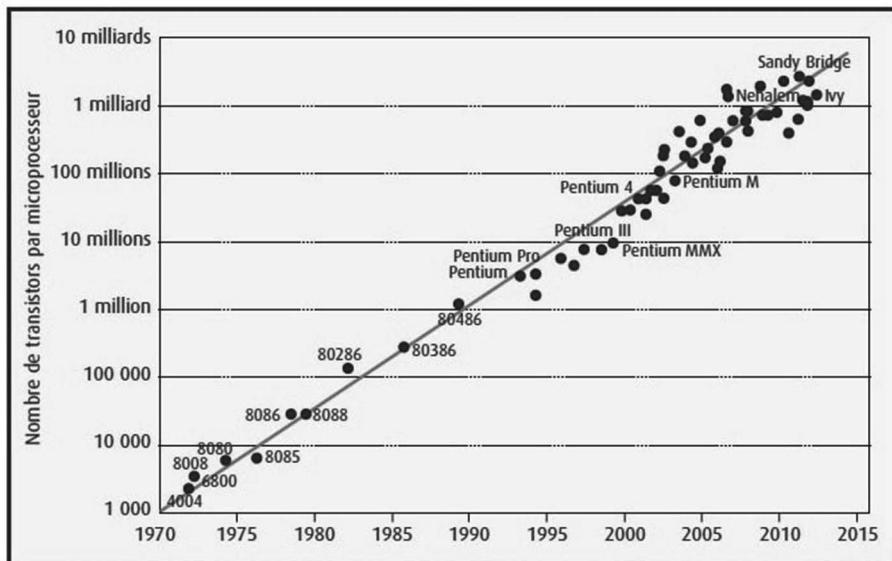
La prospective de l'intelligence artificielle est toutefois rendue difficile par le fait que les progrès exponentiels des technologies NBIC les rendent largement imprévisibles, y compris dans leurs conséquences sociales et économiques. Par exemple, le progrès soudain de l'IA et des réseaux de neurones à partir de 2012 n'avait été anticipé par personne. Les derniers processeurs comportent sept milliards de transistors et effectuent un milliard d'opérations par seconde contre une opération toutes les 30 secondes pour la calculatrice mécanique de Blaise Pascal. La « loi de Moore » pourra-t-elle se prolonger encore longtemps ?

---

pouvoir. Plus l'IA se développe, plus il devient impératif d'accepter de parler de l'intelligence, pour défendre, promouvoir et développer les spécificités humaines-biologiques-naturelles.

<sup>87</sup> Il a fallu 550 millions d'années pour passer du premier neurone au premier transistor, inventé en 1947, puis seulement 70 ans pour en arriver aux technologies d'apprentissage profond. La puissance informatique a été multipliée par 95 millions de milliards en 70 ans.

<sup>88</sup> Les technologies NBIC sont les nanotechnologies, les biotechnologies, l'informatique et les sciences cognitives.



Un autre frein à une meilleure appréhension du futur, qui est en même temps un défi à relever pour les scientifiques souhaitant s'adresser au grand public, est qu'aujourd'hui les mêmes canaux de communication diffusent les connaissances, l'information, la désinformation, les commentaires, les croyances et les opinions. D'ailleurs, y compris les chercheurs s'expriment sur les réseaux sociaux. Ces divers contenus, qui ont des qualités objectives et des statuts différents mais qui se confondent dans des conteneurs identiques, sont souvent placés au même niveau et assimilés les uns aux autres par le grand public. Ainsi des amalgames, des brouillages, des incompréhensions et de la malinformation se produisent-ils. Par conséquent, l'enjeu pédagogique n'a de cesse de se renforcer. Il est dommageable que les divers statuts des contenus se contaminent ou même fusionnent. Pour éviter cela, il faudrait mieux connaître nos connaissances afin de pouvoir les expliquer en cas de discussion avec quelqu'un qui oppose des discours non scientifiques mais jouant sur les sentiments et les peurs des populations pour convaincre, à base de discours spectaculaires et démagogiques.

Toujours est-il que Sergey Brin, co-fondateur de Google, en vient à promettre la création prochaine de « machines qui raisonnent, pensent et font les choses mieux que les êtres humains ». Si le réalisme de pareille prophétie n'est guère assuré, il est en tout cas certain que l'IA, ensemble de technologies fondamentalement transversal et dont les limites du potentiel restent inconnues, impacte et impactera l'ensemble des disciplines. Elle remodèle et remodèlera la politique, l'économie, l'éducation, l'industrie, la santé, la police, la culture, le journalisme, la philosophie, le droit etc. Elle oblige et

## *L'intelligence artificielle*

obligera à réinterroger ce qu'est et comment fonctionne une société ; et à réinterroger ce qu'est et comment fonctionne un homme.

En somme, l'intelligence artificielle est le grand sujet d'avenir et il faudra attendre longtemps pour que les recherches en la matière s'essoufflent. La problématique de la cohabitation entre intelligence biologique et intelligence technologique, entre intelligence naturelle et intelligence artificielle, est installée pour des décennies voire des siècles. Le plus passionnant est que tout ou presque reste à faire, à dire et à écrire. Le futur est incertain, vertigineux, extraordinaire. Pareille affirmation d'un nouveau champ scientifique immense et de telles perspectives multi- ou trans-disciplinaires, allant des sciences les plus « dures » aux sciences les plus « souples », et dont ce livre fournit un témoignage, sont des choses rares. Aux chercheurs d'en profiter et d'en faire profiter le plus grand nombre d'individus, le plus grand nombre de citoyens, le plus grand nombre d'hommes. La grande aventure de l'intelligence artificielle doit être et demeurer une grande aventure de l'intelligence humaine.



# L'intelligence artificielle

*Dimension historique*



# Une histoire de l'intelligence artificielle

Frédéric Fürst

L'intelligence artificielle (IA) est une idée polymorphe. S'agit-il vraiment d'un projet rationnel et réaliste ? N'est-ce pas plutôt un thème un peu éculé de la littérature et du cinéma de science-fiction ? Ou bien un des noms de l'utopie prométhéenne de maîtrise de la nature, voire son paroxysme ? Et si on considère l'intelligence artificielle comme une science, n'est-elle qu'une branche de l'informatique ? S'agit-il plutôt d'une discipline autonome, à l'intersection de la psychologie, de l'informatique, de la linguistique, de la neurologie et de la robotique ? Ou bien une simple application technique de travaux relevant de toutes ces disciplines ? Selon la nature qu'on prête à l'intelligence artificielle et les bornes qu'on lui fixe, on peut en raconter des histoires très différentes. Mais il est un moment qui cristallise toutes ces histoires et donne son nom à l'IA.

## I. La naissance de l'intelligence artificielle

En 1956, aux États-Unis, quatre scientifiques lancent un appel à toutes les bonnes volontés pour participer à une école d'été dont les travaux porteront sur ce qu'on appelle à l'époque les « machines pensantes ». Les deux instigateurs de cet événement sont John MacCarthy (1927-2011) et Marvin Minsky (1927-2016). Tous deux sont jeunes, 29 ans, ont une formation de mathématiciens, mais se sont rapidement passionnés pour cette idée de faire « penser » les machines. Ils travaillent dans des universités de Nouvelle-Angleterre, MacCarthy au Dartmouth College dans le New-Hampshire et Minsky à Harvard University dans le Massachussets.



*John MacCarthy*



*Marvin Minsky*

Ils se sont rencontrés en fréquentant la communauté des cybernéticiens, qui, dans le sillage de Norbert Wiener (1894-1964), étudient depuis les années 1940 le fonctionnement de l'esprit humain avec l'espoir de le répliquer dans les machines. Cet été de 1956, Minsky et MacCarthy vont baptiser un nouveau domaine de recherche, dont ils seront les animateurs, certains diront même les papes, pendant des années. Les deux autres signataires de l'appel, plus âgés, sont Claude Shannon (1916-2001), fondateur de la théorie de l'information, et Nathaniel Rochester (1919-2001), qui a inventé le premier langage de programmation assembleur. Leur renommée permet à Minsky et MacCarthy, inconnus à l'époque, de décrocher une aide de 7 500 dollars auprès de la fondation Rockefeller pour organiser le séminaire.

Le texte de l'appel, qu'on appelle depuis l'appel de Dartmouth, propose de rassembler pour les deux mois d'été, au Dartmouth College, ceux qui veulent

avancer dans la résolution du problème de l'intelligence artificielle. Cette expression forgée par MacCarthy désigne, dans le texte de l'appel et jusqu'à aujourd'hui, l'ensemble des techniques développées dans le but de simuler dans une machine tous les aspects de l'intelligence humaine. Des exemples de capacités à simuler sont donnés, en particulier l'utilisation du langage, la manipulation d'idées, la résolution de problèmes, la créativité. Mais les auteurs ne se hasardent pas à définir ce que veut dire exactement le mot intelligence.

L'intelligence peut être définie comme l'ensemble des capacités qui, quand un humain en fait preuve, le font considérer comme intelligent. Cette définition ne correspond cependant pas vraiment à l'intelligence artificielle. Ainsi, les travaux de recherche qui visent à écrire des programmes capables de conduire une voiture de façon autonome sont considérés comme relevant de l'IA. Mais personne ne pense qu'un humain qui conduit une voiture fait preuve d'intelligence. Inversement, un humain capable d'effectuer des calculs de façon très rapide sur des nombres très grands pourra être considéré comme possédant une intelligence hors-norme. Mais les ordinateurs nous battent à plate couture dans le domaine du calcul, sans pour cela utiliser la moindre technique relevant de l'intelligence artificielle. Le mot intelligence recèle un biais élitiste qu'il faut évacuer pour bien saisir l'objectif du projet lancé par MacCarthy et Minsky. Il ne s'agit pas de faire réaliser aux machines des tâches réputées difficiles pour les humains, ou réservées aux humains les plus « intelligents », mais bien de faire réaliser aux machines toutes les tâches dont sont capables les humains. Et, contrairement à certaines idées reçues, les tâches considérées comme les plus intelligentes ne sont pas forcément les plus difficiles à automatiser ; c'est ce qu'on appelle le paradoxe de Moravec. On sait déjà réaliser des machines qui battent tous les humains au jeu d'échecs, mais on ne sait toujours pas rendre un robot capable de balayer correctement une pièce poussiéreuse.

L'intelligence étant un phénomène mal défini et très vaste, l'objectif de l'intelligence artificielle est donc à la fois vague et extrêmement ambitieux. Il l'est d'autant plus que rien ne permet d'affirmer que simuler l'intelligence humaine dans une machine est théoriquement possible. Dans l'appel de Dartmouth, Minsky et MacCarthy se sont d'ailleurs sentis obligés de supposer qu'il est possible de décrire tous les aspects de l'intelligence de façon si précise qu'une machine peut les simuler. Ce postulat sera beaucoup discuté, et l'est toujours aujourd'hui. Mais, dans les années 1950, l'heure est à l'optimisme. Les premiers ordinateurs électroniques viennent de voir le jour et ils ouvrent des horizons jusque-là inconnus. Plus largement, le développement technique s'est emballé avec la Seconde Guerre mondiale et, de l'énergie atomique aux antibiotiques, des plastiques aux fusées, le champ des possibles ne cesse de s'élargir. Alors que les organisateurs prévoient de faire venir 10 personnes à Dartmouth, il y aura 20 participants.

Parmi eux, Alan Newell (1927-1992), mathématicien, et Herbert Simon (1916-2001), économiste, qui s'intéressent à la façon dont les humains prennent des décisions sans avoir toutes les informations pour évaluer parfaitement les différents choix possibles. Cette rationalité limitée, comme l'appelle Simon, est à l'œuvre dans la plupart de nos décisions, que ce soit pour démontrer un théorème mathématique, pour jouer aux échecs, pour conduire une voiture ou pour balayer une pièce. L'idée de Newell et Simon est que nos choix sont basés sur des heuristiques, des règles empiriques, qui peuvent se révéler fausses, mais qui nous guident généralement vers la meilleure décision.



Alan Newell et Herbert Simon

Par exemple, parmi les différentes files aux caisses d'un supermarché, on peut choisir la file où il y a le moins de clients. Généralement, c'est la file qui avancera le plus vite, mais il est aussi possible que cette file soit bloquée parce que la caisse tombe en panne, ou parce qu'un client ne trouve pas son portefeuille. Pour valider leur idée, Newell et Simon viennent de développer le *Logic Theorist*, un programme capable de démontrer automatiquement des théorèmes de logique mathématique, en se basant sur des règles heuristiques pour ne pas se noyer dans l'écheveau inextricable des raisonnements possibles. *Logic Theorist*, considéré depuis comme le premier logiciel d'intelligence artificielle, trouve tout seul les démonstrations de dizaines de théorèmes et, pour certains théorèmes, le programme a même découvert des démonstrations plus élégantes que celles données par les mathématiciens humains ! La présentation du *Logic Theorist* à Dartmouth conforte les participants dans l'idée que la réalisation d'une intelligence artificielle est à portée de main.

Ce séminaire estival ne produira pourtant pas d'idée nouvelle. Son intérêt a été de poser les bases d'une communauté de recherche, qui grossira ou

maigrira selon l'engouement ou la déception suscités par les projets de machines pensantes. Car l'IA est, sans doute plus que d'autres, un domaine chargé d'émotions et de fantasmes, d'espoirs et de peurs. Construire une machine qui se comporte comme nous est sûrement un rêve qui remonte à la fabrication des premiers appareils un peu complexes, comme les treuils ou les moulins. Aristote (384 av. JC-322 av. JC) imagine ainsi, plus de 300 ans avant notre ère, dans *La Politique*, des métiers à tisser ou des instruments de musique qui fonctionnent seuls, et l'amélioration des conditions de vie qui en résulterait. Mais, si l'idée des machines pensantes n'est passée de l'état de fantasme à celui de projet que dans les années 1950, ce n'est pas un hasard. Il fallait pour cela que plusieurs conditions soient réunies.

## II. La pré-histoire de l'intelligence artificielle

Pour imaginer réaliser une machine pensante, il faut d'abord être convaincu que la matière peut penser. Cette idée nous semble évidente aujourd'hui, mais, durant des millénaires, la plupart des sociétés humaines considéraient que les esprits de ceux qui mourraient survivaient à la destruction des corps. Il n'était donc pas possible d'imaginer que c'était le corps lui-même qui produisait la pensée. L'âme humaine relevait d'un autre monde, immatériel, et il paraissait aussi saugrenu, voire blasphématoire, de faire penser une machine que d'aller sur la Lune. L'impossibilité d'imaginer un voyage spatial résultait d'ailleurs aussi de l'idée selon laquelle les planètes n'étaient pas faites de roches ou d'eau comme la Terre. Les corps célestes qui parcourraient notre ciel étaient de nature divine, se mouvaient dans un espace différent du nôtre, donc inaccessible. De même qu'il a fallu des siècles pour qu'on admette que les planètes visibles ne sont que des sœurs de la Terre, il faudra attendre le XIXe siècle pour que le matérialisme s'impose dans le domaine de la pensée et qu'il devienne commun de considérer que l'esprit est un produit de l'activité du corps.

La deuxième condition nécessaire à l'idée d'une intelligence artificielle, c'est de disposer d'outils permettant de décrire la pensée, pour la simuler dans une machine. Nous devons à Aristote la première tentative de description de la pensée humaine, au IVe siècle av. J.-C.. Aristote ne croyait pas que la pensée était produite par la matière, mais cela n'empêchait pas de la décrire. Il a donc inventé des principes et des mécanismes de modélisation des raisonnements, qu'il a appelé Logique. Aristote considère que nous appréhendons le monde à travers des Catégories (les humains, les mortels, les vases etc), portant des propriétés (les humains sont mortels, les vases sont fragiles etc). Nos connaissances sur le monde s'expriment à travers des

propositions, comme par exemple « tout homme est mortel », qui combinent ces catégories et propriétés à l'aide d'opérateurs logiques, comme la négation ou la disjonction (le « ou »). En appliquant des règles de raisonnement à des propositions vraies, nous pouvons produire de nouvelles propositions vraies. Un exemple célèbre sont les syllogismes, du genre « Socrate est un homme et tout homme est mortel, donc Socrate est mortel ».

Bien sûr, les outils développés par Aristote sont très loin de permettre de décrire tout ce qui nous passe par la tête. Mais il était tellement en avance sur son temps que sa logique restera quasiment inchangée jusqu'au XIXe siècle. Elle sera étendue au XXe siècle au travers des logiques dites non classiques, qui, sans changer les principes de base de la logique aristotélicienne, lui ajoutent des valeurs de vérité (une proposition peut être autre chose que vraie ou fausse) ou des opérateurs (par exemple des opérateurs temporels qui permettent de tenir compte du fait que la validité d'une proposition peut changer au cours du temps).

Cependant, l'innovation majeure depuis Aristote dans la description de la pensée, du point de vue de sa simulation dans les machines, date de la fin du XIXe siècle. La logique était pratiquée jusque là en utilisant des langues humaines (le grec, le latin etc), bien trop complexes, ambiguës et verbeuses pour décrire de façon claire et rigoureuse les raisonnements humains. Le même problème s'est posé pour le calcul numérique. Sur les tablettes de Mésopotamie, les papyrus de l'Égypte antique et jusqu'aux parchemins du Moyen-Âge, les calculs étaient écrits en langage humain, et ce qu'on note maintenant simplement  $3 \times 4 / 5$  s'écrivait par exemple « je rajoute deux fois le nombre quatre à lui-même et je retire les quatre cinquièmes au résultat obtenu ». Heureusement, grâce aux chiffres arabes, diffusés en Europe au XIIIe siècle, puis à l'introduction de la notation algébrique par François Viète au XVIe siècle, l'écriture des calculs va être considérablement simplifiée.

Formaliser ainsi les calculs, c'est-à-dire les réduire à des manipulations de symboles, rend possible leur automatisation. En appliquant des règles simples, que nous avons apprises à l'école primaire, on peut additionner, soustraire, multiplier, diviser des nombres, ou résoudre des équations, en répétant des opérations élémentaires de façon purement « mécanique ». Cette approche, qu'on appelle algorithmique, du nom du mathématicien Al-Khawarizmi (~780-~850) qui l'a systématisée, va inspirer la formalisation de la logique. Pour rendre plus rigoureuse la pratique des démonstrations mathématiques et permettre de raisonner de façon non ambiguë et parfaitement vérifiable, le logicien Gotlob Frege (1848-1925) invente en 1879 un langage symbolique d'écriture de la logique d'Aristote. En notation moderne, le syllogisme cité au-dessus s'écrira  $(\text{homme}(\text{Socrate}) \wedge (\forall x \text{ homme}(x) \Rightarrow \text{mortel}(x))) \Rightarrow \text{mortel}(\text{Socrate})$ . Sous cette forme, il devient possible de mécaniser les raisonnements logiques, qui sont, comme les calculs numériques, réduits à des

manipulations élémentaires de symboles, organisées sous forme d'algorithmes.

On en arrive ainsi à la troisième condition nécessaire à la réalisation d'une intelligence artificielle : disposer d'une machine capable de simuler la pensée en manipulant mécaniquement les descriptions symboliques. La première machine réalisant des calculs numériques de façon automatique est due à Blaise Pascal, en 1642. Sa Pascaline effectue des additions et des soustractions en reproduisant dans des rouages mécaniques les manipulations de symboles que nous effectuons quand nous réalisons ces calculs à la main. Jusqu'au début du XXe siècle, les machines à calculer vont se perfectionner, mais sans permettre autre chose que le calcul numérique. Pour mécaniser également les raisonnements logiques, il faut aller au-delà, jusqu'aux ordinateurs.

Le coup de génie qui va mener aux ordinateurs est l'œuvre du mathématicien Alan Turing (1912-1954). En 1936, Turing effectue sa thèse de doctorat à l'université de Cambridge. Il travaille sur le Problème de la décision (*Entscheidungsproblem* en allemand), posé en 1922 par le mathématicien David Hilbert : peut-on toujours décider de façon parfaitement rigoureuse (donc mécanisable) si un énoncé est vrai ou pas ? Les préoccupations de David Hilbert concernant la rigueur en mathématique sont dans la droite ligne des travaux de Frege qui, quelques années avant, avait formalisé la logique d'Aristote justement pour donner aux mathématiciens des outils pour une pratique plus rigoureuse de leur métier.



*Alan Turing*

Quand Turing s'y intéresse, le Problème de la décision a déjà reçu une réponse négative, en 1931, l'année où Kurt Gödel a démontré son fameux théorème d'incomplétude : dans toute théorie mathématique, il existe un énoncé dont on ne peut démontrer qu'il est vrai, ni démontrer qu'il est faux. Certains énoncés ne sont donc pas démontrables. Mais une question reste pendante : comment démontrer de façon parfaitement rigoureuse les énoncés démontrables ? Turing propose en 1936 un modèle de traitement de symboles qui définit à la fois ce qui est calculable, au sens numérique, et ce qui est décidable ou démontrable, au sens logique. Ce modèle peut se résumer grossièrement à la lecture et l'écriture de symboles sur un ruban divisé en cases. Pour définir un traitement particulier, par exemple une addition, ou un syllogisme, il faut écrire un algorithme qui décrit la séquence adéquate d'écritures, de lectures et de déplacements le long du ruban. Ce modèle est appelé *Machine de Turing*, mais il ne s'agit pas d'une machine physique, simplement d'un modèle mathématique. La *Machine de Turing* est un mécanisme à la fois très simple et extrêmement puissant, tellement puissant qu'il couvre toutes les manipulations symboliques qu'on a pu imaginer jusqu'ici. L'hypothèse a même été émise, sous le nom de *Thèse de Church-Turing*, qu'il est humainement impossible d'imaginer un calcul ou une méthode de résolution de problème qui ne soit pas réalisable par une *Machine de Turing*. Et la simplicité du modèle de Turing permet de construire des machines réelles, qui effectuent automatiquement n'importe quel traitement mécanique de symboles. Ces machines, qu'on appelle ordinateurs, ne font basiquement que lire et écrire des symboles dans une mémoire électronique découpée en cases. Tous les ordinateurs ne sont, ni plus ni moins, que des incarnations de la *Machine de Turing*.

Le travail d'Alan Turing est si fondamental qu'il est considéré comme le père de l'informatique et que l'équivalent du « prix Nobel de l'informatique » a été nommé *prix Turing*. Mais son rôle dans l'histoire de l'intelligence artificielle ne se réduit pas à avoir inventé le modèle des machines permettant de « simuler n'importe quel aspect de l'intelligence humaine ». Convaincu que le fonctionnement de son modèle reflète le fonctionnement de l'esprit humain, Turing imagine que les ordinateurs naissants vont rapidement devenir aussi intelligents que nous et se pose la question de savoir comment déterminer quand ce stade sera atteint. Il invente pour cela un test qui repose sur le principe qu'une machine est aussi intelligente qu'un humain si et seulement si elle est capable de se faire passer pour un humain, sans qu'on puisse déceler la supercherie. Le *Test de Turing* se limite aux aspects intellectuels, car il suppose que l'humain qui teste la machine communique avec elle textuellement grâce à un clavier et un écran, ou oralement par téléphone. Cette idée est très proche de la définition que donneront plus tard MacCarthy et Minsky. Une machine capable de simuler n'importe quel aspect de l'intelligence humaine pourra facilement se faire passer pour un humain.

Inversement, pour se faire passer pour un humain, il faut nécessairement pouvoir en simuler tous les aspects.

Ainsi, dans les années 1950, toutes les conditions pour envisager la réalisation de machines pensantes sont réunies :

- (1) l'idée que la matière peut produire une pensée ;
- (2) une description formelle de la pensée, limitée cependant à certains raisonnements logiques ;
- (3) des machines permettant de faire tourner ces modèles de la pensée.

Ces trois conditions, nécessaires à l'émergence de l'IA, ne sont cependant pas suffisantes, très loin de là, pour réaliser effectivement une machine aussi intelligente que nous. Les pionniers du domaine allaient rapidement le constater. Pourtant, à cette époque, les machines intelligentes se banalisent dans la littérature de science-fiction. À partir des années 1940, Issac Asimov (1920-1992) développe dans ses nouvelles et romans le thème des robots humanoïdes, inventant même le terme de robotique. Philip Dick (1928-1982) commence à écrire, dans les années 1950, sur le même thème des machines à apparence et comportement humains, qu'il appelle des androïdes.

Il faut dire que l'informatique n'en est alors qu'à ses balbutiements. Dans ce domaine, tout reste à faire, donc tout paraît possible. Les idées fusent et rapidement de nombreux exemples de programmes voient le jour, capables de simuler certains aspects de l'intelligence humaine, en particulier la réalisation de démonstrations mathématiques et la traduction de textes. En outre, la guerre froide vient de débuter, et le gouvernement américain finance largement tous les projets susceptibles d'applications militaires. Les fusées de la conquête spatiale sont d'abord mises au point comme missiles intercontinentaux. Les premiers projets de traitement automatique des langues (TAL) seront consacrés à la traduction automatique du russe vers l'anglais, ou inversement. Par contre, si l'URSS va s'investir tout autant que les USA dans la conquête spatiale, il faut remarquer que les recherches en IA resteront pratiquement cantonnées aux États-Unis jusqu'à la fin des années 1960. Et encore, ces recherches seront menées essentiellement dans une poignée d'universités de Nouvelle-Angleterre et, à partir des années 1960, de Californie.

La principale raison à cela, c'est le manque d'ordinateurs et d'informaticiens. Les machines de l'époque occupent chacune une salle entière, sont très chères, et de surcroît tombent souvent en panne. Il existe très peu de spécialistes capables de les utiliser, de les réparer et de les programmer. On ne trouve donc des ordinateurs que dans certaines universités assez riches pour se les payer, dans des administrations, et dans quelques grandes entreprises, dont, bien sûr, celles qui les fabriquent, surtout IBM. Comparé à nos tablettes et smartphones du XXI<sup>e</sup> siècle, les ordinateurs de l'époque nous paraissent extrêmement rudimentaires et presque inutilisables. Mais, dans les

années 1950, disposer de machines capables de réaliser n'importe quel traitement algorithmique, même à des vitesses qui nous semblent aujourd'hui insignifiantes, relève presque de la magie. Une ambiance d'optimisme débridé, des financements abondants, des premières réussites encourageantes, les chercheurs en intelligence artificielle vivent alors une période bénie. La quinzaine d'années qui sépare 1956 du début des années 1970 sera appelée « âge d'Or de l'IA » et verra éclore une grande partie des idées qui sont aujourd'hui encore à la base des recherches en intelligence artificielle.

### III. L'âge d'Or de l'intelligence artificielle

Dès les débuts des travaux en IA, on voit émerger les deux grandes approches utilisées pour rendre les machines intelligentes. L'approche symbolique, héritière d'Aristote et de Frege, vise à décrire le fonctionnement de l'esprit humain grâce à des modèles logiques. C'est une approche « par le haut », qui veut simuler les manifestations conscientes de la pensée, sans se soucier du mécanisme matériel qui la produit. Le premier exemple en est le *Logic Theorist* de Newell et Simon. Ces deux-là vont généraliser leur approche et développer en 1957 le *General Problem Solver*, un programme censé pouvoir résoudre n'importe quel problème, du moment qu'il est formalisé correctement.

La deuxième approche utilisée en IA, qu'on appelle connexioniste, est une approche « par le bas », issue de la cybernétique. Son principe est de simuler dans les machines le fonctionnement du mécanisme matériel qui produit la pensée, c'est-à-dire le cerveau et ses neurones. En 1943, les neurologues Warren McCulloch et Walter Pitts introduisent le neurone formel (ou artificiel) comme modèle des neurones biologiques : il s'agit d'une fonction mathématique qui prend plusieurs nombres en entrée et calcule une valeur de sortie. Les nombres en entrée et en sortie sont l'équivalent des potentiels électriques en entrée et en sortie des neurones biologiques. Le premier modèle de réseau de neurones artificiels, le Perceptron, où plusieurs neurones formels sont connectés et fonctionnent ensemble, a été élaboré en 1957 par Franck Rosenblatt (1928-1971).



*Franck Rosenblatt*

Mais, de même que les modèles symboliques ne reflètent qu'une part très réduite de notre activité intellectuelle, les réseaux de neurones artificiels ne sont que des simulations très simplifiées du cerveau. Alors que, dans nos têtes, les neurones sont câblés de façon très complexe, les réseaux de neurones sont organisés de façon très simple, en couches parallèles où chaque neurone d'une couche envoie sa sortie à tous les neurones de la couche suivante. La transmission d'une valeur de sortie d'un neurone sur l'entrée d'un autre est modulée par des paramètres qui peuvent faire augmenter ou diminuer la valeur transmise. Ces paramètres sont l'équivalent des neurotransmetteurs, les molécules qui, entre deux neurones biologiques, inhibent ou facilitent le passage des flux électriques.

Dans le cerveau humain, plus on pense à quelque chose, plus le taux de neurotransmetteur augmente le long du trajet neuronal correspondant à ce souvenir, et plus facilement on pourra y penser de nouveau. Dans un réseau de neurones artificiels, on va modifier les paramètres de transmission entre neurones pour faire « apprendre » quelque chose au réseau. Le Perceptron est un réseau à une seule couche, mais il permet déjà de simuler le phénomène d'apprentissage et surtout d'oubli : à force de modifier les paramètres pour apprendre de nouvelles valeurs, le réseau « oublie » les anciennes valeurs. Cependant, en 1957, on ne connaît pas d'algorithme permettant de réaliser cet apprentissage artificiel dans des réseaux possédant plusieurs couches de neurones. Cette difficulté théorique du connexionisme et la réussite éclatante du *Logic Théorist* vont pousser les pionniers à privilégier l'approche symbolique.

Les systèmes de raisonnement symbolique ont malheureusement plusieurs défauts. Un premier problème, que Newell et Simon ont tenté de régler avec leurs règles heuristiques, est celui de l'explosion combinatoire. Dans un système de raisonnement logique, même avec peu de règles de raisonnement, si on applique toutes les règles à tous les faits connus au départ, puis toutes les règles à tous les nouveaux faits produits, et ainsi de suite, le nombre de chemins de raisonnements possibles devient rapidement ingérable, quelle que soit la puissance de l'ordinateur utilisé. Ce phénomène se produit par exemple dans le jeu d'échecs : malgré le peu de pièces et de cases dans ce jeu, il est totalement impossible d'explorer toutes les parties réalisables à partir d'une position donnée, sauf dans les fins de partie quand il ne reste plus que quelques pièces sur l'échiquier. Pour déterminer le meilleur coup à jouer, il est donc nécessaire de négliger certains coups possibles en recourant à des heuristiques, du genre « ne jamais prendre en considération un coup qui conduit à perdre la dame » (même si, parfois, sacrifier la dame mène à la victoire).

Les systèmes logiques posent un problème encore plus fondamental. Les utiliser pour simuler tous les aspects de l'intelligence humaine suppose qu'on soit capable de modéliser logiquement tous ces aspects. Or, la plus grande part de notre intelligence est, pour le moment du moins, impossible à représenter ainsi. Comment représenter les émotions, les mécanismes inconscients ou même nos mouvements de marche bipède avec des formules logiques ? Cette approche logique va cependant trouver un terrain favorable dans un domaine facilement formalisable : les mathématiques. À la suite du *Logic Theorist*, le premier programme de démonstration automatique de théorèmes de géométrie est produit en 1958 par Herbert Gelernter. En 1961, James Slagle, un étudiant de Minsky, développe un programme capable de traiter des équations intégrales. Le programme est capable de réaliser les exercices de manipulation d'équations demandés aux étudiants de première année du Massachusetts Institute of Technology (MIT). Automatiser le travail mathématique est sans doute la seule vraie réussite de l'IA dans les années 1950-1960. Elle débouchera dans les décennies suivantes sur les logiciels de calcul formel, qui sont actuellement des outils largement utilisés par les mathématiciens pour faciliter les calculs symboliques et les démonstrations de théorèmes.

Il paraît donc possible de doter les machines de nos capacités de démonstration mathématique, au moins partiellement. Mais, dans beaucoup d'autres domaines, nous ne savons pas écrire d'algorithme qui permette à une machine de simuler notre comportement. En 1959, Arthur Samuel (1901-1990) écrit CHECKERS, un programme de jeu de dames. Il n'est pas très compliqué de faire en sorte que le programme respecte les règles du jeu. Mais Samuel ne sait pas comment faire en sorte que le programme choisisse des coups gagnants et évite les coups perdants, comme le font les bons joueurs humains. Pour contourner ce problème, il invente l'apprentissage artificiel (*machine learning*).



*Arthur Samuel*

Il donne à son programme la capacité d'analyser les parties jouées, pour construire automatiquement un modèle statistique des coups qui permettent de gagner et de ceux qui mènent à la défaite. Au départ, CHECKERS est facile à battre, mais, après avoir joué des milliers de parties contre des humains et surtout contre lui-même, il est capable de battre la grande majorité des humains auxquels il est confronté. L'apprentissage artificiel apparaît ainsi comme une technique permettant de doter les machines de n'importe quelle capacité humaine, du moment que cette capacité peut s'exprimer sous forme de données à analyser.

Une autre technique de base de l'IA naît en 1959, lorsqu'Oliver Selfridge (1926-2008) introduit le principe des systèmes multi-agents : un comportement complexe peut résulter des interactions entre une multitude d'agents aux comportements simples. Ainsi, dans une fourmilière, chaque fourmi réalise une tâche simple, qui concourt à la vie et au développement d'un système très complexe. Selfridge utilise ce principe dans son logiciel PANDEMONIUM d'analyse d'image : des programmes simples détectent des lignes, d'autres utilisent ce travail pour détecter des formes composées de plusieurs lignes, et le programme de plus haut niveau peut détecter des objets complexes comme des maisons ou des voitures.

Une autre idée inspirée par la biologie est celle de la programmation génétique, introduite en 1966 par Lawrence Fogel (1928-2007). Quand on ne sait pas écrire un programme qui réalise correctement une certaine tâche, on va faire évoluer une population de programmes qui réalisent cette tâche de façon partielle, en utilisant les mécanismes de la sélection darwinienne : à chaque étape, on ne garde que les programmes qui donnent les meilleurs résultats, on les croise pour obtenir une nouvelle génération de programmes, et on recommence jusqu'à obtenir un programme qui réussit parfaitement la

tâche visée. La programmation génétique peut être vue comme un cas particulier d'apprentissage artificiel.

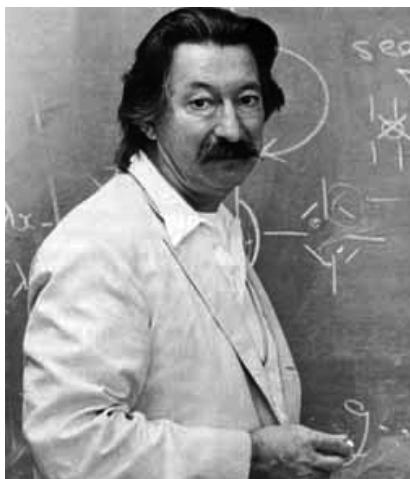
Certains chercheurs s'intéressent plutôt à la simulation de capacités cognitives particulières. Par exemple, en 1957, le linguiste Noam Chomsky (né en 1928) introduit la notion de grammaire formelle pour rendre compte de façon logique des structures de toutes les langues humaines et pouvoir faire manipuler ces langues par les machines. L'utilisation de ces grammaires va rapidement se heurter à l'incroyable complexité de la communication humaine : pour comprendre le sens d'une phrase, il ne suffit pas d'analyser sa structure et les types des mots qui la composent. Il faut aussi connaître le contexte de la phrase (les phrases qui l'ont précédée, le ton de celui qui parle, sa culture etc). Un exemple fameux est celui de la traduction automatique d'anglais en russe, puis en anglais, de la phrase *The spirit is willing but the flesh is weak* (« l'esprit est fort mais la chair est faible »). En sortie, la phrase est devenue *The vodka is strong but the meat is rotten* (« l'eau-de-vie est forte mais la viande est pourrie »).

L'idée qu'un mot ne prend sens qu'en lien avec d'autres, et que les notions (au sens des catégories d'Aristote) se définissent les unes par rapport aux autres, donne naissance à l'idée des réseaux sémantiques (*semantic networks*), introduite à la fin des années 1960. Par exemple, si on veut que la machine possède complètement la notion humaine de baleine, il faudrait aussi lui donner les notions de mammifère, d'océan, de communication sonore, de krill, de pêche, de protection de la nature etc., ainsi que toutes les notions liées à celles-ci. En bref, il faudrait mettre dans la machine toutes les connaissances qui nous viennent à l'esprit quand on prononce le mot « baleine » et, de fil en aiguille, la plus grande partie, si ce n'est la totalité, des connaissances humaines.

En 1963, Tom Evans, un étudiant de Minsky, écrit le programme ANALOGY, qui simule la détection d'analogies entre figures géométriques. La reconnaissance d'analogies est à la base de beaucoup de tests de QI (Quotient Intellectuel) et ANALOGY réussit parfaitement les tests de QI d'entrée au lycée. Mais ce programme est bien entendu totalement incapable de suivre les cours, ce qui remet en cause l'intérêt des tests de QI. ANALOGY est emblématique de nombreux programmes d'IA qui, tout en échouant à avancer vers la réalisation d'une machine intelligente, nous en apprennent pourtant long sur l'intelligence humaine.

En 1965, Joseph Weizenbaum (1923-2008) écrit le programme ELIZA, le premier agent conversationnel (*chat-bot*), destiné à simuler le comportement d'un psychothérapeute. Les utilisateurs interagissent avec ELIZA grâce au clavier de l'ordinateur. ELIZA se contente de faire parler l'utilisateur, en utilisant des phrases toutes faites du genre « Et vous, qu'en pensez-vous ? », ou bien en détectant des mots-clés. Par exemple, si l'utilisateur emploie le mot

« mère » dans une phrase, ELIZA va répondre « Dites m'en plus sur votre mère ». ELIZA n'est pas seulement l'ancêtre de tous les programmes auxquels nous sommes actuellement confrontés dans les services d'assistance par téléphone ou sur le Web. Il est aussi le premier exemple de programme à la Frankenstein, ayant poussé son créateur à arrêter ses travaux en intelligence artificielle et à militer pour l'arrêt des recherches dans ce domaine.



*Joseph Weizenbaum*

En effet, Weizenbaum va être choqué de voir certaines personnes tester ELIZA et, tout en sachant parfaitement qu'il s'agit d'un programme, l'utiliser comme un véritable psychothérapeute, en lui exposant leurs problèmes de couple par exemple, et, pire, en tirer un vrai réconfort. Il consacrera le reste de sa vie à réfléchir aux impacts de l'IA et de l'informatique en général sur les sociétés humaines.

À la fin des années 1950, ces premières réussites enthousiasment tellement la plupart des chercheurs que nombre d'entre eux sont persuadés que le projet de Dartmouth va aboutir en quelques années. Newell et Simon prédissent qu'avant 10 ans un ordinateur deviendra champion du monde d'échecs, composera de la musique et découvrira un nouveau théorème mathématique. Il a cependant fallu attendre 1997 pour qu'un champion du monde d'échecs, Gary Kasparov, perde face à une machine. Et les deux autres prédictions ne se sont jamais réalisées.

Cela n'empêche pas l'IA, comme communauté de recherche, de continuer à se développer et à se structurer. Née dans les universités de la côte est des États-Unis, cette discipline va d'abord essaimer en Californie. En 1962, en désaccord avec Minsky sur l'approche à adopter, MacCarthy part à

l'université de Stanford, près de San Francisco, créer son propre laboratoire. MacCarthy devient le chef de file de la mouvance des *neats* (soignés), qui pensent que tout travail sérieux en IA doit reposer sur la logique et les mathématiques, et qui restent fidèles au costume-cravate. Minsky, toujours au MIT, est à la tête des *scruffies* (débraillés), qui se soucient peu de preuve mathématique, n'hésitent pas à piocher des idées du côté des sciences humaines ou de la philosophie et s'autorisent les jeans et les chemises à fleur.

Le premier laboratoire européen d'IA voit le jour à Édimbourg en 1965. Et la première édition de la conférence IJCAI (*International Joint Conference on Artificial Intelligence*) a lieu en 1969 à Washington. Cette conférence, parrainée par l'association américaine d'intelligence artificielle (AAAI), ayant son siège en Californie, et plutôt proche de la mouvance *neats*, sera concurrencée dès 1974 par l'ECAI (*European Conference on Artificial Intelligence*), de tendance *scruffies*, soutenue par la suite par l'association européenne pour l'IA (ECCAI puis EurAI). Cependant, cette rivalité s'estompera au début du XXI<sup>e</sup> siècle. Cet essor de l'IA va pourtant connaître un arrêt brutal à la fin des années 1960, et une décennie de stagnation qu'on appellera le premier hiver de l'intelligence artificielle.

## IV. Le premier hiver de l'intelligence artificielle

Plusieurs causes sont à l'origine de ce coup d'arrêt. Tout d'abord, l'enthousiasme initial s'est rapidement heurté aux réalités scientifiques. L'approche connexioniste bute sur l'impossibilité de simuler des réseaux de neurones à plusieurs couches, ce qui limite énormément son intérêt pratique. L'approche symbolique, malgré ses succès dans l'automatisation des démonstrations mathématiques, n'arrive pas à dépasser le problème de l'explosion combinatoire. Le traitement automatique des langues, en particulier la traduction automatique, se heurte aux inextricables problèmes de polysémie (un mot a plusieurs sens selon le contexte) ou aux infinies variations des figures de style.

L'écart de plus en plus grand entre les prédictions euphoriques de certains chercheurs et les résultats très limités va conduire les bailleurs de fonds, au premier rang desquels figurent les agences militaires américaines, à limiter drastiquement les crédits des laboratoires. En 1973, aux USA, la DARPA (*Defense Advanced Research Projects Agency*) supprimera ainsi tous les financements concernant les projets d'intelligence artificielle, pour se concentrer sur des domaines moins fantomatiques. À la même date, le BSRC (*British Science Research Council*) annule tous ses crédits pour l'intelligence artificielle.

L'évolution de l'IA suit là encore celle de sa discipline jumelle, l'astronautique. Voyager dans l'espace et réaliser une machine pensante sont deux antiques fantasmes humains. Ils n'ont pu se concrétiser que dans les années 1950, une fois levés les obstacles psychologiques, théoriques et techniques à leur réalisation. Pour envisager la conquête spatiale, il fallait que soient réunies des conditions similaires à celles nécessaires à la naissance de l'IA :

- (1) considérer que l'espace et les planètes sont de même nature physique que la Terre ;
- (2) disposer d'un modèle des phénomènes physiques impliqués dans la navigation spatiale, en l'occurrence la mécanique newtonienne ;
- (3) mettre au point les machines permettant d'exploiter ces phénomènes pour nous emmener dans l'espace, les fusées.

À ses débuts, l'astronautique, comme l'IA, va rapidement produire des résultats époustouflants. En une dizaine d'années, les premières vont se succéder : premier satellite en 1957, première sonde interplanétaire en 1959, premier homme dans l'espace en 1961, premier pas sur la Lune en 1969. Mais, rapidement, l'exploration spatiale marque le pas. Et l'idée qu'il est possible de voyager aussi loin que l'on veut dans l'espace reste une simple hypothèse, tout comme le postulat de base de l'IA qui suppose qu'il est possible de simuler dans une machine tous les aspects de l'intelligence humaine. Bien que nous ne cessions d'en rêver, peut-être n'irons-nous jamais au-delà du système solaire, et peut-être n'aurons-nous jamais une machine intelligente.

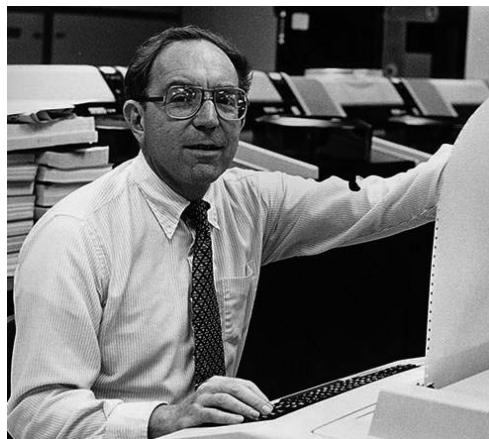
Alors que débute le premier hiver de l'intelligence artificielle, les gouvernements vont dès la fin des années 1960 renoncer à leurs ambitions spatiales initiales. En URSS, l'idée d'aller sur la Lune est abandonnée. Aux États-Unis, les missions Apollo 18, 19 et 20 sont annulées et les budgets spatiaux largement réduits. Paradoxalement, c'est aussi à cette époque que se banalisent dans la littérature et surtout au cinéma les voyages spatiaux et les machines intelligentes. Ces deux thématiques se rejoignent même souvent, comme dans le film *2001 l'Odyssée de l'espace*, où l'ordinateur contrôlant le vaisseau spatial se révolte contre l'équipage humain. Ce film illustre une particularité de l'IA par rapport à l'astronautique : les dangers que ce projet peut faire courir à l'humanité. Plus largement, les années 1970 verront l'émergence de nombreuses critiques éthiques et philosophiques de l'IA. Parmi d'autres, Joseph Weizenbaum, le créateur d'ELIZA, questionne la pertinence morale et sociale du projet : *quid* des impacts de la réalisation d'une machine pensante sur le comportement humain, sur l'emploi ou même sur la pérennité de l'humanité ?

Mais la faisabilité même du projet est aussi discutée. Un des principaux critiques, le philosophe Hubert Dreyfus (1929-2017), affirme que si les recherches n'aboutissent pas, c'est que l'idée selon laquelle on peut réduire

l'intelligence humaine à des manipulations de symboles est fausse. Établissant un nouveau parallèle avec l'astronautique, il compare les travaux en IA au fait de monter dans un arbre pour atteindre la Lune : en procédant ainsi, on peut avancer rapidement au début, mais on sera vite bloqué sans espoir de progresser. Dreyfus établit aussi un parallèle entre l'IA et l'alchimie, qui promettait des merveilles sans avoir jamais pu les produire. Pendant des siècles, alchimie et chimie étaient confondues comme disciplines ayant pour but d'étudier la matière et de produire de l'or ou des artefacts merveilleux comme la pierre philosophale. Bien que poursuivant des objectifs chimériques, les alchimistes ont contribué au développement des sciences de la matière. Puis, l'émergence de la démarche scientifique et rationnelle a conduit à la séparation entre les deux disciplines, et à la disparition de l'alchimie. De même, au milieu du XXe siècle, informatique et IA sont presque confondues. Ceux qui s'occupent de traitement automatique de données sont souvent intéressés par l'idée des machines pensantes, et les spécialistes d'IA sont experts en informatique. Les innovations des chercheurs en IA sont souvent recyclées en informatique « classique ». Ainsi, McCarthy crée en 1958 le langage LISP et introduit du même coup certains mécanismes devenus basiques en programmation. Les grammaires de Chomsky sont utilisées pour compiler les programmes, c'est-à-dire les traduire dans une forme exécutable par les machines. Inversement, les progrès de l'informatique aident les chercheurs en IA à progresser. Mais, dès les années 1970, la majorité des informaticiens ne se préoccupent plus d'intelligence artificielle, et ils ne sont aujourd'hui qu'une infime fraction à travailler à rendre les machines intelligentes.

On peut donc se demander si l'intelligence artificielle ne suivra pas le même chemin que l'alchimie et s'il ne s'agit pas d'un simple rêve, guère plus réaliste que la recherche d'un elixir de jouvence. Mais, contrairement à l'alchimie dont les ambitions sont depuis longtemps contredites par les connaissances scientifiques, rien ne prouve pour l'instant l'impossibilité de réaliser une machine intelligente, même s'il existe de nombreuses objections métaphysiques à l'IA. John von Neuman, l'inventeur de l'architecture des processeurs, se demandait par exemple s'il nous était possible de concevoir une machine aussi complexe que nous-mêmes. Ada Lovelace, la première personne à avoir écrit un programme destiné à tourner sur une machine, ce qui en fait la première informaticienne, faisait déjà remarquer au XIXe siècle qu'une machine ne fait qu'exécuter les instructions qu'elle a déjà en mémoire. Alors qu'aucun ordinateur n'existait à son époque, Ada Lovelace avait mis le doigt sur ce qu'on appelle maintenant le problème du cadre (*frame problem*) : une machine ne pourra faire face qu'aux situations prévues, directement ou non, par le programmeur, alors que l'intelligence humaine est en bonne partie affaire d'imagination et de capacité à faire face à l'inconnu et à l'imprévu.

Face aux problèmes théoriques et techniques, à la baisse des financements et aux critiques politiques ou philosophiques de l'intelligence artificielle, les chercheurs en IA des années 1970 vont revoir leurs ambitions à la baisse et privilégier des applications immédiates. Edward Feigenbaum, ancien étudiant d'Herbert Simon, émet dans les années 1960 l'idée que, pour faire face au problème de l'explosion combinatoire, il faut restreindre les systèmes de raisonnement symbolique à des domaines très limités des connaissances humaines, et à des types de raisonnement simples. Il nomme « systèmes experts » les logiciels qui visent, non pas à simuler toute l'intelligence humaine, mais uniquement à reproduire des capacités d'expertise technique, telles que le diagnostic médical ou l'identification de pannes.



*Edward Feigenbaum*

Le premier système expert, DENDRAL, créé en 1965 par Feigenbaum et Bruce Buchanan, est capable d'identifier des molécules chimiques à partir de spectroscopies. En 1972, Edward Shortliffe, un étudiant de Buchanan, développe MYCIN, un système expert dédié au diagnostic médical. MYCIN contient les connaissances des médecins sous forme de centaines de règles logiques du genre « SI le patient a un mal de crâne aigu ET que le mal de crâne s'est déclaré de façon brutale ET que sur une échelle de 0 à 4 le mal de crâne à une intensité supérieure à 3 ALORS la probabilité qu'il s'agisse d'une méningite bactérienne est supérieure à 60 % ET la probabilité qu'il s'agisse d'une méningite virale est supérieure à 40 % ET la probabilité d'une hémorragie méningée est supérieure à 60 % ». Au-delà du travail purement informatique, réaliser un système expert consiste à interroger les experts humains pour leur faire exprimer de façon précise et logique les connaissances qu'ils mettent en œuvre lorsqu'ils accomplissent la tâche qu'on veut faire réaliser à la machine. On retrouve ici le postulat de base de l'IA, posé en 1956

dans l'appel de Dartmouth : mettre l'intelligence humaine dans une machine suppose qu'on soit capable de décrire cette intelligence de façon suffisamment précise pour que la machine puisse la simuler.

Mais, même dans une tâche aussi technique que le diagnostic médical, l'expertise humaine se limite-t-elle à des règles logiques et conscientes telles que celle ci-dessus ? On peut en douter. Pourtant, en 1978, lors d'un test, MYCIN réalise de meilleurs diagnostics que les médecins humains. C'est la preuve que les techniques développées en intelligence artificielle peuvent se montrer efficaces et utiles. MYCIN ne sera cependant jamais utilisé par les médecins, qui considèrent que ce genre de logiciel pourrait les déposséder d'une part importante de leur travail, et que les problèmes juridiques et éthiques posés par leur utilisation sont supérieurs aux avantages qu'on pourrait en retirer en terme de rapidité et de précision de diagnostic. Si un médecin utilisant MYCIN fait une erreur de diagnostic, qui sera à blâmer : le médecin, ceux qui ont écrit le logiciel, ceux qui l'ont vendu, les autorités qui l'ont autorisé... ?

Ces problèmes éthiques n'empêcheront pas les systèmes experts de devenir un outil banal dans l'industrie, appliqué aujourd'hui à de nombreux domaines, du diagnostic de panne de locomotive à la prospection minière, de la conduite des usines sidérurgiques à la spéculation boursière. C'est encore un exemple des retombées pratiques de la recherche en IA, qui est une source d'innovation pour toute l'industrie informatique, de même que les inventions issues de l'astronautique irriguent toutes les industries qui contribuent au projet d'exploration spatiale.

La démarche de Feigenbaum est emblématique de cette époque de l'IA. Puisque la simulation de l'intelligence humaine dans toute sa complexité paraît utopique, il ne s'agit plus désormais que de simuler certains de ses aspects, dans un cadre restreint. Les systèmes experts sont ainsi un exemple de micro-monde, notion introduite par Terry Winograd (né en 1946), qui a développé en 1968 le programme SHRDLU, capable de gérer un univers virtuel constitué uniquement de blocs géométriques colorés. Les connaissances liées à un univers aussi simple sont très réduites, ce qui limite l'explosion combinatoire, et les énoncés concernant un tel monde sont dénués d'ambiguïté. Par exemple la phrase « poser la pyramide rouge sur le cube vert » n'offre pas beaucoup d'interprétations différentes.

Les années 1970 voient ainsi s'infléchir le projet initial de MacCarthy et Minsky. Ce qu'on appelle l'IA forte, visant à produire une machine simulant toute l'intelligence humaine, laisse place à l'IA faible, qui n'ambitionne plus que la réalisation d'imitations partielles des différentes manifestations de cette intelligence. Cette approche est confortée par certains travaux en sciences humaines, qui montrent que notre intelligence est à la fois diverse et compartimentée. Au début des années 1980, le psychologue Howard Gardner

développera l'idée des intelligences multiples : logico-mathématique, linguistique, musicale, sociale, spatiale, spirituelle etc. Ces capacités sont relativement indépendantes, de sorte que, suite à un traumatisme crânien par exemple, on peut perdre certaines capacités linguistiques (comme le fait de comprendre ce qu'on nous dit), sans pour autant perdre aucune capacité à compter ou à nouer des relations humaines.

Après l'euphorie irrationnelle des années 1960, les seventies sont l'occasion pour les chercheurs en IA de prendre réellement conscience de l'ampleur des difficultés du projet de l'appel de Dartmouth. L'intelligence humaine est bien plus complexe que prévue. En outre, elle entretient des liens étroits avec le monde physique. On sait que l'intelligence des enfants grandit grâce aux interactions avec le monde qui les entoure, et certains pensent même que la complexité de notre intelligence n'est que le reflet de la complexité du monde. Par exemple, le cheminement d'une fourmi dans la forêt peut paraître très complexe, mais il résulte souvent d'une stratégie très simple (par exemple suivre la piste des phéromones, et tourner à droite si un obstacle se présente). La complexité apparente des trajets des fourmis reflète le chaos du paysage et non la subtilité du comportement des fourmis.

Hubert Dreyfus a été l'un des premiers à pointer ce qu'on appelle en IA ce problème de l'enracinement (*grounding problem*) : une machine ne peut devenir intelligente que si elle interagit avec le monde physique et qu'elle apprend de ses expériences. La robotique et l'apprentissage apparaissent ainsi de plus en plus comme des clés pour l'IA. L'importance de l'apprentissage dans l'intelligence est aussi soulignée par des travaux en sciences cognitives. On sait, par exemple, que les capacités intellectuelles des chimpanzés adultes sont équivalentes à celle des enfants humains de deux ans. Or l'enfance chez les chimpanzés dure justement deux ans, alors qu'elle dure presque vingt ans chez les humains. Nos capacités intellectuelles singulières ne résulteraient-elles pas d'une enfance, et donc d'une période d'apprentissage, anormalement longue dans le règne animal ?

Durant l'hiver de l'IA, de nouvelles idées vont cependant voir le jour. Par exemple, Alain Colmerauer (1941-2017) développe en 1972 à Marseille la programmation logique, permettant d'écrire les programmes sous forme d'énoncés logiques plutôt que d'instructions à exécuter. Il crée le langage Prolog, qui deviendra un des principaux outils utilisés en IA, concurrençant le langage LISP créé en 1958 par MacCarthy. Les *scruffies* proposent également de nouveaux modèles censés aller plus loin que la logique d'Aristote. Par exemple, en 1974, Minsky invente le modèle des *Frames* : tout objet y est décrit par des propriétés intrinsèques et par les actions qu'il permet. Ainsi, une table est décrite par ses dimensions, son poids, sa couleur etc., et par la possibilité de la déplacer et de poser des objets dessus. Ce modèle est encore un exemple des apports de l'IA à l'informatique, car il est à l'origine du

paradigme de la programmation objet, qui est actuellement le paradigme dominant en développement logiciel.

En 1975, Roger Schank (né en 1946), constatant que l'interprétation des phrases n'est possible qu'en connaissant les situations dans lesquelles elles sont énoncées, propose de décrire toutes les activités humaines sous forme de scripts. Par exemple, aller au restaurant est une activité qui peut être décrite par des actions élémentaires comme entrer, demander une table, commander, manger, payer l'addition. Cependant, il paraît pratiquement impossible de décrire ainsi de façon précise toutes les situations que vivent les humains dans leur vie courante. Et quand bien même ce serait possible, le traitement de l'ensemble de ces descriptions dans une machine serait impraticable à cause du problème de l'explosion combinatoire. Comme dans le cas des systèmes experts, l'idée de Schank n'est utile que si elle est appliquée à des cas précis et très limités des connaissances humaines. Ainsi, en 1978, un programme utilisant cette technique est capable de comprendre et d'analyser des articles de journaux relatant des accidents de la circulation.

Après une traversée du désert d'une dizaine d'années, marquée par le manque de financements, les questionnements sur les différentes façons et même sur la possibilité de rendre une machine intelligente, et aussi par les désertions (par exemple, Winograd rejoindra Dreyfus dans sa critique de l'IA), la recherche en intelligence artificielle va connaître dans les années 1980 le début d'un nouveau cycle.

## V. Le deuxième cycle de l'IA

Le succès des systèmes experts redonne de l'intérêt à l'IA et, en 1982, le gouvernement japonais lance son projet d'ordinateur de 5e génération, doté de 500 millions de dollars, avec comme objectif final le développement d'une réelle intelligence artificielle. Les États-Unis, qui veulent rester en pointe dans ce domaine, répliquent en lançant en 1983 le programme *Strategic Computer Initiative*, doté d'un milliard de dollars. Cette rivalité va relancer pour dix ans les recherches en IA.

Au début des années 1980, plusieurs chercheurs proposent un nouvel algorithme (la rétropropagation du gradient) permettant enfin de simuler l'apprentissage dans les réseaux de neurones à plusieurs couches. Les limites du Perceptron sont ainsi surmontées, ce qui ouvre la voie à de nombreuses applications industrielles des réseaux de neurones, en particulier dans l'analyse d'image. Un réseau profond (à plusieurs couches) peut ainsi apprendre à reconnaître l'écriture manuscrite, à identifier des personnes sur des vidéos, à diagnostiquer des maladies à partir d'images médicales etc. Le

terme d'apprentissage profond (*deep learning*) est introduit en 1986 par Rina Dechter pour qualifier les nouveaux développements de l'approche connexioniste.

D'autres idées vont également voir le jour du côté de l'IA symbolique. Schank élabore au début des années 1980 le principe du raisonnement à partir de cas, qui consiste à copier notre faculté à tirer parti de nos expériences pour affronter de nouvelles situations. Pour résoudre un nouveau problème, on va chercher dans les cas déjà résolus celui qui s'en approche le plus, puis la solution déjà connue pour ce cas est adaptée au nouveau problème. Le nouveau problème et sa solution sont alors ajoutés à l'ensemble des cas déjà résolus, et ainsi de suite.

De nouvelles techniques d'apprentissage artificiel symbolique voient également le jour. En 1983, Douglas Lenat (né en 1950), un étudiant de Feigenbaum, veut donner aux programmes utilisant des heuristiques la possibilité d'améliorer ces dernières. Lenat écrit le programme EURISKO qui peut modifier automatiquement ses heuristiques de jeu, en utilisant des règles, heuristiques elles-aussi, d'un niveau supérieur. Il inscrit EURISKO à un tournoi de jeu de rôle où les joueurs doivent concevoir des vaisseaux spatiaux pour gagner des batailles. EURISKO gagne deux années de suite le tournoi en créant des vaisseaux très différents de ceux imaginés par les joueurs humains, puis les organisateurs obligent Lenat à ne plus participer au tournoi. En inventant automatiquement des formes de vaisseaux auxquelles aucun humain n'avait pensé, EURISKO est le premier exemple de « créativité artificielle ».

L'idée de Lenat est qu'il faut mettre dans les machines non seulement les connaissances humaines concernant le monde qui nous entoure, mais également les connaissances qui nous permettent de faire évoluer nos connaissances, ce qu'on appelle des métacatégories. Par exemple, nous avons tous appris à utiliser une bibliothèque, ou à interagir avec d'autres humains, pour apprendre de nouvelles choses ou pour remettre en cause des connaissances acquises.

Ces quelques avancées ne suffisent pourtant pas à justifier les sommes dépensées aux USA ou au Japon pour faire tourner les laboratoires d'IA. En 1987, les crédits du *Strategic Computer Initiative* dédiés à l'IA sont supprimés, et le projet japonais de 5e génération est stoppé en 1992 sans avoir permis de progresser vers la mise au point d'une machine pensante. Ces crédits publics représentent alors l'essentiel des ressources de la recherche en IA. Très peu d'entreprises privées s'intéressent à ce domaine, à commencer par les mastodontes, comme IBM qui s'est toujours refusé à investir dans ce projet flou, voire fumeux.

Un nouvel hiver de l'IA commence donc, alors que les pionniers des années 1950/1960 partent à la retraite. Les jeunes qui se lancent dans la recherche en IA n'auront pas connu l'effervescence de l'âge d'Or ni

l’indécroitable optimisme de ses acteurs. L’IA leur apparaît comme une idée séduisante, mais pas forcément concrétisable. Nombres de chercheurs ne veulent même plus revendiquer l’expression d’« intelligence artificielle ». Ils se considèrent comme des spécialistes en ingénierie des connaissances, en traitement automatique des langues, en vision par ordinateur ou en résolution de problèmes. Certains vont même proposer qu’IA ne signifierait plus qu’Informatique Avancée (*Advanced Informatics*), considérant que le but des recherches dans ce domaine ne consisterait plus qu’à résoudre des problèmes d’informatique (au sens de traitement automatique de données), cela par des méthodes non conventionnelles.

Leurs aînés s’étaient lancés tête baissée dans une entreprise fascinante, avec une foi absolue dans la possibilité de comprendre notre propre esprit et de le répliquer dans des machines. Mais notre esprit nous reste largement mystérieux. Et les ordinateurs, qui apparaissaient 40 ans plus tôt comme des machines fabuleuses pouvant potentiellement se substituer aux humains, voire à dieu (*cf.* la nouvelle d’Asimov *La dernière question*), ne sont plus, avec le développement de l’informatique grand public, que des appareils banals du quotidien. Pourtant, un incroyable regain d’intérêt pour l’IA est en cours depuis le début des années 2010.

## VI. L’IA au XXI<sup>e</sup> siècle

Depuis les années 2000, la massification des données informatiques, en particulier sur le Web, constitue une formidable opportunité pour l’IA. D’abord, elle introduit des problèmes nouveaux. Par exemple, pouvoir chercher efficacement des informations sur le Web requiert de mettre au point des moteurs de recherche « intelligents ». Les moteurs actuels, bien qu’utiles, sont extrêmement peu satisfaisants par rapport à ce qu’est capable de faire un libraire ou un bibliothécaire. Contrairement à un moteur de recherche, un libraire comprend des requêtes complexes, peut interagir avec nous, ne référence pas n’importe quoi, mais seulement les documents qui ont vraiment un intérêt, et ne nous conseille pas seulement les documents les plus consultés. Le projet de Web sémantique lancé par l’inventeur du Web, Tim Berners-Lee, en 1999, vise ainsi à ajouter au Web une couche de connaissances, exprimées par des formules logiques, pour rendre les moteurs de recherche et toutes les applications en ligne moins rudimentaires.

Mais les données massives sont surtout une des raisons des récents succès de l’approche connexioniste. L’apprentissage des réseaux de neurones profonds ne peut en effet donner de bons résultats que si beaucoup de données sont utilisées. De nouvelles idées sur la structuration des réseaux neuronaux,

et l'augmentation de la puissance des ordinateurs, vont en outre permettre d'utiliser des réseaux de grande taille et d'obtenir des résultats inattendus dans l'analyse d'image, la traduction automatique ou la prédition statistique de divers phénomènes comme la météorologie ou la criminalité. Ces avancées ont d'abord provoqué une nouveauté : les entreprises privées, au premier rang desquelles les GAFAM (Google, Amazon, Facebook, Apple, Microsoft), vont se mettre à investir massivement dans ces techniques. Cet intérêt de l'industrie informatique pour l'IA va déclencher un grand retour de l'IA dans les préoccupations des gouvernements et donc dans les laboratoires publics. Mais s'agit-il d'un vrai décollage de l'IA, ou plus vraisemblablement du début d'un troisième cycle de l'histoire de l'IA où l'effervescence laissera bientôt place à la frustration ?

L'exploration spatiale, pendant longtemps exclusivement réalisée par les États, est actuellement aussi de plus en plus menée par des entreprises privées. Encore une fois, l'histoire de l'IA est parallèle à celle de l'astronautique. Et, qu'il s'agisse de voyager dans l'espace ou de créer une machine pensante, l'aboutissement de ces deux projets est toujours suspendu à la même question : nous sera t-il possible un jour de comprendre suffisamment le monde pour réaliser tous nos rêves ? Pourtant, concrétiser ces rêves n'est peut-être pas le plus important. On considère souvent qu'explorer l'espace, même de façon très limitée, a comme principal intérêt de nous permettre de poser un autre regard sur notre planète. Les recherches en IA, qu'elles aboutissent ou pas, ont peut-être pour mérite essentiel de nous offrir de nouvelles perspectives sur notre intelligence et sur ce que nous sommes.

## Références

- Daniel Crevier. *À la recherche de l'intelligence artificielle*. Flammarion, 1999.
- Jean-Gabriel Ganascia. *L'âme machine, les enjeux de l'intelligence artificielle*. Seuil, 1990.
- Vernon Pratt. *Machines à penser, une histoire de l'intelligence artificielle*. Presses Universitaires de France, 1995.
- Franck Rose. *L'intelligence artificielle, histoire d'une recherche scientifique*. Payot, 1986.



# L'intelligence artificielle

*Dimension politique*



# Nouvelle gouvernementalité et fin du politique Libéralisme, contrôle social et privatisation du pouvoir

**Charles Thibout**

En première approximation, l'intelligence artificielle s'assimile à un répertoire de techniques visant à permettre à des machines d'accomplir des tâches traditionnellement dévolues à des animaux ou à des êtres humains. Rien n'indique donc que l'intelligence artificielle inclue *a priori* quelque dimension politique que ce soit. Néanmoins, force est de constater qu'elle produit des effets politiques : presque tous les États économiquement avancés ont élaboré des stratégies de développement de l'intelligence artificielle, suivant des considérations intérieures ou internationales, économiques, sécuritaires ou géopolitiques. À telle enseigne qu'un pays comme la Chine a placé l'intelligence artificielle au centre de sa stratégie de puissance et au cœur de son régime de contrôle sociopolitique, le même ensemble technologique devant *in fine* briser tout à la fois l'hégémonie américaine et les tentatives de rébellion à l'intérieur de ses frontières. Ces techniques occupent aujourd'hui des positions privilégiées dans des secteurs clés, du commerce à la justice, jusqu'à la santé et à la défense — en témoigne le contrat passé entre le département américain de la Défense et Google pour que l'entreprise fournisse certaines de ses technologies (en l'occurrence de « vision par ordinateur ») à l'armée et aux services de renseignement.

Il serait cependant réducteur d'examiner la dimension politique de l'intelligence artificielle à l'aune unique de ses effets. Ses conditions de production, théoriques et pratiques, sont déterminées par un ensemble idéologique et utopique qu'en première instance nous pouvons rattacher au libéralisme et à ses différents courants de pensée (social-libéralisme,

néolibéralisme et libertarianisme). Il est remarquable que l'un des auteurs les plus influents parmi les cadres dirigeants de la Silicon Valley soit Ayn Rand, théoricienne de l'objectivisme, un mouvement contigu au libertarianisme selon lequel les prérogatives de l'État doivent strictement se limiter au maintien de l'ordre et à la défense du territoire, toute autre activité étant soumise à la compétition des égoïsmes que seules les libertés individuelles (au premier rang desquelles le droit de propriété) peuvent légitimement circonscrire. L'inscription de l'intelligence artificielle dans le champ du libéralisme oblige, ce faisant, à interroger les effets de sa conception et de sa mise en œuvre sur les rapports entre l'État et la société. Si l'intelligence artificielle s'institue, de fait, en instrument de gouvernementalité (Michel Foucault), elle induit un changement de nature du politique.

C'est selon trois niveaux de lecture que nous examinerons donc les rapports de l'intelligence artificielle et du politique : idéologique d'abord, en ce que l'IA constitue, dans le sillage de la cybernétique, une tentative de surmonter les trois grandes blessures narcissiques infligées au libéralisme par les pensées marxienne, nietzschéenne et freudienne ; gouvernemental, ensuite, étant entendu qu'un certain nombre d'acteurs étatiques lui accordent une fonction sociorégulatrice de premier ordre ; régional, enfin, en ceci qu'elle attribue aux firmes numériques un rôle structurant dans l'action publique et impose, finalement, une réflexion sur leurs desseins politiques.

## I. La technique au secours du libéralisme

L'intelligence artificielle, fondée sur les principes techniques et philosophiques de la cybernétique, s'inscrit dans une entreprise de plus grande envergure qu'un simple élargissement du champ des possibilités humaines. Elle constitue avant tout une tentative de restauration de l'édifice conceptuel du libéralisme, mis à mal au tournant des XIXe et XXe siècles par trois puissants réquisitoires formulés par Marx, Freud et Nietzsche, dont la cible principale fut le postulat de l'autonomie individuelle et son corollaire, le libre arbitre. Sans viser l'exhaustivité, il nous faut restituer les arguments fondamentaux opposés par la cybernétique au marxisme, au nietzschéisme et à la psychanalyse.

## A. La communication contre l'aliénation

« *Ce n'est pas la conscience qui détermine la vie, mais la vie qui détermine la conscience.* » Cette phrase de Marx et Engels, extraite de *l'idéologie allemande*, initie la critique en profondeur du principe d'autonomie et de volonté libre, fondamental dans la pensée libérale. Celle-ci infère le droit de propriété de l'autonomie individuelle, qui implique *eo ipso* la propriété de soi, de sa conscience, de son corps et de ce que le corps produit ou transforme, notamment par le travail<sup>89</sup>. À cela Marx répond que le droit de propriété, qui fonde la propriété des moyens de production, débouche sur un procès de production de survaleur, c'est-à-dire l'excédent de valeur produit par le salarié pendant son temps de travail global, une fois qu'il a reproduit la valeur de sa force de travail (son salaire). Or l'existence même de la survaleur signifie l'aliénation de l'individu qui, en régime capitaliste, ne possède ni son outil de travail, ni le fruit de son travail, sa production : l'égalité des droits, et notamment du droit de propriété, est annihilée par le droit de propriété lui-même, en tant que le droit de propriété du capitaliste déborde celui du salarié, si bien que le travail salarié, sa production, appartiennent à un autre, et que le salarié lui-même est propriété d'un autre. Par extension, c'est le principe et la jouissance de l'autonomie individuelle qui sont refusés à une partie de l'humanité. Marx en déduit que « la société bourgeoise moderne [...] n'a pas aboli les antagonismes de classes. Elle n'a fait que substituer de nouvelles classes, de nouvelles conditions d'oppression, de nouvelles formes de luttes »<sup>90</sup>, le tout fondé sur une philosophie générale qui condamne *a priori* un tel état de fait.

Le mouvement cybernétique, qui a posé les fondements théoriques de l'intelligence artificielle au cours des années 1940 et 1950, entend surmonter la critique marxienne de la société libérale, conçue comme entité (re)productrice d'antagonismes, en déplaçant le problème du champ de la production vers celui de la communication, considérée comme l'agent intégrateur par excellence des sociétés humaines à la fois par les cybernéticiens et par certains philosophes libéraux ultérieurs, comme Jürgen Habermas. L'un des pères fondateurs de la cybernétique, le mathématicien américain Norbert Wiener, écrit ainsi en 1950 : « *Je ne veux pas dire que le sociologue ignore l'existence et la nature complexe des communications dans la société, mais, jusqu'à une date récente, il a eu tendance à oublier à quel*

---

<sup>89</sup> J. Locke, *Traité du gouvernement civil*, chapitre V « De la propriété des choses », 44.

<sup>90</sup> K. Marx et F. Engels, *Le manifeste communiste*, Œuvres, t. 1, « Bourgeois et prolétaires », Paris, Gallimard (Bibliothèque de la Pléiade), 1965, p. 162.

*point elles sont le ciment qui donne sa cohésion à l'édifice social.* »<sup>91</sup> Ce faisant, les maux dont souffre la société, les convulsions sociales et les tensions politiques qui la traversent sont réduits en dernière instance à une défaillance technique, au manque ou à la mauvaise qualité de la communication entre les élites dirigeantes et leurs administrés, les propriétaires de capitaux et leurs salariés, et entre les citoyens eux-mêmes. Autre fondateur de la cybernétique, le physicien et philosophe Heinz von Foerster assimile ainsi les problèmes sociaux à des « *dysfonctions* » et indique qu'« *au niveau individuel, ceci est péniblement ressenti en termes d'apathie, de méfiance, de violence, de discontinuité, d'impuissance, d'aliénation etc.* [...] *Il n'est pas difficile de voir que la cause essentielle de ces dysfonctionnements est l'absence pour l'individu de données adéquates lui permettant d'interagir avec la société* »<sup>92</sup>. Par conséquent, le processus d'aliénation ne réside non plus dans la production capitaliste, mais dans les carences communicationnelles de la société qui interdisent l'autorégulation optimale de ses différentes composantes (homéostasie). De nécessaire, selon la lecture marxienne, la lutte des classes revue par la cybernétique devient donc contingente, autrement dit dépassable.

## B. Contre Nietzsche et Freud : le triomphe de la raison et de la liberté

Les cybernéticiens et, à leur suite, les théoriciens promoteurs de l'intelligence artificielle s'appliquent aussi à réhabiliter la conscience individuelle, rationnelle et libre, attaquée par Nietzsche et Freud. Sans entrer dans le détail, Nietzsche critique la définition transcendante de la logique et réduit le « Moi rationnel » à un mythe : la raison, qui fonde la responsabilité individuelle, est une construction ; elle participe de croyances qui ressortissent, en dernier ressort, à l'expérience, à des impressions, c'est-à-dire au corps. Le choix rationnel, la capacité de parvenir à une connaissance exacte et approfondie du monde (la science), constituent *in fine* des illusions : « [...] nous nous figurons par conséquent qu'*intelligere* est quelque chose de conciliant, de juste, de bien, quelque chose d'essentiellement opposé aux instincts ; tandis que ce n'est en réalité qu'*un certain rapport des instincts*

---

<sup>91</sup> N. Wiener, *The Human Use of Human Beings. Cybernetics and Society* [1950], Londres, Free Association Books, 1989, p. 27.

<sup>92</sup> H. von Foerster, « Les responsabilités de la compétence », dans E. Andreewsky et R. Delorme (dir.), *Seconde cybernétique et complexité. Rencontres avec Heinz von Foerster*, Paris, L'Harmattan, 2006, p. 112.

*entre eux.* »<sup>93</sup> Pour sa part, Freud attaque le principe de responsabilité individuelle par la voie du libre arbitre, auquel il substitue un déterminisme psychique connu sous le concept d'inconscient. La responsabilité individuelle, chère au libéralisme, est rendue inane, puisqu'elle n'est plus sise qu'à un mythe : la volonté libre.

La cybernétique et l'intelligence artificielle entendent donc réhabiliter le Moi rationnel et libre déconstruit par Nietzsche et Freud. L'altération de la raison pure sous le coup des passions et la modulation de la vie psychique par des processus insaisissables et inaccessibles à la conscience trouvent leur résolution dans une conception renouvelée du comportement humain. En effet, la cybernétique s'applique à édifier une science générale du fonctionnement de l'esprit humain et à reproduire ce dernier dans une machine. Ce projet suppose que l'esprit humain ou animal serait, en dernière analyse, réductible à l'activité mécanique d'une machine. Surtout, affranchi des passions et de l'inconscient, cet esprit (dés)incarné dans un dispositif mécanique peut désormais accéder sans entrave à la Raison et à la pleine disponibilité de ses moyens d'action logiques. L'être de pure rationalité discrédiété par Freud et Nietzsche trouve alors sa voie de salut dans la technique : « Théoriquement, si nous pouvions construire une structure mécanique remplissant exactement toutes les fonctions de la physiologie humaine, nous obtiendrions une machine dont les capacités intellectuelles seraient identiques à celles des êtres humains. »<sup>94</sup>

Cette analogie fonctionnelle entre l'homme et la machine requiert néanmoins une définition particulière de la machine, conçue comme un automate homéostatique, c'est-à-dire une entité mécanique autorégulatrice — régulée par ses propres mécanismes de communication internes. De là, Wiener et les autres cybernéticiens appellent de leurs vœux l'édification d'une homéostasie sociale, fondée sur une politique et une morale de la bonne communication. Cette conception sera reprise par les principaux théoriciens du libéralisme, dans la seconde moitié du XXe siècle, au premier rang desquels Jürgen Habermas et sa théorie de l'agir communicationnel, et Friedrich Hayek et sa notion d'ordre spontané, où la communication demeure le principe central d'ordonnancement de la société, et le « signal-prix » son code élémentaire.

---

<sup>93</sup> F. Nietzsche, *Le gai savoir*, « Qu'est-ce que connaître ? », § 333.

<sup>94</sup> N. Wiener, *The Human Use of Human Beings: Cybernetics and Society* [1950], Londres, Free Association Books, 1989, p. 57.

## II. La gouvernementalité selon l'intelligence artificielle

L'œuvre philosophique de la cybernétique est importante à plusieurs égards. Premièrement, elle réhabilite le libre arbitre et la raison rationnelle qui, s'ils sont limités par les passions humaines, peuvent trouver leur parfaite réalisation dans et par les machines. Deuxièmement, elle institue la possibilité d'une concorde sociale, tout spécialement grâce à une meilleure communication entre les individus permise par la technique. À un niveau politique, l'intelligence artificielle va plus loin en posant le possible avènement *hic et nunc* d'une société harmonieuse, fondée sur des règles objectivées par la science. De ce point de vue, elle s'inscrit dans le grand mouvement de « politisation de la technologie »<sup>95</sup>, né au XVIIIe siècle. Néanmoins, l'intelligence artificielle fournit à la gouvernementalité contemporaine les moyens d'agir au plus près des individus, dans leur intimité même, suivant une manière de « technologie politique des corps » poussée à son paroxysme.

### A. Extension du domaine de l'administration technicienne

Le principe d'une organisation mécanique de la société s'est répandu au rythme du progrès technique. Si le temps de la cybernétique est révolu, ses théories et son esprit demeurent foncièrement ancrés dans l'architecture conceptuelle de l'intelligence artificielle et sa mise en œuvre. D'abord populaire aux États-Unis, où Stanley Hoffman décèle une propension singulière à la résolution par la technique des problèmes humains — tendance qu'il conceptualisa sous l'expression « pensée experte »<sup>96</sup> —, c'est désormais en Chine que l'intelligence artificielle accuse un développement sans précédent, sous la conduite d'un Parti-État particulièrement allergique aux soubresauts sociaux et décidé à combler son retard face à la puissance américaine.

L'intérêt politique premier de l'intelligence artificielle repose sur sa capacité, plus ou moins imaginaire<sup>97</sup>, à ériger en science le gouvernement des

---

<sup>95</sup> N. L. Kaplan, *Le Pain, le peuple et le roi. La bataille du libéralisme sous Louis XV*, Paris, Perrin, 1986.

<sup>96</sup> S. Hoffmann, *Gulliver empêtré. Essai sur la politique étrangère des États-Unis*, Paris, Seuil, 1971.

<sup>97</sup> Ch. Thibout, « Intelligence artificielle, une géopolitique des fantasmes », 2018-1, n° 5, *Religiosité technologique*, p. 105-115.

hommes, au même titre que la physique ou les mathématiques. Elle adhère aux développements du libéralisme de la seconde moitié du XXe siècle, qui tentent d'évincer l'idéologie (au sens libéral) du champ politique pour circonscrire l'action publique aux règles et protocoles technoscientifiques<sup>98</sup>. En postulant la capacité de l'intelligence artificielle à traiter des quantités phénoménales de données, avec une rapidité et un niveau de précision sans équivalent, le personnel politique des principales puissances se considère en mesure de gouverner scientifiquement les populations, dans la longue tradition de l'« administration technique »<sup>99</sup> née dans l'Europe du XVIIIe siècle. À la « raison classificatoire »<sup>100</sup> issue du Siècle des Lumières est par ailleurs venue se surajouter une raison prédictive produite par les propriétés statistiques et probabilistes des technologies d'intelligence artificielle : il ne s'agit plus seulement d'ordonner la connaissance des territoires et des populations afin d'en extraire des instruments de gouvernementalité à un instant donné, mais également de prévoir l'évolution future des comportements individuels et collectifs et d'en inférer logiquement les politiques publiques idoines.

De telles utilisations de l'intelligence artificielle sont déjà à l'œuvre. Aux États-Unis, depuis les années 1990, la police utilise le logiciel PredPol, fondé sur la technique du *machine learning* (apprentissage automatique). À partir des données traitées par le logiciel (procès-verbaux divers, registres d'arrestation, appels au 911 etc.), il s'agit de prédire où et quand certains types de crime sont le plus susceptibles d'être perpétrés et, partant, de mieux définir les zones d'affectation des patrouilles. Quel que soit l'objectif à atteindre, le *modus operandi* est identique : la police de Chicago utilise la *Strategic Subject List* (SSL), plus connue sous le nom de « *Chicago Heat List* », un outil fondé également sur la technique du *machine learning* qui s'applique à prédire quelles personnes sont enclines à être impliquées dans une fusillade, en tant qu'agresseurs ou victimes.

Si ce type d'utilisation de l'IA, loin d'être parfait<sup>101</sup>, n'en est qu'à ses balbutiements, il indique une voie : celle de la rationalisation de l'action publique par la technique, entendue comme le mode d'intervention optimum des autorités jusque et y compris au cœur des fonctions régaliennes du pouvoir. C'est l'accomplissement de ce que Stanley Hoffmann appelait la « pensée experte » ou de ce que le chercheur américano-biélorusse Evgeny

---

<sup>98</sup> D. Bell, *The End of Ideology*, New York, The Free Press, 1965.

<sup>99</sup> M. Antoine, *Le gouvernement et l'administration sous Louis XV. Dictionnaire biographique*, Paris, Éditions du Centre national de la Recherche scientifique, 1978.

<sup>100</sup> P. Tort, *La Raison classificatoire. Quinze études*, Paris, Aubier Montaigne, 1989.

<sup>101</sup> J. Asher et R. Arthur, « Inside the Algorithm That Tries to Predict Gun Violence in Chicago », *The New York Times*, 13 juin 2017.

Morozov qualifie de « solutionnisme technologique »<sup>102</sup> : la technique, en l'occurrence l'intelligence artificielle, est vue non seulement comme un instrument efficace de résolution des problèmes sociaux, mais elle instille également dans l'esprit des dirigeants l'idée suivant laquelle tous les problèmes afférents à l'activité humaine seraient réductibles, en dernier ressort, à un dysfonctionnement, un grippage dans la mécanique sociale conçue *in abstracto* sur le mode d'une harmonie préétablie qu'il conviendrait de faire advenir à nouveau. Là où l'intelligence humaine du personnel politique atteint ses limites, la machine viendrait ainsi suppléer à ses défaillances, accédant dès lors au vœu de Norbert Wiener : « Nous pouvons rêver à un temps où la *machine à gouverner*<sup>103</sup> viendrait remédier — pour le meilleur ou pour le pire — à l'insuffisance manifeste aujourd'hui du cerveau lorsqu'il est impliqué dans les mécanismes politiques habituels. »<sup>104</sup>

## B. De la technologie politique des corps au « Panoptique inversé »

Les appareils d'État tendent à exploiter les propriétés techniques de l'intelligence artificielle non seulement dans l'administration générale du territoire et des populations, mais ils s'attachent encore à infléchir, par l'algorithme, les comportements sociaux et individuels selon leurs agendas particuliers. Entre les mains des autorités, l'intelligence artificielle constitue ainsi tendanciellement un instrument de contrôle social, c'est-à-dire un instrument de conformation des comportements à certaines normes — ou de réaction à la déviance vis-à-vis de ces normes. Pour ce faire, deux configurations des algorithmes sont possibles<sup>105</sup>. Soit leurs concepteurs n'ont programmé aucune orientation particulière, mais les résultats d'analyse de l'IA seront utilisés dans un second temps pour nourrir des procédés plus classiques de manipulation (propagande, marketing politique, censure). Soit leurs concepteurs leur ont délibérément assigné certains objectifs, suivant un programme donné, ce qui a pour fonction d'aiguiller les comportements dans un sens particulier (par exemple, un moteur de recherche qui, à la requête « politique », classerait systématiquement et arbitrairement un homme politique en tête du classement).

---

<sup>102</sup> E. Morozov, *Pour tout résoudre, cliquez ici*, Paris, Fyp Éditions, 2014.

<sup>103</sup> En français dans le texte original.

<sup>104</sup> N. Wiener, *The Human Use of Human Beings...*, op. cit., p. 179.

<sup>105</sup> D. Cardon, « Le pouvoir des algorithmes », *Pouvoirs*, 164, 2018, p. 63-73.

Le système de manipulation des comportements le plus fameux, même si son développement n'est pas achevé, est sans doute le « Système de crédit social » chinois. Pensé dès le début des années 2000, ce système est surtout devenu une priorité après l'arrivée au pouvoir de Xi Jinping en 2012. Il vise à évaluer, en leur attribuant des notes, l'ensemble des citoyens et des entreprises chinois en fonction de leurs attitudes examinées au fil de l'analyse continue de leurs données. Disposant d'un capital de départ de mille points, citoyens et entreprises voient leurs notes personnelles croître ou décliner selon que leurs actions concordent ou non avec la morale commune édictée par le Parti-État, et selon que leur attitude fait d'eux des individus ou des entreprises en qui l'État peut avoir « confiance » (« 信用 » : crédit, fiabilité). De cette note dépend une position particulière dans un classement (par exemple, AAA (plus de 1050 points) identifie un « citoyen exemplaire » ; D (moins de 550 points) désigne un citoyen malhonnête<sup>106</sup>). À cette position correspondent un certain nombre de droits, de récompenses et de sanctions : facilités ou restrictions administratives et médicales, accès ou non à certains emplois publics, droit ou non d'acheter un billet d'avion, parmi d'autres exemples. Pour lors, il n'existe pas de système de crédit social unique couvrant l'ensemble du territoire, quoique l'État central se fixe cet objectif à l'horizon 2020. Il s'agit davantage d'une myriade d'initiatives locales (43 municipalités ont initié à ce jour un système de cette nature, associé systématiquement à un projet de *smart city*), dont le type d'évaluation et de régime de récompenses et de sanctions varie à la marge.

L'aspiration à une évaluation totale des comportements individuels n'est possible qu'avec l'assistance des techniques d'intelligence artificielle. Les autorités recueillent les informations nécessaires sous la forme de données, collectées grâce à différents vecteurs (réseaux sociaux, caméras de vidéosurveillance, objets connectés...), qu'analysent ensuite des techniques algorithmiques, assistées de techniques d'apprentissage (*machine learning*), qui peuvent ainsi permettre de reconnaître un visage capté par une caméra et transformer les données en informations (identité juridique et biométrique, comportement, réseau de sociabilité...). Ces informations permettent alors aux technologies d'intelligence artificielle de noter les individus identifiés, mais aussi de préjuger en termes probabilistes leurs comportements futurs. Il n'est pas anodin qu'une des premières versions de ce système ait été conçue par Alipay, un service de paiement dématérialisé appartenant à Alibaba, l'une des grandes firmes numériques nationales, pour analyser la capacité de ses clients à rembourser leurs créances. Sur fond de moralisation des citoyens et des entreprises, ce système vise à rendre la société chinoise « plus civilisée » et « plus harmonieuse » sous le haut et bienveillant patronage du pouvoir

---

<sup>106</sup> Ce classement correspond à la version du système de crédit social de la municipalité de Rongcheng.

chinois, qui se pose ainsi en unique source de concorde et de civilisation de la société chinoise<sup>107</sup>, grâce à un dispositif scientifique — et donc « objectif » — d'évaluation et de régulation morales des individus.

Pour reprendre la terminologie foucaldienne, ce système se rapporte à une « technologie politique des corps », une « microphysique du pouvoir » où la surveillance et le contrôle des corps préludent à la réforme des esprits<sup>108</sup>. Néanmoins, il n'est plus ici question de simples « sociétés disciplinaires », « panoptistes », en tant que la surveillance généralisée de la population passe par des dispositifs « fluides » et qu'elle s'horizontalise à mesure que s'étend l'usage des technologies sophistiquées de l'information et de la communication. Dans le contexte chinois, le Panoptique existe bel et bien — puisque demeure un pouvoir central de régulation des comportements sociaux —, mais il adopte des formes plus « insidieuses » et plus « souples »<sup>109</sup> de coercition, notamment sous la forme de programmes algorithmiques. Par ailleurs, les autorités tendent à déléguer à la société elle-même son pouvoir de contrôle, par l'interconnexion croissante des individus au moyen des nouvelles technologies : chacun devient ainsi non seulement la sentinelle attentive de ses pairs<sup>110</sup>, mais il constitue encore la brique élémentaire d'un système de surveillance généralisée et autorégulée, suivant le vœu formulé par les cybernéticiens et les penseurs libéraux au milieu du XXe siècle.

Ce « Panoptique inversé » repose néanmoins sur des dispositifs créés par les autorités (à travers les entreprises du numérique proches du pouvoir) et contrôlés par elles (régulation étatique du cyberspace chinois). L'« ordre spontané » rêvé par Hayek demeure finement réglé par les institutions centrales. Est-ce à dire, cependant, que l'intelligence artificielle n'est qu'un instrument supplémentaire dans l'arsenal des États pour encadrer les populations ? Ne contient-elle pas en elle-même le ferment d'une gestion disciplinaire, fût-elle diffuse, des individus ? Par leur généralisation, les technologies d'IA accoutumant la population à s'autoréguler, ne serait-ce que par les systèmes de notation réciproque mis en place par nombre de plateformes, comme Uber, Facebook, Airbnb et tant d'autres. Ce faisant, l'intelligence artificielle semble davantage participer de la mise en œuvre de

---

<sup>107</sup> E. Dubois de Prisque, « Le système de crédit social. Comment la Chine évalue, récompense et punit sa population », Institut Thomas More, note 36, juillet 2019.

<sup>108</sup> M. Foucault, *Surveiller et punir*, Paris, Gallimard, 1975.

<sup>109</sup> Constatant la disparition progressive des sociétés disciplinaires traditionnelles, Michel Foucault en vint à évoquer sous ces termes le nouveau régime de pouvoir en train de se mettre en place, qu'il qualifiera de biopouvoir dans ses cours du Collège de France (1978-1979).

<sup>110</sup> Voir notamment J.-G. Ganascia, *Voir et pouvoir : qui nous surveille ?*, Paris, Le Pommier, 2009.

« sociétés de contrôle », notion développée par Gilles Deleuze et reprise par Antonio Negri. Soit des sociétés « dans lesquelles les mécanismes de commandement deviennent toujours plus “démocratiques”, toujours plus immanents au champ social, diffusés à travers les cerveaux et les corps des citoyens »<sup>111</sup>. La nouveauté introduite par les nouvelles technologies de l’information et de la communication, alliée aux apports prédictifs et adaptatifs de « l’intelligence artificielle », tient en ceci qu’un individu est à la fois surveillant, surveillé et vecteur de surveillance, et ce *a priori* de façon consentante, puisqu’il accepte et plébiscite même ces instruments. Surveillant, car il accède à la vie privée et transparente de ses semblables et fait peser sur eux une pression sociale qui, autrement, n’eût été aussi forte. Inversement, il est aussi sujet à la surveillance des autres, selon le même procédé. Enfin, vecteur, car il se place de fait en position de mouchard vis-à-vis des autorités par son utilisation sociale des nouvelles technologies de surveillance.

L’intelligence artificielle permet ainsi d’exercer en théorie un contrôle plus fin, plus leste et plus individualisé qu’aucun autre répertoire de techniques. L’autre apport de cette technologie ressortit au pouvoir de fascination qu’elle exerce, y compris sur les personnels administratif et politique. Elle contribue ainsi à accroître l’influence sociale et politique d’acteurs privés, souvent étrangers (exception faite de la Chine et des États-Unis en grande partie), qui sont les producteurs techniques de ces dispositifs et en règlent l’usage, notamment par leurs conditions générales d’utilisation.

### III. Vers la privatisation du pouvoir ?

La vision trouble que nous avons de ce Panoptique inversé, propre aux sociétés de contrôle, procède également de ce que les firmes transnationales numériques, tout particulièrement d’origine américaine, se trouvent en situation d’auxiliaires des autorités politiques en matière de maîtrise du corps social.

---

<sup>111</sup> M. Hardt et A. Negri, *Empire*, Cambridge (Mass.), Harvard University Press, 2000, p. 23.

## A. Entre nécessité et dépendance, États et firmes numériques

Depuis la mise en place du *Patriot Act* au lendemain des attentats du 11 septembre 2001, le gouvernement américain peut exiger l'accès aux données stockées par des serveurs hébergés sur le territoire national. Renforcé par les amendements au *Foreign Intelligence Surveillance Act* de 2008 et par le *Cloud Act* de 2018, l'arsenal législatif américain offre aux autorités toute possibilité d'accéder aux masses de données détenues par les firmes numériques américaines, au premier rang desquelles les GAFAM — comme dans le cadre du programme PRISM, révélé en 2013 par Edward Snowden. Or ces entreprises concentrent les procédés techniques les plus sophistiqués et les financements les plus conséquents en matière d'intelligence artificielle. Par ailleurs, elles collaborent étroitement avec les agences de renseignement américaines, notamment la National Security Agency (NSA) et la Central Intelligence Agency (CIA) : elles se sont d'ailleurs livrées une âpre concurrence pour remporter le contrat JEDI (*Joint Enterprise Defense Infrastructure*), visant à fournir une infrastructure informatique en nuage (*cloud computing*) et les services d'intelligence artificielle y afférents à l'ensemble du département américain de la Défense. Finalement, Microsoft a remporté ce contrat, le plus important de l'histoire du Pentagone, d'une valeur de dix milliards de dollars sur 10 ans.

L'instrumentalisation des GAFAM par les autorités américaines n'a pas seulement servi la politique de surveillance généralisée de la population nationale. L'extraterritorialité du droit américain, à laquelle sont soumises ces entreprises en vertu de leurs conditions générales d'utilisation, les ont également mises en situation d'exercer cette fonction hors du territoire américain, y compris pour espionner des gouvernements étrangers<sup>112</sup>. D'une part, cette situation décrit le statut des entreprises dans la compétition interétatique : elles sont traditionnellement mises à contribution par l'État au service de sa stratégie de puissance, en raison de leur subordination juridique et de leurs propres intérêts privés (contrats publics, subventions, exonérations fiscales, droit des sociétés jouent un rôle central dans la réussite ou l'échec de ces entreprises). D'autre part, cette configuration met en lumière le lien de dépendance qui unit l'État à ces firmes, en tant qu'elles produisent les technologies dont l'État a besoin pour mieux contrôler sa population et projeter son influence à l'extérieur de ses frontières. En régime capitaliste et libéral, la politique de surveillance passe nécessairement par un état de sujexion technique à l'égard des propriétaires des moyens de surveillance. Or

---

<sup>112</sup> La NSA a mené des opérations d'interception des communications personnelles de personnalités politiques de haut rang, à l'instar du président de la République française et de la chancelière allemande.

ce nouveau pouvoir en voie de constitution est difficilement maîtrisable : dans un contexte d'austérité budgétaire, lieu commun des politiques économiques d'inspiration libérale, l'État n'est pas en mesure de prendre à sa charge les investissements nécessaires à la production d'innovations technologiques ; l'État ne constraint pas l'activité d'une entreprise, sauf en cas d'atteinte à certaines règles, comme la concurrence libre et non faussée et la sécurité nationale. Or il s'avère que les démocraties occidentales renâclent à encadrer le champ d'action de ces entreprises : le règlement général sur la protection des données (RGPD) est ainsi régulièrement enfreint sans que les autorités publiques n'affichent une volonté franche d'y remédier, d'autant que d'autres textes nationaux et internationaux (*Privacy Shield*)<sup>113</sup> tendent à en minimiser les effets.

## B. La capacité d'influence politique des producteurs d'IA

Cette situation est aisément compréhensible si l'on se réfère au besoin affiché par ces pays d'accéder aux technologies développées dans le secteur privé étranger, en premier lieu l'intelligence artificielle. La majorité des États européens, et la Commission européenne elle-même, ont établi des plans de développement de l'intelligence artificielle<sup>114</sup> qui, en l'absence d'entreprises européennes suffisamment compétitives, requièrent l'apport de technologies extérieures, principalement venues de Californie. Cet état de fait révèle, de surcroît, le pouvoir d'influence de ces entreprises au sein des États, tout

---

<sup>113</sup> Entré en vigueur en 2016, le Privacy Shield est « un mécanisme d'auto-certification pour les entreprises établies aux États-Unis qui a été reconnu par la Commission européenne comme offrant un niveau de protection adéquat aux données à caractère personnel transférées par une entité européenne vers des entreprises établies aux États-Unis », selon la définition de la Commission nationale de l'informatique et des libertés (CNIL). Or, au-delà des procédures peu rigoureuse en matière de certification, ce texte est largement invalidé dans les faits par la législation américaine et n'entrave pas la transmission des données recueillies par des entreprises certifiées aux agences de renseignement américaines. Voir notamment Marc Rees, « Données personnelles : le Parlement européen très insatisfait du Privacy Shield », Next Impact, 6 juillet 2018. Disponible sur : <https://www.nextimpact.com/news/106824-donnees-personnelles-parlement-europeen-tres-insatisfait-privacy-shield.htm>

<sup>114</sup> À titre d'exemple, la Commission européenne a annoncé, en avril 2018, l'augmentation de 70 % de ses investissements annuels dans l'IA, dans le cadre du programme de recherche et d'innovation Horizon 2020. L'objectif est de parvenir à investir 1,5 milliards d'euros entre 2018 et 2020. En mars 2018, la France a également annoncé un plan d'investissement en IA de même valeur, étalé sur quatre ans.

spécialement aux États-Unis<sup>115</sup>. Les premiers linéaments de la stratégie nationale américaine en intelligence artificielle sont ainsi ébauchés par une commission *ad hoc* (*National Security Commission for Artificial Intelligence*), créée par le Congrès en mai 2018 et dont les principaux membres sont les PDG d'Amazon Web Service et d'Oracle, les directeurs de Microsoft Research Labs et de Google Cloud AI. Surtout, la commission est présidée par Eric Schmidt, l'ancien PDG de Google et de sa maison-mère, Alphabet.

Loin d'être exceptionnelle, l'intégration des GAFAM au cœur de l'État participe du fonctionnement ordinaire des structures politiques américaines — en témoignent la place privilégiée laissée aux lobbies dans les institutions publiques et la tolérance accordée aux pratiques de *revolving doors* (pantoufle et rétro-pantoufle). En 2018, Alphabet, Amazon, Facebook, Apple et Microsoft ont dépensé plus de 64 millions de dollars en lobbying. Or, durant la session parlementaire 2017-2018, 364 des 450 lobbyistes de ces firmes (81 %) occupaient auparavant un poste officiel au sein de l'État (Maison-Blanche, Congrès, départements ou agences). S'il demeure difficile d'évaluer avec précision les effets de l'immixtion des firmes numériques dans l'appareil d'État, le cas de l'intelligence artificielle à usage militaire nous en donne une première indication<sup>116</sup>. En avril 2017, le département américain de la Défense a lancé le « *Project Maven* », un programme d'équipement des drones de l'armée en logiciels de traitement automatisé de flux vidéo, basés sur les technologies d'apprentissage automatique, afin d'améliorer leurs capacités de renseignement, de surveillance et de détection de cibles. Google faisait alors partie des entreprises retenues pour fournir ce type de technologies. Or, concomitamment, Eric Schmidt cumulait les fonctions de PDG d'Alphabet et de directeur du Defense Innovation Advisory Board, un comité chargé de conseiller le Pentagone en matière d'innovation organisationnelle et d'acquisition technologique, avec une nette priorité accordée à l'IA. Par ailleurs, le responsable de ce projet, le secrétaire adjoint à la Défense, Robert O. Work, a depuis été nommé aux côtés d'Eric Schmidt pour diriger la National Security Commission on Artificial Intelligence. Les grandes entreprises d'intelligence artificielle sont ainsi en position d'influer sur le cours de l'action publique.

---

<sup>115</sup> Ch. Thibout, « La compétition mondiale de l'intelligence artificielle », *Pouvoirs*, n°170, p. 131-142.

<sup>116</sup> Ch. Thibout, « Quid de la domination technologique et scientifique », dans B. Badie et D. Vidal, *Fin du leadership américain ? L'état du monde 2020*, Paris, La Découverte, p. 132-139.

## Conclusion : Les jalons d'une politique de l'IA

Certains contempteurs de l'intelligence artificielle voient en elle le ferment d'un profond désordre social et politique. Selon certaines estimations de chercheurs d'Oxford<sup>117</sup> et de la Banque mondiale<sup>118</sup>, l'automatisation et la robotisation du travail, c'est-à-dire le remplacement tendanciel de la main-d'œuvre humaine par des machines dans le processus productif, pourrait concerter 47 % des emplois aux États-Unis, 66 % en Inde, 77 % en Chine et jusqu'à 85 % en Éthiopie, cela à l'horizon 2040-2050. En moyenne, deux tiers des emplois dans les pays en développement seraient menacés par l'intelligence artificielle, étant entendu qu'une grande partie d'entre eux, voire la quasi-totalité, ne seraient pas remplacés. Ce faisant, l'on s'achemine vers des sociétés où la majorité de la population sera sans emploi. Par ailleurs, cette situation, si elle advient, créera un substantiel déséquilibre entre les facteurs de production au profit du capital, ce qui générera en retour une forte concentration du capital et du pouvoir économique entre les mains d'un petit nombre d'individus. Or, l'histoire nous l'enseigne, cette concentration induira automatiquement corruption et inégalités, qui déboucheront nécessairement sur des conflits autour du contrôle des capitaux — source du pouvoir —, ainsi que des tensions sociales puisqu'une grande partie de la population sera exclue du processus productif et verra son niveau de vie considérablement décroître à mesure qu'une petite portion de la population s'enrichira.

Par ailleurs, la priorité accordée à l'IA dans le développement économique des pays conduira sans doute à un déséquilibre structurel, dans un pays donné, entre l'industrie de l'IA et les autres secteurs économiques. Dans l'hypothèse d'une croissance portée par cette industrie de l'intelligence artificielle, celle-ci tendra à monopoliser les exportations du pays et à les accroître. Or, l'augmentation des exportations d'un pays entraînant *ceteris paribus* le renchérissement de sa devise, les exportations des autres secteurs industriels diminueront en conséquence, ce qui ne manquera pas de générer des conflits et une hyperspecialisation sectorielle accrue autour de l'IA. Autrement dit, cette concentration sectorielle et financière autour de l'intelligence artificielle risque de marginaliser les travailleurs humains et les autres secteurs économiques et, par suite, d'être à l'origine de convulsions sociales et

---

<sup>117</sup> C. Benedikt Frey et M. Osborne, *The Future of Employment*, Working Paper, Oxford Martin Programme on Technology and Employment, 17 septembre 2013. Disponible sur : <https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf>.

<sup>118</sup> World Bank Group, *The Changing Nature of Work*, World Development Report 2019,2019. Disponible sur : <http://documents.worldbank.org/curated/en/816281518818814423/pdf/2019-WDR-Report.pdf>.

économiques aux effets potentiellement dramatiques. En outre, il n'est pas inutile d'établir une analogie entre ce qu'il adviendra d'une grande masse d'hommes sous l'effet de l'expansion industrielle de l'IA et le sort subi par les chevaux lors de la Seconde révolution industrielle : les bêtes sont devenues des bouches inutiles, incapables de se former, de se réformer et finalement de s'adapter et de rester compétitives face aux conditions économiques nouvelles. Leur remplacement par des machines a ainsi conduit à leur quasi-disparition : quand les chevaux et les mules, surtout employés à des travaux agricoles, étaient 21 millions aux États-Unis en 1900, le pays n'en comptait plus que trois millions en 1960.

Le tour de force des principaux champions contemporains de l'intelligence artificielle est de confirmer ces hypothèses tout en les parant d'atours désirables. Ainsi Ray Kurzweil, directeur de l'ingénierie chez Google et cofondateur de la Singularity University, à la fois centre de formation, *think tank* et incubateur d'entreprises, estime, dans son célèbre livre *The Singularity is Near* (2005), que les machines devraient dépasser en intelligence les êtres humains d'ici 2045, ce qu'il appelle l'avènement de la « singularité technologique ». Le progrès technologique permettra non seulement d'« augmenter » les capacités physiques et intellectuelles des hommes, mais encore de transférer leur conscience à une machine et, partant, d'abolir la mort. En droite ligne avec le projet cybernétique, il s'agit de rompre tout lien entre l'esprit et cette matérialité pesante, passionnelle et limitative qu'est le corps : « Aujourd'hui, pour un humain, lorsque le matériel [son corps] fait défaut, le logiciel [son esprit] disparaît avec lui. Mais bientôt nous deviendrons du logiciel et le matériel sera remplaçable. »<sup>119</sup> L'arrivée prochaine de la singularité technologique nécessite que les êtres humains s'adaptent et emploient la technologie à se perfectionner afin de rester compétitifs face à ces nouveaux rivaux. Parallèlement, ces transhumanistes, parmi lesquels les principaux cadres dirigeants des firmes numériques de la Silicon Valley, affichent des convictions libertariennes, qui oscillent entre le minarchisme et l'anarcho-capitalisme. À la suite d'Ayn Rand, ils estiment que la libre concurrence des individus et des organisations économiques, qui requiert la mise en retrait radicale de l'État, permet une allocation optimale des ressources. Dans ce cadre, les ressources parviennent nécessairement en priorité aux individus les plus capables, créatifs et audacieux. Les autres, laissés-pour-compte, étant ceux qui n'ont pas voulu s'adapter à ce nouveau monde.

L'intelligence artificielle, associée aux théories libertariennes, dessine ainsi un programme politique élitiste, fondé sur la libre compétition des individus hors de toute régulation de l'État. Le projet cybernétique visant à

---

<sup>119</sup> R. Kurzweil, cité dans F. Benoit, *The Valley. Une histoire politique de la Silicon Valley*, Paris, Les Arènes, 2019, p. 238.

combler les trois grandes fractures narcissiques du libéralisme serait donc en voie d'achèvement : l'harmonie sociale, établie par les seules lois du marché et fondée sur la plus pure rationalité d'êtres humains réduits à l'état de consciences numériques, peut désormais advenir. Par voie de conséquence, cette utopie dessine tout autant une fin de l'Histoire qu'une fin du politique. Pour lors, néanmoins, l'État demeure un levier d'action incontournable pour y parvenir. L'un des grands promoteurs du libertarianisme californien, Peter Thiel, n'est-il pas celui qui, grâce aux logiciels d'IA développés par sa société, Palantir, dote l'État américain et quelques autres (dont la France avec la Direction générale de la sécurité intérieure) des moyens d'exercer une pression accrue sur la population en systématisant sa surveillance et son contrôle ? Dans une perspective cette fois-ci plus néolibérale, l'État est ainsi considéré à ce jour comme le plus puissant vecteur de diffusion des idéaux libertariens.

Dès lors, deux stratégies concurrentes se font face. D'un côté, ceux qui, à l'instar de Patri Friedman, ingénieur chez Google, petit-fils du monétariste Milton Friedman et fils du chantre de l'anarcho-capitalisme David Friedman, ambitionnent de créer des îles artificielles autogérées, sortes d'hétérotopies capitalistes et technicistes débarrassées des contraintes étatiques. De l'autre, ceux qui aspirent à pénétrer les États de l'intérieur pour les refonder sur un modèle entrepreneurial, voire — ce qui n'est pas nécessairement incompatible — d'en extraire les prérogatives régaliennes pour les transférer au secteur privé. Les ambitions présidentielles de Mark Zuckerberg ou le soutien d'entreprises comme Google aux contre-sociétés numériques laissent ouvertes toutes les perspectives.



# L'intelligence artificielle

*Dimension sociologique*



# **Une intelligence artificielle au travail**

## **Cinq histoires d'Homme**

**Yann Ferguson**

Au début du XIX<sup>e</sup> siècle, un état d’insurrection plane sur l’Angleterre. Des émeutiers du Midlands, armés d’« Enochs », de lourdes masses, entreprennent de briser des machines à tricoter des bas qui menacent leurs conditions d’existence : « Nous ne voyons pas la raison pour laquelle notre métier, qui est considéré comme l’activité essentielle de notre communauté et qui, pour sa plus grande partie, est notre seul moyen d’existence, devrait être exposé à tant de maux sans disposer d’aucun moyen de défense approprié »<sup>120</sup>. Leurs actions sont généralement précédées de l’envoi d’une lettre de menace, sommant le bonnetier de renoncer à ses machines sous peine de les voir détruites. Le signataire de cette lettre : Ludd, général Ludd, ou le roi Ludd, personnage légendaire, secret. Plusieurs siècles plus tard, l’ombre du machinisme sur l’ordre social s’étend à nouveau, si bien qu’« il n’est pas exclu que, bientôt, les chauffeurs de taxi et voitures de transport brisent les premiers véhicules sans chauffeurs », imaginent les informaticiens Serge Abiteboul et Gilles Dowek<sup>121</sup>.

La soudaine montée en maturité de techniques d’intelligence artificielle (IA) déclenche en effet un état d’urgence sociétal : « Le monde du travail est à l’aube de grandes transformations et n’y est encore que peu préparé », alerte la Mission Villani, chargée de définir la stratégie française. Il apparaît « de plus en plus certain qu’elle va modifier la majorité des métiers et des organisations. Ce développement va en effet permettre l’automatisation d’un grand nombre de tâches. Nous entrons donc dans une période de transition

---

<sup>120</sup> Chevassyus-au-Louis, 2006, chapitre 2, paragraphe 9.

<sup>121</sup> Abiteboul et Dowek, 2017, chapitre 8, paragraphe 21.

technologique importante, ce qui n'est pas sans créer des inquiétudes majeures : l'histoire nous enseigne que les précédentes transitions ne se sont pas faites sans encombre et que les processus de réajustements politiques ont parfois été violents, souvent au détriment des populations déjà les plus fragiles »<sup>122</sup>.

Ces inquiétudes doivent être rapportées à la visée originelle de l'IA, formulées dans la lettre de demande de subvention du *Dartmouth Summer Project*, séminaire fondateur de l'IA, en 1956 : « Notre travail a pour base l'hypothèse que chaque aspect de l'apprentissage ou de quelque autre caractéristique de l'intelligence peut en principe être décrit avec tant de détails qu'une machine pourra être construite pour la simuler. On tentera de découvrir comment fabriquer des machines qui utiliseront le langage, formeront des abstractions et des concepts, résoudront les problèmes aujourd'hui réservés aux êtres humains et sauront s'améliorer elles-mêmes »<sup>123</sup>. La recherche en IA a donc d'emblée pour ambition de développer des programmes sollicitant l'intelligence, une propriété sémantiquement floue mais symboliquement forte pour jusqu'à présent désigner la supériorité de l'Homme sur les non-humains. Si un tel programme aboutit, ira-t-on vers des suppressions massives d'emplois ? Quels seront les métiers et les catégories sociales les plus exposées ? Quelles seront les compétences clés pour évoluer dans ce nouvel environnement technologique ? En résumé, que fait l'IA de l'Homme au travail ?

S'est alors engagé un important effort prospectiviste et spéculatif de la part d'agences, de cabinets de consultants, de missions publiques, d'Académies, de chercheurs, d'essayistes etc. en vue d'identifier et de préparer ces mutations. Car « le risque de tétonie est important face à l'ampleur du phénomène », prévient la Mission Villani<sup>124</sup>. Ces travaux échafaudent incidemment une anthropologie du travail que cet article s'emploiera à ordonner. Un récit d'hominisation, fondé sur trois scénarios et cinq figures de l'Homme se dégage en effet. Le premier scénario considère la configuration où l'IA ferait mieux que l'humain : c'est l'« Homme remplacé ». Le deuxième aborde des situations de relations Homme-IA : ce sont l'« Homme dominé », « augmenté », puis l'« humanité divisée ». Le troisième qualifie comment l'IA rétablirait le propre de l'Homme. C'est l'« Homme réhumanisé ».

---

<sup>122</sup> Villani, 2018 : 102.

<sup>123</sup> Bostrom, 2017 : 17.

<sup>124</sup> Villani, 2018 : 102.

## I. L'homme remplacé : des routines aux algorithmes, de l'IA à l'emploi

En 2013, deux chercheurs d'Oxford, Carl Benedikt Frey et Michael Osborne, publient une étude quantitative retentissante dont la sentence inaugurale saisit le lecteur d'effroi : 47 % des emplois américains seraient menacés d'être automatisés d'ici dix à vingt ans. Pour obtenir cette estimation, les auteurs ont passé plus de 700 professions au filtre de trois catégories de compétences qui marquent traditionnellement les limites de la technologie : l'intelligence créative, l'intelligence sociale et la perception/manipulation. Plus une profession mobilise ces compétences, moins elle serait automatisable. « La frontière qui sépare les métiers automatisables des autres ne recoupe plus la distinction “manuel”/“intellectuel” comme c'était le cas jusqu'ici. Ce qui rend une tâche automatisable à l'heure du digital, c'est avant tout son caractère répétitif, qu'elle soit manuelle ou intellectuelle »<sup>125</sup>, analyse le cabinet Roland Berger qui, avec la même méthode, évalue à 42 % la part des emplois français potentiellement automatisables. L'IA engagerait un « point d'inflexion » vers ce qu'Erik Brynjolfsson et Andrew Mc Afee<sup>126</sup> appellent le « Second Âge de la Machine ». Les machines du premier âge se contentaient de produire un surcroît de puissance, sans toucher au monopole humain de la décision. Désormais, l'automatisation de tâches cognitives et de systèmes de contrôle devient possible au point que certaines machines se révèleraient capables de prendre de meilleures décisions que des humains. Dans cette configuration, les machines — c'est-à-dire les ordinateurs, les logiciels, les applications ou les robots — ne sont plus de simples compléments mais peuvent constituer des substituts à part entière.

Ces travaux aux résultats spectaculaires, dits « centrés sur l'emploi », sont largement repris dans les médias, mais aussi controversés. Ils négligeraient en effet la différence des activités réellement exercées derrière l'intitulé d'une profession. Or un même métier cache une grande diversité de tâches et de pratiques sollicitant des compétences variées. Il en résulte des travaux dits « centrés sur la tâche ». Selon David Autor, Levy et Murnane<sup>127</sup>, les tâches substituables sont les tâches routinières, aussi bien manuelles que cognitives, c'est-à-dire un nombre limité de tâches qui peuvent être définies avec les règles explicites d'un programme. *A contrario*, pour les tâches non routinières, plus complexes, le capital informatique se révèle plutôt complémentaire que substituable au travailleur. Avec cette approche, les travaux de Nicolas Le Ru pour l'agence gouvernementale France Stratégie considèrent que 3,4 millions

---

<sup>125</sup> Frey et Osborne, 2014 : 6.

<sup>126</sup> Brynjolfsson et Mc Afee, 2015.

<sup>127</sup> Autor, Levy et Murnane, 2003.

des emplois français sont fortement automatisables, contre 9,1 millions d'emplois *a priori* peu automatisables.

La plupart des études s'intéressent aux destructions potentielles d'emplois davantage qu'aux créations, toujours plus difficiles à estimer. L'étude d'Arntz, Gregory et Zierahn<sup>128</sup>, qui examine 18 États membres de l'OCDE, dont la France, sur la période 1990-2012, conclut que les investissements en technologies de l'information et de la communication n'ont pas d'effets négatifs sur l'emploi dans ces pays, au niveau agrégé, compte tenu des phénomènes de compensations. Cognizant Technology Solutions Corps<sup>129</sup> prévoit ainsi la création de 21 millions d'emplois aux États-Unis, pour des raisons simples : « Les machines peuvent faire plus, mais il y a toujours plus à faire. Une machine peut-elle (sous sa forme logicielle ou matérielle) se créer elle-même, se commercialiser, se vendre elle-même ? Se livrer ? Se nourrir ? Se nettoyer ? Se réparer ? Les machines sont des outils et les outils ont besoin d'être utilisés. Par les gens »<sup>130</sup>. De nouveaux métiers devraient donc apparaître « gestionnaires d'équipes homme-machine », « détectives de données », « sherpa de magasin virtuel », « agent de la diversité génétique », « conservateur de la mémoire individuelle », « responsable du sourcing éthique »<sup>131</sup> etc. S'intéressant aux États-Unis, l'économiste Robert Atkinson<sup>132</sup> montre que l'amélioration de la productivité s'est toujours accompagnée d'une augmentation de l'emploi : « quand une machine remplace un travailleur, un effet de second ordre se produit : l'organisation qui emploie cette machine économise de l'argent, et cet argent est injecté dans notre économie, soit parce que les prix baissent, soit parce que les salaires des employés restants augmentent, ou que l'entreprise réalise des profits plus importants. Dans tous les cas, cet argent est dépensé, ce qui stimule la demande des autres entreprises, qui, elles, embauchent donc plus d'employés ». Le chômage technologique ne résulterait donc pas des gains de productivité procurés par le progrès technique mais de son intégration tardive qui en ralentirait les bénéfices indirects et induits.

Ce rapide tour d'horizon dresse un tableau contrasté des effets quantitatifs de l'automatisation associés à l'IA. Un consensus se forme toutefois autour de la transformation inévitable du contenu des emplois, que le Conseil d'Orientation pour l'Emploi<sup>133</sup> évalue à 50 % pour la France. Ces transformations dessinent nos quatre autres « Hommes ».

---

<sup>128</sup> Arntz, Gregory et Zierahn, 2016.

<sup>129</sup> Cognizant Technology Solutions Corps, 2017.

<sup>130</sup> p. 3.

<sup>131</sup> p. 54.

<sup>132</sup> Atkinson, 2013.

<sup>133</sup> Conseil d'Orientation pour l'Emploi, 2017.

## II. L'homme dominé : singularité technologique, passivité humaine

La figure de l'« Homme dominé » est profondément ancrée dans les représentations collectives en raison d'un puissant imaginaire. Le rapport Villani rappelle en ce sens que « La (science) fiction, les fantasmes et les projections collectives ont accompagné l'essor de l'intelligence artificielle et guident parfois ses objectifs de long terme [...] », et que « c'est probablement cette alliance entre des projections fictionnelles et la recherche scientifique qui constitue l'essence de ce qu'on appelle l'IA »<sup>134</sup>. C'est d'ailleurs dans son roman *Un feu sur l'abîme*<sup>135</sup> que le scientifique Vernor Vinge a proposé le concept de « singularité technologique ». Il postule que l'avènement inéluctable d'une IA générale (ou forte) déclenchera un emballement de la croissance technologique porteur de changements imprévisibles sur la société humaine, qui perdrait alors le contrôle de son destin. Pour l'éviter, Ray Kurzweil<sup>136</sup>, fondateur de l'Université de la singularité, préconise la fin de l'humanité actuelle au profit du « transhumanisme », une nouvelle configuration où l'homme et les réseaux technologiques s'interpénètreront et se renforceront réciproquement.

Cette perspective est loin d'être partagée par l'ensemble de la communauté des experts : « rien dans l'état actuel des techniques d'intelligence artificielle n'autorise à affirmer que les ordinateurs seront bientôt en mesure de se perfectionner indéfiniment sans le concours des hommes, jusqu'à s'emballer, nous dépasser et acquérir leur autonomie », estime Jean-Gabriel Ganascia<sup>137</sup>. Mais son emprise sur les représentations collectives détourne l'opinion publique des effets des algorithmes actuels, développés dans le cadre d'une « IA faible » (ou étroite). Ces algorithmes se répandent dans de nombreux métiers, faisant émerger de nouvelles formes de collaboration homme-machine potentiellement aliénantes : « obéir aux ordres d'une intelligence artificielle, perdre le contrôle sur les processus, déléguer les décisions à la machine sont autant de modes de complémentarité qui, au niveau individuel et au niveau collectif, seront susceptibles de créer de la souffrance au travail »<sup>138</sup>. Cette aliénation se répandrait suivant une « logique rationalisante »<sup>139</sup> où le travail, plus ou moins piloté par des algorithmes, se passerait de l'intelligence des travailleurs.

---

<sup>134</sup> Villani, 2018, p. 9.

<sup>135</sup> Vinge, 1992.

<sup>136</sup> Kurzweil, 2007.

<sup>137</sup> Ganascia, 2017 : 55.

<sup>138</sup> Villani, 2018 : 186.

<sup>139</sup> Zacklad, 2017.

Constatant qu' « un degré extraordinaire d'ingéniosité humaine a été mis au service de l'élimination de l'ingéniosité humaine », l'activiste sociale Barbara Garson voit ainsi dans l'IA une résurgence de la rationalisation taylorienne : « transférer le savoir, les compétences et les capacités de décision des employés aux employeurs »<sup>140</sup>. Jean-Gabriel Ganascia met lui en garde contre la « démission des hommes » provoquée par une « prise de pouvoir passive des machines »<sup>141</sup>. Dès lors que le travailleur sera équipé d'un programme prétendument meilleur que lui, comment ne pas tomber dans l'écueil d'une délégation totale qui ressemble à une soumission à la décision automatique ? Qui prendra le risque de déconnecter la machine pour suivre son intuition ? Quel manager acceptera cette pratique quand les premières erreurs humaines, évitables par la machine, seront commises ? Les vertus facilitatrices de certaines machines intelligentes pourraient alors insidieusement déployer une forme de « paternalisme technologique »<sup>142</sup> aux multiples visages : alertes, recommandations, rappels à l'ordre, blocages, interdictions. C'est pourquoi Éric Sadin s'inquiète du « pouvoir injonctif » revêtu par l'IA, où « le libre arbitre de notre faculté de jugement et d'action se trouve substitué par des protocoles destinés à infléchir chacun de nos actes »<sup>143</sup>. Ces injonctions connaîtraient différents niveaux : « incitatif », « prescriptif », jusqu'à « coercitif ». Le groupe de travail sur l'acceptabilité sociétale de l'IA de France Stratégie identifie un risque de « prolétarisation des savoirs et des savoir-faire » : « À force de ne plus être entièrement partie prenante dans l'élaboration et l'expression d'une décision dans l'environnement de travail, les humains risquent de se voir dépossédés de leur expertise en terme de *know-how*, et de perdre un ensemble de capacités et de compétences, qui non seulement peuvent être utiles à la société, mais qui contribuent aussi à alimenter le respect de soi »<sup>144</sup>. Plutôt qu'une domination des machines par une hypothétique singularité technologique, Nicholas Carr dénonce une domination rampante des machines par l'abêtissement des humains. Contre le « mythe de la substitution » selon lequel à chaque fois que nous faisons appel à « un algorithme pour nous décharger dans notre travail, nous nous émancipons pour viser un objectif plus élevé et qui exige un degré supérieur d'ingéniosité et d'intelligence »<sup>145</sup>, il constate dans les métiers qui s'automatisent un appauvrissement des tâches cognitives, une déqualification, qui altèreraient notre façon d'agir et de penser.

---

<sup>140</sup> Garson, 1989 : 120.

<sup>141</sup> Ganascia, 2017 : 57.

<sup>142</sup> France Stratégie, 2017 : 8.

<sup>143</sup> Sadin, 2018 : 16.

<sup>144</sup> France Stratégie, 2017 : 6.

<sup>145</sup> Carr, 2017 : 71.

### III. L'homme augmenté : former des « centaures »

En 1997, Garry Kasparov subit une retentissante défaite contre Deep Blue, supercalculateur d'IBM. D'abord profondément meurtri, le champion se reprend en imaginant une nouvelle expérience de jeu : « Pourquoi ne pas jouer comme des partenaires au lieu de nous affronter ? » : « L'idée était de créer le plus haut niveau d'échecs jamais joué, une synthèse du meilleur de l'homme et de la machine »<sup>146</sup>. Il appela ces nouveaux joueurs des « Centaures », dotés simultanément de l'intuition, de l'empathie et de la créativité humaine et de la force de calcul brute d'un ordinateur. Pour le paléoanthropologue Pascal Picq<sup>147</sup>, l'avènement de l'IA constitue l'occasion d'enfin reconnaître et associer toutes les formes d'intelligence, contre le postulat cartésien selon lequel l'intelligence humaine serait l'étalon de toutes les intelligences, voire la seule. La fin de ce prétendu monopole serait la condition pour engager une nouvelle « co-évolution ».

Laurent Alexandre propose déjà d'élaborer un « quotient de complémentarité avec l'Intelligence Artificielle » qui « pourrait devenir l'indicateur phare de l'employabilité »<sup>148</sup>. La mission Villani en retient une version socialement plus acceptable, suggérant de « développer un indice de bonne complémentarité avec l'ensemble des parties prenantes (syndicats, État, monde de la recherche...), et en produisant des informations et de la documentation à l'intention des entreprises et des partenaires sociaux »<sup>149</sup>. Levy et Murnane<sup>150</sup> dessinent les contours de cet indice : à l'IA reviendrait l'exécution et donc les compétences techniques et les connaissances associées ; à l'Homme les compétences analytiques et de résolution de problèmes ainsi qu'un renforcement des aspects relationnels de l'emploi.

Dans les métiers du secteur juridique, par exemple, les applications capables d'effectuer des analyses juridiques ou de rédiger des actes juridiques basiques, « déplacent la valeur ajoutée des professions juridiques vers d'autres tâches telles que la plaidoirie, l'écoute du client, la constitution d'un réseau d'influence... »<sup>151</sup>. Le « collaborateur augmenté » verrait, lui, « son opérationnalité renforcée tout au long de sa carrière, et bénéficierait grâce à l'IA d'un appui à l'acquisition/au maintien de ses compétences »<sup>152</sup>. Pour le « DRH augmenté » la liste des opportunités serait longue : « Meilleure

---

<sup>146</sup> Kasparov, 2010.

<sup>147</sup> Picq, 2019.

<sup>148</sup> Alexandre, 2017 : 121.

<sup>149</sup> Villani, 2018 : 122.

<sup>150</sup> Levy et Murnane, 2013.

<sup>151</sup> Roland Berger, 2014 : 8.

<sup>152</sup> BCG-Malakoff Médéric, 2018 : 26.

compréhension des facteurs clés de succès dans un emploi donné et capacité à prédire la probabilité de succès à terme d'un candidat dans un emploi », « Optimisation du processus de recrutement grâce au tri préalable des CV », « Meilleure appréhension de la diversité grâce à une analyse plus fine des données, et suppression des biais cognitifs dans les processus d'entretiens grâce à une programmation adéquate de l'IA », « Meilleure rétention des talents clés grâce à des indicateurs prédictifs »,<sup>153</sup> etc. Le « manager augmenté »<sup>154</sup> pourra échanger dans de nombreuses langues grâce à la traduction automatisée, mieux gérer les talents grâce à des IA qui évaluent l'intérêt de chacun pour sa tâche, analyser et faire des reportings sur la posture des collaborateurs (analyse de mails, de comportements en réunion, recommandations pour s'améliorer dans sa vie professionnelle et privée), renforcer sa capacité d'analyse grâce au tri et à la mise en relation de données et d'idées au service de l'innovation et faciliter le travail en équipe grâce à des outils collaboratifs. Mais il devra pour cela apprendre à problématiser, savoir « interpréter les données et faciliter les flux », « développer un esprit critique », « éduquer et apprendre des IA »<sup>155</sup>.

L'implémentation de l'IA nécessiterait dès lors de repenser la formation professionnelle et le système éducatif. Ainsi, le futur « lab public de la transformation du travail », appelé par la mission Villani, devra expérimenter des « nouvelles modalités d'apprentissage et la manière de structurer une offre de formation professionnelle qui corresponde aux besoins difficiles à couvrir parce qu'ils ne relèvent pas directement de compétences métiers : créativité, transversalité, compétences cognitives générales... »<sup>156</sup>. Laurent Alexandre alloue des objectifs similaires à l'école. Selon lui, elle formerait actuellement les enfants à des métiers qui ne leur permettront pas d'être compétitifs face à l'IA. Or l'IA ne sait pas utiliser un savoir pour faire autre chose (apprentissage par transfert) et analyser transversalement un sujet. Laurent Alexandre préconise dès lors de déprofessionnaliser l'enseignement au profit des humanités, du développement de l'esprit critique, de l'apprentissage du travail en groupe, des savoir-faire transversaux et de toutes les approches multidisciplinaires. On retrouve ces inspirations dans la mission Villani : « Pour assurer la complémentarité de l'humain avec l'intelligence artificielle, ce sont les compétences cognitives transversales, mais également les compétences sociales et relationnelles et les compétences créatives qui doivent être développées »<sup>157</sup>. Contre une « une stricte conception "adéquationniste" » qui consiste à « construire des cursus éducatifs

---

<sup>153</sup> p. 48.

<sup>154</sup> Desjoux, 2018.

<sup>155</sup> p. 24.

<sup>156</sup> Villani, 2018 : 190.

<sup>157</sup> p. 114.

strictement indexés aux besoins d'un bassin d'emplois — et qui formeraient aujourd'hui des personnes dont l'emploi pourra être automatisé à peine quelques années plus tard », il faudra revaloriser « la compétence matricielle dans un monde en perpétuelle évolution : la créativité ». La mission préconise une profonde mutation des modes d'enseignement en faveur d'une capacité de « développement personnel » via « des pédagogies nouvelles qui soient orientées davantage vers le développement de l'esprit critique et de la coopération ». La promesse économique de l'*« Homme augmenté »* a donc un prix, celui de redéfinir les savoirs utiles et les modalités de leur acquisition. Il pourrait en résulter des opportunités professionnelles pour les bénéficiaires et des gains de compétitivité et de performance pour leurs employeurs. À condition, exprime la figure de l'*« humanité divisée »*, de faire partie de l'élite en capacité de se lier à l'IA.

## IV. L'humanité divisée : l'intelligence au cœur de la fracture sociale

Dans un de ses nombreux textes prophétiques, « Conséquences sociales du machinisme automatisé »<sup>158</sup>, Karl Marx pressentait comment les machines automatiques, « organes du cerveau humain créé par la main de l'homme », pourraient engendrer une situation sociale inédite où « les conditions du procès vital de la société sont soumises au contrôle de l'intelligence générale et portent sa marque »<sup>159</sup>. Au temps de l'IA, cette situation sociale prend la forme d'une polarisation du marché du travail entre, d'une part, une « aristocratie de l'intelligence » ayant un fort niveau de complémentarité avec l'IA et occupant des emplois fortement qualifiés et stimulants, et, d'autre part, des « inutiles »<sup>160</sup> aux emplois faiblement qualifiés, précaires et peu intéressants.

Pour saisir ce qui se joue socialement, Ian Goldin ose un parallèle avec la Renaissance : tout en portant une transformation radicale des sciences, des systèmes politiques et de l'humanisme, elle s'est accompagnée les siècles suivants d'une montée des extrémismes, de l'intolérance envers les scientifiques, les intellectuels, les étrangers et les minorités ethniques, ainsi que de guerres de religions. Or nous entrerions dans une « seconde Renaissance » : « elle amènera d'immenses bénéfices et de profonds désordres. Nous devons nous attendre à davantage d'extrémisme et de risques

---

<sup>158</sup> Marx, 1857-1858.

<sup>159</sup> p. 124.

<sup>160</sup> Harari, 2017.

potentiellement catastrophiques »<sup>161</sup>. Dans *La fin du travail*, Jérémie Rifkin affirmait déjà que « le logiciel remplace le personnel »<sup>162</sup>. Dans sa prévision, le secteur du savoir est le seul à émerger, et il n'absorberait pas les centaines de millions d'individus balayés par l'automatisation. La séparation mondiale entre une élite de manipulateurs d'abstraction et une masse croissante de travailleurs précarisés serait « le prix du progrès »<sup>163</sup>.

L'IA ferait dramatiquement monter ce prix. Selon Laurent Alexandre, le quotient intellectuel, qui serait déjà la principale source des inégalités sociales et économiques, deviendrait encore plus prépondérant à l'heure de l'IA en raison d'une courbe QI/revenus exponentielle. Car la complémentarité avec l'IA ne sera ouverte qu'aux cerveaux dotés de hauts QI, synthétisant notre plasticité neuronale, notre adaptabilité intellectuelle. Or cette capacité serait très inégalement distribuée socialement. Dès lors, l'intelligence pourrait cristalliser la lutte des classes, entre une élite désireuse de conserver sa domination et des inutiles qui, au nom de l'égalité, pourraient faire valoir « un droit opposable à l'intelligence » pour avoir une chance de continuer à exister socialement. En réponse, l'industrialisation de l'IA doit impérativement s'accompagner d'une « démocratisation de l'intelligence biologique »<sup>164</sup>, sous peine d'être socialement insoutenable.

Pour de nombreux experts, la polarisation est déjà engagée. Dans une étude pionnière, Nir Jaimovich et Henry E. Siu<sup>165</sup> notent que, durant les dernières récessions américaines, ce sont les emplois « solides » de la classe moyenne qui ont été les plus durablement touchés. À l'inverse, ceux qui ont tendance à être créés durant les reprises sont concentrés dans les secteurs d'emplois à bas salaires, comme le commerce de détail, l'hôtellerie et la restauration rapide, et, dans une moindre mesure, dans les professions qui nécessitent de longues études et de solides compétences. Cette polarisation entraînerait un marché du travail en forme de sablier où les travailleurs « ubérisés », incapables d'accéder aux emplois les plus attrayants, finissent au bas de l'échelle. Matthew Crawford y voit les conséquences structurelles de l'extension de la rationalisation taylorienne aux emplois de bureau qui, comme les ouvriers, subissent une dégradation de leur activité par la séparation du faire et du penser : « La part cognitive de ces tâches est “expropriée” par le management, systématisée sous formes de procédures abstraites, puis réinjectée dans le procès de travail pour être confiée à une nouvelle couche d'employés moins qualifiés que les professionnels qui les précédait. Loin d'être en pleine

---

<sup>161</sup> Goldin, 2017 : 327.

<sup>162</sup> Rifkin, 1997 : 24.

<sup>163</sup> p. 327.

<sup>164</sup> Alexandre, 2017 : 79.

<sup>165</sup> Jaimovich et Siu, 2012.

expansion, le véritable travail est en voie de concentration aux mains d'une élite de plus en plus restreinte »<sup>166</sup>.

Cette polarisation en cours appellerait la constitution de « solides filets de sécurité » en faveur des pauvres, des personnes et des familles fragilisées par la technologie, sous peine d'effriter la cohésion sociale et de voir monter les mouvements populistes. Un dispositif revient avec récurrence : le revenu minimum de base. Pour un de ses plus célèbres promoteurs, Rutger Bregman, « si nous voulons conserver les bienfaits de la technologie, il n'y a qu'une option en fin de compte, c'est la redistribution. Une redistribution massive. Redistribution d'argent (revenu minimum), de temps (réduction du temps de travail), de l'imposition (sur le capital plutôt que sur le travail) et, bien sûr, de robots »<sup>167</sup>. Mais, à l'image de sa réception lors des élections présidentielles françaises de 2017, le revenu minimum de base est très critiqué. D'abord, il semble se dresser contre ce « fait social total » qu'est le travail, consacré comme le meilleur canal de l'expression de soi en plus d'engendrer le lien social. Ensuite, analyse Harari, la révolution numérique aura lieu au niveau global, générant d'importantes inégalités géographiques. Des emplois disparaîtront au Honduras, alors que d'autres apparaîtront dans la Silicon Valley : « Et les Américains ne seront certainement pas enclins à payer des impôts pour les reverser au Honduras »<sup>168</sup>.

Les trois figures précédentes de l'Homme invitent donc hommes et société à se transformer pour éviter les scénarios dystopiques. Dans notre dernier profil, il s'agit moins de transformation que de redécouverte par la grâce des propriétés libératrices de l'IA.

## V. L'homme réhumanisé : la redécouverte du proprement humain

Une dernière histoire raconte comment l'IA redéfinit l'essence de la condition humaine. Le *think tank* The Future Society affirme que « l'IA incite les humains à réévaluer leur identité et leur place dans la société » et que « nous sommes en train de réinventer la condition humaine »<sup>169</sup>. Selon Philippe Breton, ce mécanisme se répéterait à chaque fois que l'Homme tente de se représenter dans ses créations et de se reproduire : « Là sans doute réside leur véritable signification, car, à travers elles, l'homme se contemple et tente

---

<sup>166</sup> Crawford, 2016, chapitre 2, paragraphe 14.

<sup>167</sup> Bregman, 2017 : 188.

<sup>168</sup> Harari, 2017 : 327.

<sup>169</sup> The Future Society, 2018 : 46.

de discerner les contours exacts de son humanité »<sup>170</sup>. Les créatures artificielles constituent en ce sens une « interrogation de l’homme sur lui-même »<sup>171</sup>.

À cette interrogation, l’informaticien Jean-Gabriel Ganascia répond que l’IA participe « d’un nouvel humanisme qui vise à mieux connaître l’homme, et à utiliser cette connaissance de l’homme pour mieux maîtriser son destin »<sup>172</sup>. Cette promesse de maîtrise est formulée dès la pièce de Karel Čapek, *R.U.R.*<sup>173</sup>, où apparaît pour la première fois le mot « robot » (issu du tchèque « *robota* », qui signifie « corvée »). Domin, le directeur de l’usine de production des robots, énonce ainsi les vertus émancipatrices de sa machine : « Désormais, Adam ne mangera plus son pain à la sueur de son front, il ne connaîtra ni la soif, ni la faim, ni la fatigue, ni l’humiliation, il reviendra au paradis où la main du Seigneur le nourrissait. Il sera libre et souverain. Son unique tâche, son unique travail et souci sera d’être le meilleur possible. Il sera enfin le maître de la création »<sup>174</sup>. Près d’un siècle plus tard, les mots de Domin se retrouvent dans le rapport Villani, qui décrit une complémentarité « libératrice » entre l’Homme et l’IA au travail : « l’automatisation des tâches et des métiers peut constituer une chance historique de désautomatisation du travail humain : elle permet de développer des capacités proprement humaines (créativité, dextérité manuelle, pensée abstraite, résolution de problèmes) »<sup>175</sup>. Libérer l’Homme de son automate marquerait donc ici un retour à « la qualité d’homme » dans la pensée rousseauiste où la liberté est constitutive de la nature humaine.

Le proprement humain se définit aussi par ce que l’IA ne sait pas faire, ces compétences devenant alors les symboles de l’humanité et consécutivement des indicateurs d’employabilité. Le Conseil d’Orientation pour l’Emploi retient ainsi cinq compétences : la flexibilité, la capacité d’adaptation, la capacité à résoudre des problèmes, les interactions sociales et la précision. Plus un emploi solliciterait ces facultés, plus il serait « humain », et moins il serait susceptible d’être automatisé. « Par exemple, si vous vous engagez dans la médecine, ne choisissez pas d’être le radiologue qui analyse les images médicales et sera remplacé par Watson, l’IA d’IBM, mais le docteur qui prescrit la radio, en discute les résultats avec le patient et décide du traitement », explique Max Tegmark<sup>176</sup>. Le Conseil d’Orientation pour l’Emploi alerte toutefois sur les « *a priori* inhérents aux critères retenus » qui

---

<sup>170</sup> Breton, 1995 : 68.

<sup>171</sup> p. 79.

<sup>172</sup> Ganascia, 2017, chapitre 4, paragraphe 4.

<sup>173</sup> Čapek, 1920.

<sup>174</sup> p. 41.

<sup>175</sup> Villani, 2018 : 105.

<sup>176</sup> Tegmark, Tegmark, 2018 : 146.

pourraient conduire à une sous-estimation du nombre d'emplois à risque d'automatisation : « Le fait qu'aujourd'hui telle ou telle tâche soit impossible à automatiser ne signifie pas que, demain, cela sera encore le cas »<sup>177</sup>. Le « proprement humain » serait alors à nouveau redéfini. Avec l'IA, l'Homme n'a décidément pas fini de se redécouvrir...

Cette redécouverte du proprement humain pourrait enfin le conduire à réaffirmer sa singularité. L'argument de la singularité humaine<sup>178</sup> postule que les principes évolutionnistes sont inadéquats pour expliquer la propriété centrale de l'espèce humaine : son ultra-sociabilité. Une propriété que l'Homme retrouverait grâce à l'IA, devenant encore plus Homme. Pour Bill Gates, l'avènement de l'IA est une opportunité pour « libérer les ouvriers, nous laisser améliorer la manière dont nous tendons la main aux personnes âgées, avoir des classes avec moins d'élèves, aider les enfants qui ont des besoins spécifiques. Toutes ces choses où l'empathie et la compréhension humaine sont encore tout à fait irremplaçables »<sup>179</sup>. Cette réhumanisation par réhabilitation de notre ultra-sociabilité est solidaire d'une revalorisation de la place des émotions dans l'intellect. Contre la doxa cartésienne, Damasio<sup>180</sup> soutient que les émotions font partie de la raison : le cerveau a pour caractéristique de permettre d'anticiper l'avenir et de former des plans d'action, ceci en s'appuyant sur l'orchestration fine de l'émotion. Le Conseil National de l'Ordre des Médecins acquiesce : « il est maintenant reconnu que c'est une erreur de chercher à éliminer les émotions du cursus médical. Il faut, au contraire, apprendre à les apprivoiser, à savoir les canaliser »<sup>181</sup>.

## Conclusion : Descartes et Taylor automatisés

Libéré de son automate, l'Homme au travail serait ainsi réhumanisé par l'anoblissement de ses tâches, la réhabilitation de ses instincts sociaux et de ses émotions. Cela suppose toutefois de repositionner deux penseurs structurants de la condition de l'Homme moderne, Descartes et Taylor, qui, loin de disparaître, passeraient de l'Homme à l'algorithme. Les figures de l'Homme remplacé, dominé, et de l'humanité divisée expriment comment ce transfert pourrait renforcer leur tutelle sur le travail. Simon Head<sup>182</sup> évoque en

---

<sup>177</sup> Conseil d'Orientation pour l'Emploi, 2017 : 69.

<sup>178</sup> Marchery, 2004.

<sup>179</sup> Cité par Hebert, 2017.

<sup>180</sup> Damasio, 1995.

<sup>181</sup> Conseil National de l'Ordre des Médecins, 2018 : 32.

<sup>182</sup> Head, 2014.

ce sens la montée d'un « management numérique » assimilant les êtres humains à des « représentations électroniques »<sup>183</sup> affichées sur les écrans des managers, l'ensemble accroissant l'emprise des processus au détriment de la pratique. Encoder Descartes et Taylor dans des machines marquerait la victoire définitive du mariage de l'intelligence logico-mathématique et de la rationalisation du travail et serait la source du drame social mondial à venir. Les figures de l'Homme augmenté et réhumanisé décrivent au contraire comment le travailleur s'émancipe de ce dogme, tout en profitant de sa version automatisée, par symbiose ou par délégation de tâches. Cette nouvelle configuration anthropotechnique conduirait à un nouvel humanisme, fondé sur le double déplacement de la raison aux émotions et de la répétition à la création. Dans ce cas, l'IA ne proposerait finalement rien d'autre à l'Homme qu'une remédiation pour sortir des carcans de la modernité.

## Références

- Abiteboul, S., Dowek, G. (2017), *Le temps des algorithmes* [version kindle], Paris : Le Pommier.
- Alexandre, L. (2017), *La guerre des intelligences. Intelligence Artificielle versus Intelligence Humaine*, Paris : JC Lattès.
- Arntz, M., Gregory, T., Zierahn, U. (2016), « The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis », OECD Social, Employment and Migration Working Papers, No. 189, OECD Publishing, Paris.
- Atkinson, R. (2013), « Stop Saying Robots Are Destroying Jobs — They Aren't », in MIT Technologie Review. <https://www.technologyreview.com/s/519016/stop-saying-robots-are-destroying-jobs-they-arent/>
- Autor, D., Levy, F. , Murnane, J.-M. (2003), « The skill content of recent technological change: an empirical exploration », in Quarterly Journal of Economic, 118(4), p. 1279-1333.
- BCG, Malakoff Mederic, (2018), « Intelligence Artificielle et capital humain. Quels défis pour les entreprises ? ».
- Benedikt, C., Osborne, M. (2013), « The Future of Employment : How Susceptible are Jobs to Computerisation ? » in Technological Forecasting and Social Change, vol. 114, issue C,, 254-280.

---

<sup>183</sup> p. 16.

- Berger Roland (2014), « Les classes moyennes face à la transformation digitale. Comment anticiper ? Comment accompagner ? ».
- Bostrom, N.(2017), *Superintelligence* [version kindle], Paris : Dunod.
- Bregman, R. (2017), *Les utopies réalistes* [version kindle], Paris : Seuil.
- Breton, P. (1995), *À l'image de l'Homme. Du Golem aux créatures virtuelles*, Paris : Seuil.
- Brynjolfsson, E., MC Afee, A. (2015), *Le Second Âge de la Machine*, Paris : Odile Jacob.
- Carr, N. (2017), *Remplacer l'humain. Critique de l'automatisation de la société*, Paris : L'Echappée.
- Chevassus-au-Louis, N. (2006), *Les briseurs de machines de Ned Ludd à José Bové* [version kindle], Paris : Seuil.
- Cnom (2018), « Médecins et patients dans le monde des datas, des algorithmes et l'Intelligence Artificielle »,
- COE (2017), « Automation, numérisation et emploi, tome 1 : Les impacts sur le volume, la structure et la localisation de l'emploi ».
- Cognizant (2017), « 21 jobs of the future. A guide to getting –and staying– employed over the next 10 years ».
- Crawford, M. (2016), *Éloge du carburateur* [version kindle], Paris : La Découverte.
- Damasio, A. (1995), *l'erreur de Descartes*, Paris : Odile Jacob.
- Dejoux, C. (2018), « Manager augmenté avec l'Intelligence Artificielle ? », in La Revue RH&M, n.68, 24-25.
- France Stratégie (2016), *Anticiper les impacts économiques et sociaux de l'Intelligence Artificielle. Annexe I : L'intelligence Artificielle en quête d'acceptabilité et de confort*.
- Ganascia, J.-G. (2017), *Le mythe de la Singularité. Faut-il craindre l'Intelligence artificielle* [version kindle], Paris : Seuil.
- Ganascia, J.-G (2017), *Intelligence Artificielle. Vers une domination programmées (Idées reçues)* [version kindle], Paris : Seuil.
- Garson (B.), *The Electronic Sweatshop : How Computers Are Transforming the Office of the Future into the Factory of the Past*, Simon & Schuster, N.Y., 1988.
- Goldin, I. (2017), « The second Renaissance », in Nature, vol. 550, 327-329.
- Harari, Y.N. (2017), *Homo Deus. Une brève histoire de l'avenir*, Paris : Albin Michel.
- Harari, Y.N. (2017), « Reboot for AI revolution », in Nature, vol. 550, 324-327.

Hebert, D. (2017), « Coup de bol pour Hamon : Bill Gates défend la taxe robot, Elon Musk le revenu universel », in L'Obs.

Heudin, J.-C. (2015), *R.U.R. Les Robots Universels de Rossum : D'après Karel Capek* [version kindle], Paris : Science eBookHead, S. (2014), Mindless : Why Smarter Machines are Making Dumber Humans, Basic Books : New York.

Levy, F., Murnane, R. (2013), « Dancing with robots Human skills for Computerized Work », NEXT.

Meda, D. (1998), Le travail, une valeur en voie de disparition, Paris : Champs Flammarion.

Jaimovich, N., Siu, H. (2012), « The Trend is the Cycle : Job Polarization and Jobless Recoveries », NBER Working Paper, 18334.

Kurzweil, R. (2007), *Humanité 2.0 : La Bible du changement*, Paris : M21 Editions.

Leru, N. (2016), « L'effet de l'automatisation sur l'emploi. Ce qu'on sait et ce qu'on ignore », in La note d'analyse, n. 49.

Machery, E. (2004), *Culture et singularité humaine. Les Matérialismes et leurs contradicteurs*, Paris : Editions Syllepse.

Mission Villani (2018), « Donner un sens à l'Intelligence Artificielle ».

Picq, P. (2019), *l'intelligence artificielle et les chimpanzés du futur* [version kindle], Paris : Odile Jacob.

Rifkin, J. (1997), *La fin du travail*, Paris : La Découverte.

Sadin, E. (2018), *l'intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical*, Paris : L'Echappée.

Tegmark, M. (2018), *La vie 3.0. Être humain à l'ère de l'Intelligence Artificielle*, Paris : Dunod.

The Future Society (2018), « A Global Civic Debate on Governing the rise of Artificial Intelligence ».

Zacklad, M. (2017), « Intelligence Artificielle : représentations et impacts sociétaux », CNAM.

# L'intelligence artificielle

*Dimension économique*



# Vers une quatrième révolution industrielle ? L'IA comme moteur, les données comme carburant

**Boris Barraud**

Une révolution industrielle est un processus historique lors duquel l'économie d'un pays, d'une région puis du monde change profondément de visage, entraînant par ricochet de lourds changements dans la société, la politique, le droit ou encore l'environnement<sup>184</sup>. Le terme « révolution » suppose une rupture brutale, mais il s'agit, en réalité, davantage d'une évolution progressive et continue. L'intelligence artificielle et les données pourraient néanmoins être les moyens d'une quatrième révolution industrielle relativement soudaine, profitant du déploiement du web et autres services de l'internet au début du XXIe siècle<sup>185</sup>. Fruit de progrès et d'innovations

---

<sup>184</sup> L'expression a été utilisée pour la première fois en 1837 par l'économiste français Adolphe Blanqui, dans son ouvrage *Histoire de l'économie politique*, puis elle a été reprise dans les années 1840 par Friedrich Engels.

<sup>185</sup> La première révolution industrielle date du milieu et de la fin du XVIIIe siècle, intervenant tout d'abord au Royaume-Uni. Après une lente phase de proto-industrialisation, elle correspond à l'apparition de la mécanisation et à la substitution de l'industrie à l'agriculture en tant que structure économique cadre de la société. L'exploitation massive du charbon associée à l'invention de la machine à vapeur ont donné l'impulsion à tout le processus, suivies par le développement des réseaux ferroviaires et la multiplication des échanges économiques, humains et matériels. La deuxième révolution industrielle s'est produite près d'un siècle plus tard, à la fin du XIXe, tout d'abord aux États-Unis et en Allemagne. La découverte de nouvelles sources d'énergie en est à l'origine : électricité, gaz et pétrole. La création du moteur à explosion a notamment permis d'exploiter le pétrole. Dans le même temps, la sidérurgie s'est développée à mesure de l'explosion de la demande d'acier. La chimie a également progressé fortement à ce moment là, avec la production de nouveaux

scientifiques trouvant de multiples applications économiques, cette révolution engendrerait bel et bien des conséquences à la fois sociales, politiques, juridiques et environnementales.

Il s'agirait de la première révolution industrielle à ne pas trouver son origine dans la découverte et l'exploitation d'une nouvelle source d'énergie matérielle. Le numérique et les données sont une énergie immatérielle<sup>186</sup>. La quatrième révolution industrielle, qui n'en est qu'à ses prémices, dont l'avènement futur demeure d'ailleurs en tout ou partie incertain<sup>187</sup> mais dont les développements actuels sont exponentiels — le marché de l'intelligence artificielle pourrait être multiplié par dix d'ici 2025 —, repose sur l'interconnexion permanente des ressources et des moyens de production, sur la communication entre les différents acteurs et objets des nouvelles lignes de production. « *Cloud* », « *Big Data* », « internet des objets », « intelligence

---

engrais. Les moyens de communication, enfin, ont été révolutionnés par les inventions du télégraphe et du téléphone, de même que les moyens de transport avec l'apparition de l'automobile puis de l'avion. Ces bouleversements macroéconomiques ont été permis par la centralisation de la recherche et des capitaux, structurés autour d'une économie et d'une industrie se basant sur les nouvelles « grandes usines », modèles d'organisations productives imaginés par Taylor et par Ford.

La troisième révolution industrielle est intervenue dans la seconde moitié du XXe siècle, aux États-Unis et au Japon en premier lieu, avec l'utilisation du nucléaire civil, nouvelle énergie au potentiel immense. L'avènement de l'électronique et celui de l'informatique sont au cœur de cette dernière révolution, spécialement avec les inventions du transistor et du microprocesseur. Il s'agit, dans une large mesure, d'une révolution des télécommunications et de l'informatique que l'internet, inventé dans la seconde moitié du XXe siècle mais réellement exploité avec le web à partir des années 1990, a très substantiellement enrichi. Dès lors, on produit des objets miniatures qui permettent l'entrée dans l'ère de l'automatisation de la production, avec le recours à des automates et à des robots.

<sup>186</sup> L'un des avantages de cette quatrième révolution industrielle serait ainsi d'être beaucoup moins énergivore que les précédentes car les données ne comptent pas au nombre des énergies non renouvelables — bien que certains services soient fortement consommateurs d'électricité. En outre, elle entre en résonance avec ce que l'on présentait, il y a quelques années, comme la perspective heureuse de la mondialisation : l'économie de la connaissance.

<sup>187</sup> Si l'on constate que d'innombrables entreprises investissent dans l'IA et s'équipent en IA, afin d'améliorer leur productivité et répondre à une nouvelle demande, il reste difficile de percevoir un réel signal macro-économique. Le même phénomène a déjà été observé, par exemple, avec l'électrification. Il faut du temps pour que des technologies et des ressources nouvelles entraînent une réorganisation du système. Il importe notamment, pour cela, que les entreprises adoptent toutes l'IA au même niveau.

artificielle » et, plus généralement, technologies « NBIC »<sup>188</sup> sont les maîtres-mots de l'économie 4.0 qui se dessine progressivement mais rapidement.

La quatrième révolution industrielle se caractériserait par une fusion des technologies qui brouille les lignes et qui perturbe les marchés. Elle promet de transformer des systèmes entiers de production, de gestion et de gouvernance. Le secteur industriel connaît déjà de nombreuses applications : maintenance prédictive, amélioration des prises de décision en temps réel, anticipation des stocks en fonction de l'avancement de la production, meilleure coordination entre les métiers etc. L'IA serait ainsi l'instrument de base de l'optimisation et du renouvellement des outils de production dans une économie toujours plus globale et interconnectée.

Cette quatrième révolution industrielle constitue — ou, du moins, constituera, une fois qu'elle sera mieux établie — un champ de recherche considérable pour les économistes, qui doivent et devront comprendre et expliquer les marchés et le système macroéconomique grâce à des indicateurs adaptés à leurs nouvelles structures. Pour l'heure, ce sont surtout des prospectivistes qui spéculent sur le devenir de l'économie mondiale et des économies nationales à mesure que l'IA et le big data se trouvent toujours plus en leurs cœurs. En ces lignes, on prendra plutôt le pari que l'intelligence artificielle et les données devraient effectivement bientôt impacter lourdement le monde économique et, par suite, les sociétés, les politiques, les droits, l'environnement etc. On répondra donc par l'affirmative à la question « Vers une quatrième révolution industrielle ? », cela en voyant dans l'IA son moteur et dans les données son carburant.

Une caractéristique importante des changements économiques actuels et à venir est que, pour la première fois de l'histoire, non seulement la France et l'Europe n'en sont pas à l'origine mais, en plus, elles sont fortement distancées, à tel point que d'aucuns considèrent que leur retard serait déjà irattrapable. Les États-Unis, avec leurs GAFAM<sup>189</sup>, et la Chine, avec ses BATX<sup>190</sup>, sont clairement les leaders de l'économie de l'IA et des données. Ce sont eux qui développent les algorithmes les plus performants et qui disposent des réservoirs de données les plus importants. Les talents ne manquent pourtant pas en France et en Europe, où les start-up fleurissent partout, mais les meilleurs chercheurs et les entreprises les plus innovantes sont trop souvent aspirés par les multinationales étrangères.

La généralisation des machines intelligentes pourrait donc induire une mutation économique aussi profonde que celle qui a vu l'industrie matérielle

---

<sup>188</sup> Les technologies NBIC sont les nanotechnologies, les biotechnologies, l'informatique et les sciences cognitives.

<sup>189</sup> Google, Amazon, Facebook, Apple, Microsoft.

<sup>190</sup> Baidu, Alibaba, Tencent, Xiaomi.

céder la place aux services. L'IA et les données constituerait le secteur quaternaire, un secteur dont le financement, la stimulation et l'encadrement deviendraient économiquement prioritaires. Après le passage de l'agriculture à l'industrie, puis celui de l'industrie aux services (banque, enseignement, administration, santé etc.), l'heure sonnerait d'une transition vers l'IA et les données. Toutefois, ces dernières impactent peut-être plus l'industrie et les services qu'elles ne les remplacent. Il ne s'agirait donc pas nécessairement d'un nouveau secteur à part entière — ce qui ne constitue pas un obstacle à ce qu'elles soient au cœur d'une quatrième révolution industrielle basée sur la création de chaînes de valeur globales autour, notamment, de l'internet des objets et de la robotique. « *AI is new electricity* », répète à l'envi Andrew Ng, le créateur de Google Brain. À mesure qu'elles traitent des quantités massives de données, les IA deviennent toujours plus efficaces et renforcent la dynamique d'innovation et de restructuration de l'économie prenant corps autour d'elles.

## I. L'explosion de la demande d'IA et du marché de l'IA

Ce sont de véritables guerres économiques autour de l'intelligence artificielle et des données qui s'annoncent<sup>191</sup>. Il y aura des vainqueurs et des vaincus, des gagnants de l'IA et des perdants de l'IA. Selon une étude datant de janvier 2019, 63 % des dirigeants d'entreprise estiment que l'IA aura un plus grand impact qu'internet sur l'économie et sur l'emploi<sup>192</sup>. En premier lieu, on s'attend à ce que le chiffre d'affaires global généré par l'intelligence artificielle dans le monde explose d'ici à 2025.

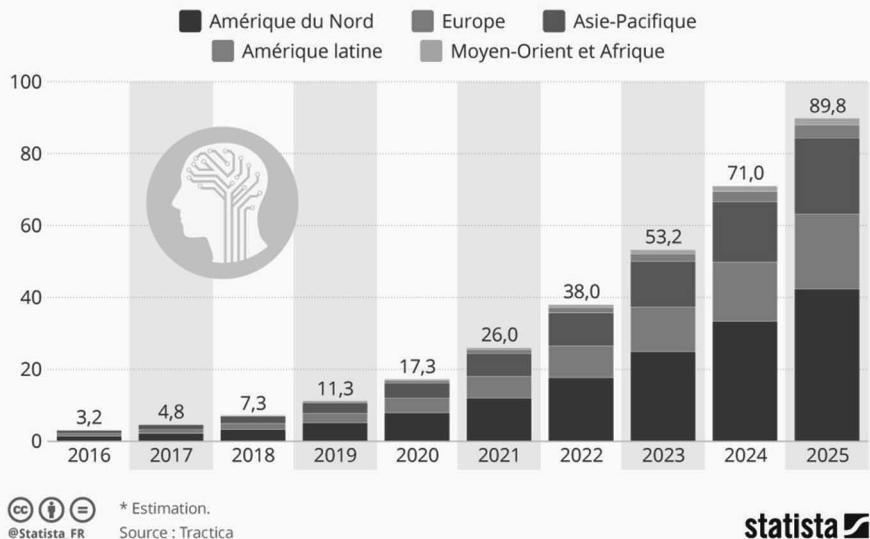
---

<sup>191</sup> B. Bérard, B. Pahud, C. Fayolle, *Guerres économiques pour l'intelligence artificielle*, VA Press, 2019.

<sup>192</sup> Enquête de PwC menée en janvier 2019 auprès de 1378 dirigeants d'entreprise dans divers pays du monde.

## L'IA, un marché qui vaut des milliards

Chiffre d'affaires généré par l'intelligence artificielle dans le monde, en milliards de dollars \*



L'IA devrait stimuler la demande grâce à des gains de productivité et à l'amélioration de la qualité, mais aussi en permettant de nouvelles propositions de valeur et de nouvelles expériences clients. En moyenne, parmi les entreprises du S&P500<sup>193</sup>, l'investissement en IA devrait permettre de dégager 6 milliards d'euros de chiffre d'affaires supplémentaires et un accroissement des bénéfices de l'ordre de 700 millions d'euros d'ici à 2030. Et cette technologie pourrait rapporter plus de 15 000 milliards de dollars à l'économie mondiale sur la même période<sup>194</sup>.

L'intelligence artificielle doit permettre aux consommateurs d'économiser beaucoup de temps. Les véhicules autonomes, l'internet des objets ou les outils de gestion financière, par exemple, serviraient à libérer du temps utilisable autrement, en consommant de nouveaux biens et services, en s'adonnant à de nouvelles activités. Cela devrait mécaniquement influencer la consommation. L'IA pourrait également doper la consommation grâce à une personnalisation accrue des produits. Or des biens et services plus personnalisés sont davantage attractifs. En outre, l'IA pourrait bientôt être capable d'anticiper et même prédire les demandes des consommateurs. Cela permettrait de réduire le temps de livraison, qui constitue aujourd'hui une

<sup>193</sup> Il s'agit des 500 sociétés les plus importantes au monde en termes de valorisation boursière.

<sup>194</sup> Enquête de PwC menée en janvier 2019 auprès de 1378 dirigeants d'entreprise dans divers pays du monde.

pierre d'achoppement pour le commerce à distance. De même, dans les secteurs de la santé, de l'éducation ou de l'assistance à distance, l'IA permettrait de décharger et donc de limiter les carences de systèmes dysfonctionnels, pénalisés par exemple par de longues files d'attente. Les chantres de l'IA mettent ainsi en avant l'amélioration de la qualité de service qu'elle favoriserait.

Le principal levier de croissance du PIB mondial pourrait donc résider autant dans une révision en profondeur des habitudes de consommation que dans des gains de productivité.

Concrètement, les secteurs dans lesquels l'IA pourrait générer le plus de chiffre d'affaires sont :

- Médias sociaux
- Reconnaissance d'images, classification et marquage
- Détection d'objets mobiles, identification et évitement
- Traitement de données médicales et analyse d'images médicales
- Algorithmes d'affaires
- Stratégie de trading algorithmique, amélioration des performances
- Localisation et cartographie
- Prévention contre les menaces de cybersécurité
- Conversion de documents papier en données numériques
- Systèmes de recrutement intelligent, ressources humaines
- Maintenance prédictive

Les cas d'usage, touchant principalement aux domaines de la vision, du *machine learning* et du *natural language processing*, témoignent de la diversité des recours possibles à l'IA, allant de la sécurité à la santé en passant par de nombreuses relations commerciales<sup>195</sup>. L'intelligence artificielle sert

---

<sup>195</sup> On peut citer, par exemple :

- Un vendeur de prêt-à-porter et d'ameublement a créé une application mobile permettant de proposer des produits similaires de son catalogue à partir d'une photo d'un objet acheté ou mis dans le panier.
- Un assureur a développé des agents conversationnels (chatbots) permettant l'interaction avec les clients et la fourniture d'informations commerciales plus pertinentes aux réseaux de vente.
- Un fabricant de logiciels a mis au point RoadPredict, une IA d'analyse prédictive visant à renforcer la sécurité routière et à réduire le temps d'intervention des secours sur les autoroutes.
- Un producteur de films a développé différentes solutions de détection vidéo qui permettent d'alerter les agents de sécurité des aéroports en cas de bagages oubliés, de

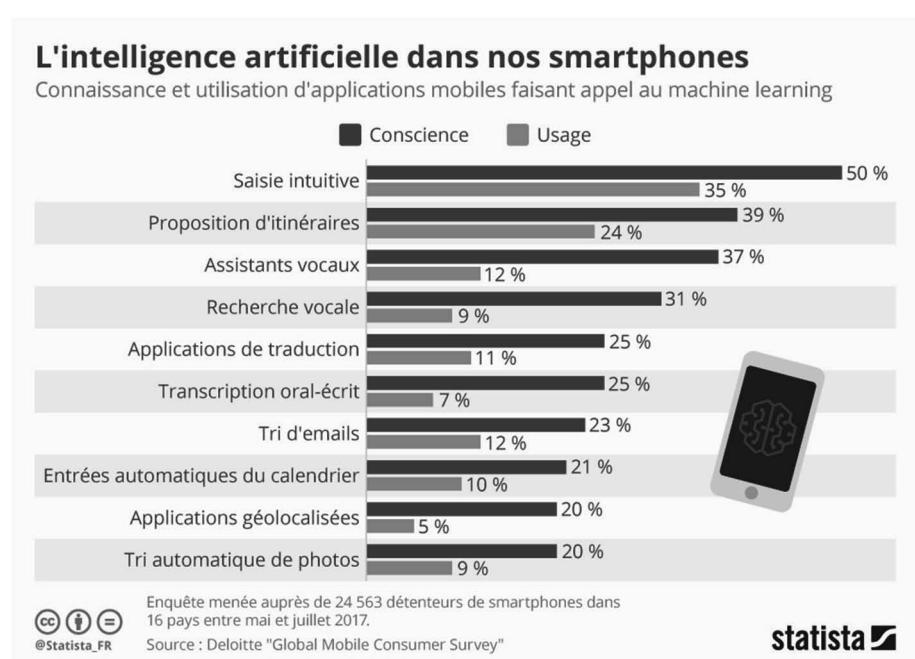
notamment le développement qualitatif et quantitatif de la robotique. Des robots thérapeutiques interviennent déjà dans certaines maisons de retraite. Des robots ménagers sont, quant à eux, présents depuis longtemps dans les foyers pour aider au quotidien. Des robots d'aide au déplacement permettent de pallier à certaines déficiences physiques. Des robots de téléprésence détectent des situations anormales ou rappellent l'heure de la prise de médicaments. Et des robots-compagnons, sécurisants et divertissants, arrivent sur le marché, tandis qu'on devrait bientôt voir apparaître des robots majordomes au service des plus fortunés. Il faut aussi ajouter tous les robots non humanoïdes, tels que les véhicules autonomes ou les drones de livraison — qui pourraient, par exemple, livrer les repas et les colis et donc remplacer les livreurs à vélo et les postiers. Tout cela dessine, semble-t-il, les contours d'une nouvelle économie.

Mais peut-être cette économie trouve-t-elle dans les ordinateurs de poche (ex-téléphones portables) son principal lieu de développement. C'est surtout là que l'intelligence artificielle infiltre le quotidien de chacun et donc les habitudes de consommation. Les smartphones, utilisés à tout moment pour tout faire partout dans le monde, comportent des applications à base d'apprentissage automatique. Or une enquête a pu mesurer à quel point les utilisateurs de ces ordinateurs de poche sont pour l'heure encore peu ou mal conscients de la présence de ces technologies et de leur influence sur leurs choix et comportements quotidiens. Mais la population devient progressivement plus avertie de la place de l'intelligence artificielle et de ses progrès, notamment en matière de fonctions quotidiennes telles que la saisie prédictive, les suggestions d'itinéraires, l'assistance vocale ou la traduction automatique.

---

déetecter les chutes de personnes, mais aussi les émotions faciales telles que la colère, la peur ou la joie.

- Un exploitant de gaz a créé une application mobile de reconnaissance d'image des produits chimiques utilisés par ses employés, afin de prévenir d'éventuels risques d'intoxication ou pollution.



## II. L'IA, nouveau facteur de production et nouveau vecteur de croissance

L'IA devrait bouleverser les structures de coûts des entreprises et donc leur compétitivité. Celles qui placeront suffisamment en amont ces nouvelles technologies au cœur de leurs stratégies de développement, spécialement en adaptant leurs métiers, devraient gagner en compétitivité et garder un avantage par rapport à de nouveaux entrants, ainsi que faire la différence par rapport à leurs concurrents non encore équipés et donc affaiblis par un retard technologique. Par conséquent, l'introduction massive de l'IA dans l'entreprise semble devenir un grand défi, voire une nécessité, une condition de survie. Il est certain qu'au moment de l'apparition de l'électricité, mieux valait ne pas choisir de rester à la vapeur et au charbon. Peut-être la situation est-elle peu ou prou la même actuellement.

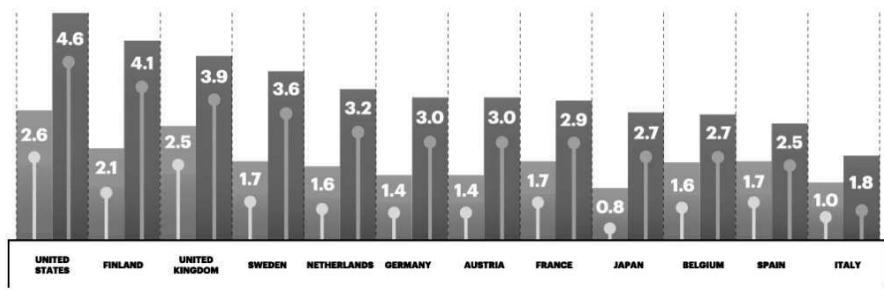
L'intelligence artificielle est une richesse pour une entreprise car elle peut apprendre en se nourrissant de l'expérience de cette entreprise, donc de son historique et des données qu'elle a produites. Cela doit permettre d'automatiser de nombreuses tâches, autant dans l'administration et la gestion que dans la chaîne de production et dans les mécanismes d'accès aux consommateurs. L'IA et les données ont donc vocation à offrir d'importants

## *L'intelligence artificielle*

gains de productivité immédiats. Et ceux-ci doivent porter tout spécialement sur la valeur créée grâce aux nouveaux services et à l'optimisation des services existants. Cela peut profiter aux marges, que ce soit avec des réductions de coûts très élevées ou avec le développement de services générant de nouvelles sources de revenus pratiquement à coût nul.

L'IA apprend des expériences et du passé de l'entreprise et cette intégration peut être réalisée dans toutes les fonctions et dans tous les métiers<sup>196</sup>. On parlera facilement d'une disruption technologique allant de pair avec une disruption économique, une disruption microéconomique et, par suite, une disruption macroéconomique. Au niveau microéconomique de l'entreprise, cette rupture est rendue possible par le déploiement de systèmes d'information dans la plupart des sociétés depuis les années 1980 et 1990. Cela a permis de produire et donc utiliser des données, du *big data*. L'introduction de l'IA s'inscrit dans la continuité tout en franchissant l'étape décisive de la recommandation et de l'action.

Les facteurs classiques de production, capital et travail, doivent donc être repensés, en y plaçant au centre le moteur de l'intelligence artificielle et le carburant des données, cela afin de pouvoir engendrer la croissance la plus élevée possible des économies nationales. Les perspectives, en la matière, semblent pour le moins intéressantes, avec une croissance susceptible de doubler d'ici à 2035, dans de nombreux pays dont la France, sous l'effet du développement des intelligences artificielles.



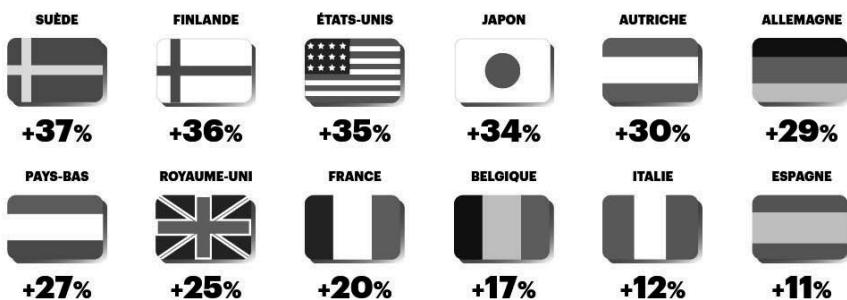
Taux de croissance annuels de valeur ajoutée brute en 2035 (proche du PIB) comparant la croissance de référence en 2035 à un scénario dans lequel l'intelligence artificielle a été intégrée à l'économie

Source : Accenture and Frontier Economics

<sup>196</sup> Par exemple, des cabinets comptables s'équipent d'IA permettant de comprendre et de classer automatiquement dans les bons plans comptables les factures de leurs clients ; et des assureurs automatisent l'allocation de certains fonds grâce à des IA reprenant l'historique des dix dernières années pour obtenir de meilleurs scores que leurs concurrents.

## Dans toutes ses dimensions

Une telle croissance reposant sur la révolution de l'IA supposerait de repenser la relation entre l'homme et la machine. L'impact de ces nouvelles technologies pourrait ainsi augmenter l'efficacité du travail de près de 40 % dans certains pays (20 % en France) en modifiant la manière d'effectuer de très nombreuses tâches et en permettant aux salariés de travailler plus efficacement et de se concentrer sur ce qui comporte la plus grande valeur humaine ajoutée.



Pourcentage d'augmentation de la productivité avec l'IA, en comparaison aux niveaux de productivité attendus en 2035  
Source : Accenture and Frontier Economics

L'intelligence artificielle conçue tel un nouveau facteur de production pourrait stimuler la croissance en jouant sur trois leviers :

1° En concevant des machines intelligentes capables de réaliser diverses tâches (beaucoup) mieux que les humains, y compris en simulant la capacité d'adaptation de l'homme ainsi que sa faculté à apprendre de ses expériences. Une telle automatisation requiert le support de salariés humains et une étroite collaboration entre salariés-humains et salariés-robots.

2° En permettant d'utiliser les facteurs de production classiques plus efficacement et notamment en amenant les employés à se concentrer sur les missions présentant la plus haute valeur ajoutée.

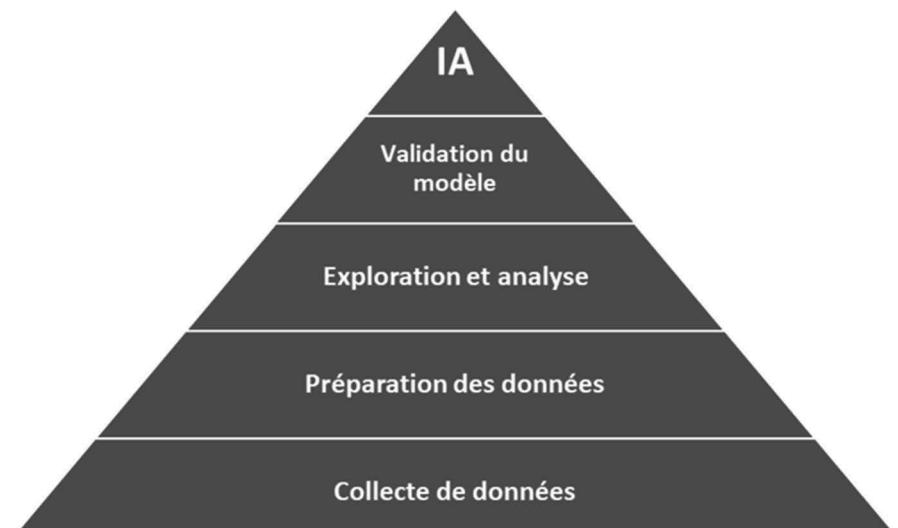
3° En diffusant les innovations à travers les sociétés. Par exemple, le développement des premières voitures autonomes a conduit de nombreux fabricants de voitures classiques à proposer leurs propres innovations et à établir des partenariats avec des spécialistes de l'intelligence artificielle.

Ces changements pourraient notamment provoquer l'éclatement de beaucoup d'entreprises. En effet, l'ensemble des process étant plus efficace, du fait de la spécialisation de certains acteurs recourant à des IA et apportant des solutions pour des besoins de plus en plus fins et précis, les entreprises devraient externaliser une partie croissante de leur travail et de leurs compétences, allégeant en conséquence leurs modes de fonctionnement. Il deviendrait donc de plus en plus nécessaire de développer de nouvelles

activités à coût modéré en allant chercher les compétences nécessaires à l'extérieur de l'entreprise<sup>197</sup>. L'IA et la quatrième révolution industrielle qu'elle accompagnerait auraient donc tendance à changer, parfois assez radicalement, les manières de faire et de penser des entreprises. D'aucuns évoquent ainsi une « plateformisation » de l'économie. Les petites entreprises, pour leur part, devraient se développer plus rapidement en s'appuyant sur des prestataires en IA capables de répondre efficacement à leurs besoins.

La création ou l'achat d'une IA n'est pas une fin en soi. Elle doit répondre à un objectif, généralement en rapport avec un *use case* métier. Ce cas d'usage, tout d'abord, indique quelles sont les données à collecter et à enregistrer pour nourrir l'IA et lui permettre d'apprendre et de progresser. Cet apprentissage est sans doute la phase la plus compliquée lors du déploiement d'une IA à des fins industrielles. Elle demande aux *data scientists* de tester plusieurs algorithmes puis d'ajuster les nombreux paramètres de celui choisi.

L'IA se retrouve ainsi au sommet de la pyramide des besoins de la science des données :



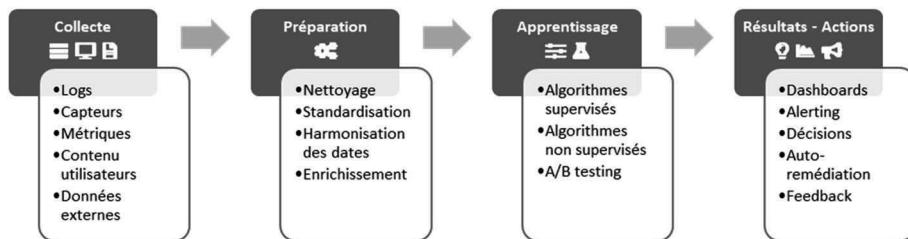
*Artik Consulting, juin 2018*

---

<sup>197</sup> Dans le secteur de la banque, par exemple, les nouveaux acteurs tels que les fintechs, qui parient sur l'IA pour proposer des solutions innovantes, concurrencent l'activité des acteurs traditionnels sur des segments importants. Les grandes banques sont ainsi amenées à construire tout un écosystème de partenaires afin de pouvoir proposer des services adaptés, personnalisés.

## Dans toutes ses dimensions

Les phases successives d'un projet d'IA peuvent être synthétisées de la manière suivante :



Artik Consulting, juin 2018

Le succès du recours à une IA dépend donc avant tout du sérieux et de l'efficacité de la collecte, de l'exploration et de la préparation des données. Sans ce socle solide qu'est la base de données, il n'est pas d'IA envisageable. La collecte des données, première phase, consiste à identifier les sources d'informations et à mettre en œuvre les moyens de se les procurer et de les stocker. La préparation, deuxième phase, consiste à nettoyer et organiser autant que possible les données collectées afin qu'elles soient exploitables malgré leurs sources possiblement hétérogènes. Il s'agit notamment de supprimer les entrées non pertinentes, de standardiser les formats, d'harmoniser les dates sur un même fuseau horaire, d'enrichir éventuellement les données avec les référentiels de l'entreprise etc. L'apprentissage, troisième phase, consiste en l'exploration manuelle et en l'apprentissage automatisé. Le *data scientist* peut explorer « manuellement » les jeux de données avant de définir l'algorithme le plus efficace pour les exploiter. Ensuite, l'apprentissage devient automatique. Le modèle issu de cet apprentissage doit alors être validé. C'est à ce moment que l'IA est opérationnelle. Débute la phase de résultats, l'IA étant alors censée répondre aux objectifs fixés au départ.

Toutes les entreprises doivent donc s'adapter aux ruptures technologiques en cours et à venir. L'investissement actuel dans l'IA portera peut-être ses fruits, à un niveau macroéconomique, dans cinq ou dix ans.

### III. Les effets de l'IA sur le marché du travail : destruction ou réorientation des emplois ?

Parmi d'autres problématiques<sup>198</sup>, la quatrième révolution industrielle façonnée autour des intelligences artificielles pose la question d'un grand renouvellement du marché du travail. Chaque révolution industrielle s'accompagne de profondes migrations des emplois : de l'agriculture vers l'industrie, de l'industrie vers les services, et aujourd'hui des services vers l'économie des données ? Les emplois dans les trois premiers secteurs seraient voués à être substantiellement remplacés par des robots et autres formes d'IA. On estime ainsi qu'il serait décisif d'envisager, dès aujourd'hui, le développement et le financement du secteur quaternaire constitué de métiers à haute valeur humaine ajoutée, dont le rôle serait notamment de s'occuper de l'humain (éducation, santé, culture, sport etc.). Si l'émancipation du secteur tertiaire et son déploiement au détriment du secteur secondaire ont été rendus possibles par l'enseignement obligatoire et la sécurité sociale, quelles mesures structurelles pourraient répondre au besoin de développement d'un tel secteur quaternaire ?

Les travailleurs devraient donc progressivement être libérés des tâches les plus simples et les plus mécaniques ou sans valeur humaine ajoutée. Cela signifierait moins de routine, de répétitivité et peut-être de pénibilité au travail et plus de missions stimulantes. Il importeraît surtout de favoriser la complémentarité entre intelligence humaine et intelligence artificielle, entre homme et machine. En rendant de nombreux outils et applications plus simples et intuitifs, l'IA devrait permettre à un plus grand nombre de personnes de les utiliser. Ceux qui travailleront aux côtés de l'IA profiteront d'un soutien décisif et seront beaucoup plus productifs. Ils acquiereront de nouvelles compétences et des aptitudes supplémentaires. Certaines théories font même de l'IA le moyen de passer à l'âge des « employés augmentés » grâce à de nouvelles formes d'assistance électronique à la décision ou à l'activité. Les hommes sont beaucoup moins efficaces que les IA pour faire des prédictions pertinentes, en raison de leurs mémoires limitées et de leurs capacités de calcul plus faibles, ainsi que du fait de biais psychologiques. Dans

---

<sup>198</sup> Par exemple, il serait aujourd'hui temps de s'interroger quant aux diverses conséquences d'une économie dans laquelle l'essentiel des besoins matériels de l'humanité serait satisfait par des robots intelligents. Dès lors, comment sera formé le prix des productions des robots si ceux-ci sont matériellement fabriqués par d'autres robots ? *Quid* alors de la théorie de la valeur travail ? La production de robots devrait être hautement capitaliste. Or, si le capital demeure très concentré, ses revenus seront partagés entre un petit nombre de ménages incapables d'absorber la totalité de la production de masse de robots par les robots. À moins que cette production de robots soit achetée par d'autres robots ?

ces conditions, les situations ne manquent pas dans lesquelles le travail de l'homme et le travail de l'IA pourraient être complémentaires.

Divers nouveaux métiers devraient être créés. Il faudra des ingénieurs et informaticiens pour concevoir les IA, des gestionnaires et autres spécialistes des données (*data scientists*), des entraîneurs et superviseurs d'algorithmes, mais aussi des techniciens capables d'entretenir et réparer tous ces outils informatiques et technologiques.

Le nombre d'emplois préservés sera fonction de l'élasticité de la demande suite à l'adoption de l'IA par les entreprises. Ainsi, les gains de productivité permettant de réduire les coûts de production et donc les prix de vente, la demande pourrait augmenter, obligeant à produire plus et donc à créer de nouveaux emplois. C'est pourquoi chaque pays se bat — ou le devrait — afin que son économie soit en avance sur ces sujets stratégiques et reste compétitive face à ses concurrentes. Plus généralement, l'économie des données et la révolution des IA permettraient d'améliorer la qualité du travail, l'innovation, la qualité de service et, par suite, la satisfaction globale, les ventes et les profits. Ces gains de compétitivité et ces nouveaux marchés auraient donc un impact positif sur l'emploi comme sur la croissance.

Mais beaucoup d'observateurs estiment que la quatrième révolution industrielle devrait engendrer plus de destructions que de créations et réorientations des emplois. La substitution de l'homme par la machine devrait frapper de très nombreux secteurs, de la comptabilité à la santé. En premier lieu, les emplois de bureau consistant à manipuler et organiser des données risquent fort de disparaître. Si le développement de l'intelligence artificielle provoque une baisse substantielle et pérenne de la demande de travail, le taux de chômage structurel pourrait augmenter. Les travailleurs les moins qualifiés et ceux dont les emplois demandent le moins de qualités humaines sont les plus exposés à ce risque de chômage technologique. L'IA, même faible, pourra effectuer toutes les tâches sans valeur humaine ajoutée (conduire un camion, tenir une comptabilité, traduire des textes ou des discours, diagnostiquer une maladie, par exemple). Cela est conforme à la théorie du progrès technologique biaisé, selon laquelle l'apparition de nouvelles technologies tend à détruire principalement les emplois peu qualifiés et à diriger la demande de travail vers le segment qualifié de l'offre de travail<sup>199</sup>. L'entrée massive des IA dans les entreprises pourrait donc accélérer, dans les pays développés, la segmentation du marché du travail. Néanmoins, si l'offre de capital humain s'ajuste en prévision des changements technologiques — ce à quoi il faut œuvrerurgemment en adaptant l'offre de formations et plus généralement tout le parcours éducatif —, les destructions brutes d'emploi pourraient rester limitées.

---

<sup>199</sup> L. Galiana, « Les conséquences économiques de l'intelligence artificielle », *Idées économiques et sociales* 2018, n° 192, p. 32.

Les réorganisations et mouvements sectoriels devraient constituer l'un des principaux effets de l'économie des données et des IA. Près d'un quart des emplois seraient concernés par une réallocation des tâches. Il faut s'attendre ou, du moins, espérer non des destructions massives d'emplois mais une réorganisation assez radicale du travail. Autant de défis pour l'éducation nationale, l'université et la formation professionnelle. Le travailleur humain devra se spécialiser dans les secteurs requérant des compétences multidisciplinaires et transversales, où il y a peu de données et de statistiques mais un besoin fort de sens commun, de souplesse, d'adaptabilité, de créativité, de culture générale, de sens critique, d'empathie, d'affectivité, d'humanité, de capacité à prendre des décisions arbitraires et à innover, ainsi que, parfois, de force physique et de dextérité. S'il n'y a plus besoin des hommes pour les métiers basiques, parce que l'IA s'en chargera plus efficacement et à un moindre coût, plus rapidement et avec moins d'erreurs et d'approximations, et si ne comptent plus que les capacités cognitives complémentaires de l'IA, alors il faudra non seulement inventer des métiers à haute valeur humaine ajoutée, mais aussi préparer les individus afin qu'ils puissent exercer ces métiers, former non des techniciens et des applicateurs mais des tacticiens et des créateurs, entre autres. Or le World Economic Forum s'inquiète du fait que les travailleurs, notamment français, seraient très mal armés pour s'adapter et affronter le changement.

La question de l'emploi est en tout cas l'une des principales clés de l'acceptation sociale de l'intelligence artificielle. Les révolutions industrielles ont toujours généré un moment d'angoisse généralisée. L'incertitude liée à la nouveauté et au brouillard entourant l'avenir, en premier lieu à un niveau micro-économique et à l'échelle de l'individu-salarié, provoque des inquiétudes. Mais les nouvelles technologies, jusqu'à présent, ont toujours bouleversé les modes de production sans pour autant entraîner des pertes d'emplois supérieures aux créations d'emplois. De nouveaux secteurs d'activité vont donner lieu à des métiers auxquels l'on n'a pas encore idée. Reste que l'adaptabilité et la souplesse, dans un tel contexte, seront les conditions d'une transition la moins chaotique possible vers l'ère de l'économie des données et des IA.

## IV. Concentrations dans le secteur de l'économie des données

La révolution de l'intelligence artificielle et des big data pose par ailleurs la question de l'évaluation de la valeur des données et des technologies telles que les algorithmes. Cette valeur ne correspond pas au prix payé par le

consommateur puisque les réseaux sociaux et autres moteurs de recherche, principaux services fonctionnant à base d'IA et de données aujourd'hui, sont proposés gratuitement en échange de l'abandon à ces acteurs des données de connexion et autres données personnelles. La valeur devrait correspondre à celle des données ainsi livrées aux multinationales du web. Mais, pour cela, il faudrait patrimonialiser ces données — c'est pourquoi on plaide pour un droit de propriété sur les données. Car, en échange de la gratuité apparente, les plateformes exploitent et revendent les données à des annonceurs à des fins de publicité ciblée. Ils les utilisent également pour leur propre compte, dans le but d'améliorer leurs services et donc de renforcer leurs positions sur les marchés. Reste que les utilisateurs profitent des effets positifs des externalités de réseau qui provoquent la concentration des usages vers un nombre restreint de fournisseurs de service. Tel est le cas, par exemple, dans le domaine des moteurs de recherche en ligne. La présence d'une structure centralisatrice diminue les coûts d'opportunité de la recherche pour les utilisateurs et garantit aux propriétaires des sites web l'accessibilité de leurs pages à tout le monde.

Un marché ordonné autour d'un tel mécanisme est qualifié de « marché à deux faces » (*« two-sided market »*). Il s'agit de la situation où une plateforme sert d'intermédiaire entre deux acteurs qui, sans elle, peineraient à s'apparier. Cette plateforme profite en théorie aux deux parties puisqu'elle réduit à zéro le coût des recherches, du point de vue de l'utilisateur, et elle diminue les asymétries d'informations sur les caractéristiques des utilisateurs des services, ce qui permet au producteur d'appliquer une tarification optimale. Ici, la valeur des données et des IA correspond au prix de réserve du fournisseur du bien ou service inclus dans la transaction, c'est-à-dire le prix ou la commission maximum que la firme est prête à payer pour pouvoir être mise en lien avec un potentiel consommateur<sup>200</sup>.

Concernant l'utilisateur, le gain de surplus correspondant à une centralisation accrue de la plateforme est difficile à cerner. Les externalités de réseau, qui baissent les coûts d'information (délicats à chiffrer), trouvent en contrepartie l'extraction d'une information permettant au consommateur de profiter d'une rente informationnelle. Cette dernière se définit comme le fait que le producteur ne connaisse pas les caractéristiques du consommateur et, ainsi, ne puisse extraire tout son surplus lors d'une transaction<sup>201</sup>. Sous l'angle des annonceurs, les bénéfices de l'exploitation de cette information présentent un coût correspondant à celui de la plateforme intermédiaire.

Une conséquence inévitable et directe des externalités de réseau est ainsi la création de structures oligopolistiques ou même monopolistiques, lorsqu'un seul ou quelques rares acteurs font main basse sur un marché. L'économie numérique est donc capturée par les GAFAM et BATX, incontournables à

---

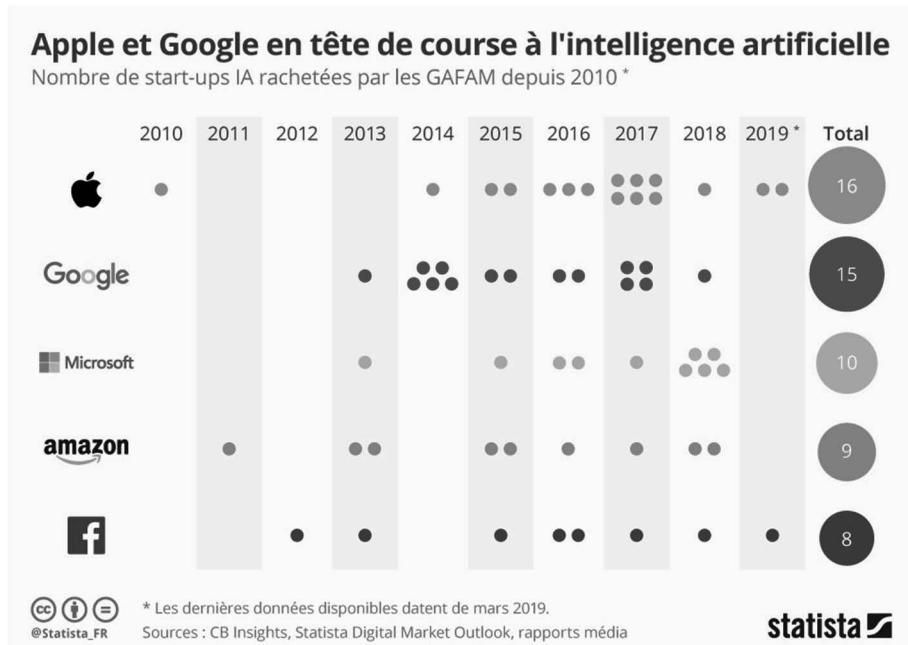
<sup>200</sup> *Ibid.*

<sup>201</sup> *Ibid.*

## *L'intelligence artificielle*

bien des points de vue. Or beaucoup de marchés des données et, par suite, des IA semblent voués à suivre un tel chemin, à moins que quelque régulateur souhaite intervenir — et qu'il possède les moyens de ses ambitions. Plus un service à base d'IA possède de données et plus il se perfectionne, si bien qu'il attire de nouveaux utilisateurs avec leurs données, lesquelles lui permettent de s'améliorer encore plus et de prendre davantage d'avance sur ses concurrents etc.

Les GAFAM sont aussi bien souvent tentés d'acheter ceux qui pourraient être leurs concurrents avant qu'ils ne le deviennent ou de racheter des start-up leur permettant d'étendre leurs empires vers de nouveaux services.



La mise en place d'une architecture de type oligopole ou monopole permet à la plateforme de s'accaparer la plus grande partie du surplus collectif de l'échange entre l'utilisateur et le fournisseur de bien ou service. La plateforme devient de la sorte l'acteur profitant le plus de la diminution des rentes informationnelles et des asymétries d'informations. Après avoir engendré une concentration du secteur autour des GAFAM, dans le monde, et des BATX, en Chine mais aussi de plus en plus dans le monde, ainsi qu'autour de quelques autres start-up devenues multinationales du numérique grâce à des applications bien pensées et utiles, les externalités du réseau tendent à

conforter la mainmise de certains acteurs sur la collecte de données<sup>202</sup>. Les plus forts deviennent toujours plus forts et sans rivaux. Cela est d'autant plus vrai que se produit un phénomène de transversalisation : ces acteurs investissent différents secteurs, pour ne pas dire tous les secteurs, de l'automobile à la santé. Auparavant, on n'aurait jamais imaginé un constructeur automobile se lancer dans la production de produits de santé, par exemple.

Le phénomène de concentration pouvant aller de pair avec la quatrième révolution industrielle résulte donc de l'utilisation massive des données et de leur exploitation par ceux qui en sont les propriétaires : les multinationales de l'IA à qui ces données sont cédées au moment de l'acceptation des conditions générales d'utilisation de leurs services. Les IA ayant besoin de très grandes quantités de données pour apprendre et donc devenir utiles à des utilisateurs potentiels, seuls quelques acteurs semblent pouvoir triompher, absorbant les autres et empêchant de nouveaux protagonistes de tenter leur chance sur les marchés émergents de l'économie des données et des IA. Quant aux principaux acteurs, ils se développent en proposant toujours plus de nouveaux services parallèles à leurs cœurs de métiers. Le secteur des données et des IA peut ainsi être qualifié d' « oligopole à frange »<sup>203</sup>. L'économie tend à s'organiser autour d'un nombre réduit d'acteurs en situation d'oligopole ou de monopole avec, gravitant autour d'eux, une galaxie de petits acteurs qui, lorsqu'ils innovent et grossissent, risquent fort de se faire absorber par les ogres qui les surveillent.

Une telle situation portant atteinte à la liberté de la concurrence et au bon fonctionnement du marché met au défi le régulateur — s'il existe. La concentration risque de s'accompagner de phénomènes de *hold-up*<sup>204</sup>, lorsqu'un acteur s'approprie l'intégralité ou presque des gains potentiels d'une transaction. Il faudrait qu'une autorité supérieure puisse arbitrer entre rémunération de l'acteur permettant les gains de surplus initiaux (gains d'efficience issus des externalités de réseau et surplus reliés aux échanges qui n'auraient pas eu lieu sans intermédiaire) et risque d'appropriation totale de ce surplus lorsque l'intermédiaire atteint une masse critique<sup>205</sup>.

---

<sup>202</sup> *Ibid.*

<sup>203</sup> G. J. Stigler, *The Organization of Industry*, Homewood, 1968.

<sup>204</sup> W. P. Rogerson, « Contractual Solutions to the Hold-Up Problem », *The Review of Economic Studies* 1992, vol. 59, n° 4, 1992, p. 777 s.

<sup>205</sup> L. Galiana, « Les conséquences économiques de l'intelligence artificielle », *Idées économiques et sociales* 2018, n° 192, p. 27.

## V. Leaders et suiveurs de l'économie des données

Traversant tous les secteurs et impactant tous les acteurs de l'économie, l'IA pourrait être, avec l'internet auquel elle est intimement liée, la rupture technologique majeure de l'époque contemporaine et du futur proche. C'est elle qui devrait faire le partage entre nouvelles et anciennes puissances économiques, redessinant le visage de l'économie mondiale. Cela concerne les entreprises, avec l'émergence en quelques années de multinationales dans des secteurs qui n'existaient pas il y a peu, mais aussi les États, certains s'imposant comme des leaders et d'autres étant cantonnés à la position de suiveurs. Partout l'on convient que l'intelligence artificielle doit être une technologie et les données une ressource qu'on maîtrise et non qu'on subit. Mais il est parfois difficile, en raison de la conjoncture et/ou du fait de choix stratégiques maladroits, de posséder les moyens de ses ambitions. Ainsi, tant au niveau des entreprises qu'au niveau des nations, certains sont dans la course, avec de l'avance sur leurs poursuivants, d'autres sont dans la course mais ont pris du retard et tentent de revenir sur la tête, quand d'autres ne sont même pas dans la course, ayant manqué le virage technologique et étant sortis de la route de la nouvelle économie.

Les capacités de stockage de données et de traitement de ces données par des algorithmes progressent en permanence. Y compris au-delà du secteur numérique, les IA et les *big data* attirent. Mais ce sont essentiellement les leaders des marchés du web, de l'informatique et des télécommunications qui profitent de cet engouement et de cette conjoncture. Leur expertise dans la collecte, le stockage et le traitement des données volumineuses, et surtout leur accès sans égal à de telles données volumineuses, leur permettent de proposer des services à base d'IA toujours plus pertinents et efficaces, renforçant leur domination sur les marchés concernés.

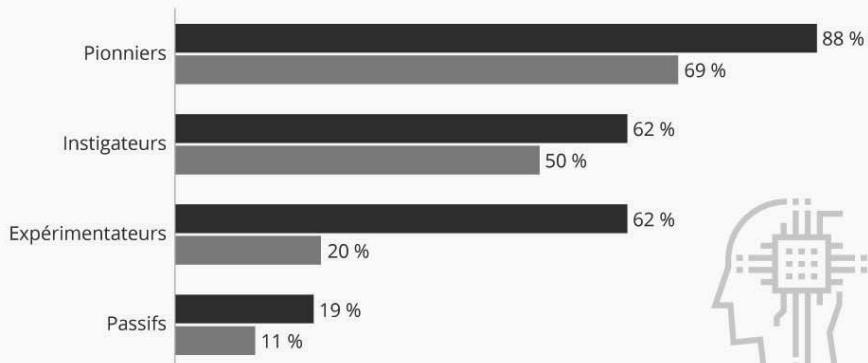
Ce sont donc essentiellement les GAFAM américains et les BATX chinois qui dominent le secteur de l'intelligence artificielle, car ils ont dès aujourd'hui un immense avantage compétitif par rapport à leurs concurrents et à d'éventuels nouveaux entrants. Déjà extrêmement puissants, ils pourraient l'être encore plus à court et moyen termes, lorsque l'IA aura réellement provoqué la rupture et une nouvelle révolution industrielle. Les données étant au cœur des nouveaux modèles économiques, plus un acteur a accès à beaucoup de données, plus il possède un potentiel de développement important. Il peut alors imposer à tous la façon dont ces données sont utilisées.

Le graphique suivant montre que les sociétés les plus en avance en matière de technologies d'intelligence artificielle sont celles qui, actuellement, investissent le plus dans l'IA et s'intéressent le plus aux enjeux de l'IA. Ainsi les écarts se creusent-ils dans la course à l'IA.

## Les écarts se creusent dans la course à l'IA

Niveau d'adoption et de compréhension de l'IA dans les groupes d'entreprises suivants \*

- Part ayant augmenté ses investissements dans l'IA l'an passé
- Part ayant amélioré sa compréhension des enjeux de l'IA l'an passé



\* Enquête mondiale menée auprès de 3 076 cadres d'entreprise. Les organisations enquêtées ont été classées en 4 groupes selon leur niveau d'adoption et de compréhension de l'IA.

Du plus faible au plus fort niveau : Passifs, Expérimentateurs, Instigateurs et Pionniers.



Source : Boston Consulting Group

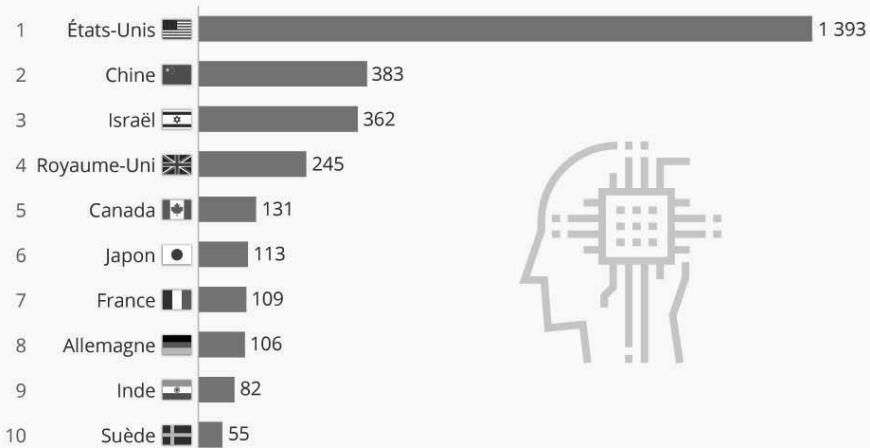
statista

De mêmes observations peuvent viser les pays : ceux qui sont déjà leaders en matière d'économie de l'IA et des données sont ceux qui investissent le plus dans le secteur, qui s'y intéressent le plus, de telle sorte que les écarts se creusent entre pays-leaders et pays-suiveurs. La Chine et les États-Unis sont ainsi ceux qui tirent les plus gros bénéfices de la diffusion de l'IA dans l'économie. La Chine, en particulier, investit massivement dans l'IA, y voyant le levier susceptible de lui permettre de devenir la première puissance mondiale et de générer un tiers du PIB mondial<sup>206</sup>. Puis viennent Israël, le Royaume-Uni et le Canada, mais loin derrière. Les autres sont encore plus en retrait, peinant à encourager et protéger de nouveaux acteurs viables dans le secteur. Le classement des pays comptant le plus de start-up spécialisées en IA reflète cette hiérarchie des puissances économiques.

<sup>206</sup> La technologie devrait ajouter 7 000 milliards de dollars (790 000 milliards de yens) à son produit intérieur brut d'ici 2030.

## Les pays pionniers de l'intelligence artificielle

Top 10 des pays comptant le plus de start-ups dans l'intelligence artificielle en 2018

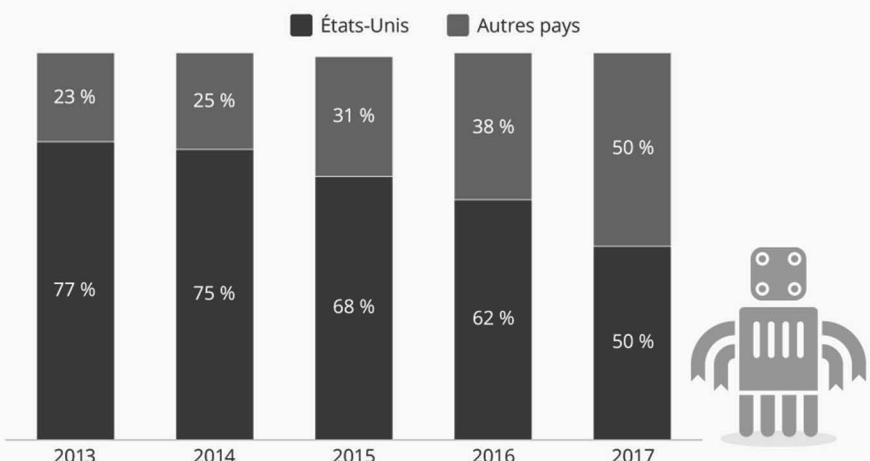


@Statista\_FR Source : Roland Berger

statista

## La fin de la mainmise des États-Unis sur l'IA

Provenance des investissements dans les start-ups travaillant sur l'intelligence artificielle



Données arrondies.  
Source : CB Insights

statista

Pour dépasser les États-Unis et dominer le monde dans le secteur de l'IA et des données, la Chine s'appuie sur trois piliers : des budgets de recherche et développement conséquents<sup>207</sup>, de nombreux spécialistes compétents et reconnus<sup>208</sup> et des données massives accessibles sans entrave grâce à la philosophie confucéenne qui conduit à penser que l'attitude condamnable et dangereuse serait de vouloir protéger ses données personnelles, au détriment du collectif, de la nation et de son économie florissante ; or le *big data* est bien la première source de compétitivité à l'ère de l'IA<sup>209</sup>. La Chine possède donc l'avantage énorme d'un environnement propice au développement de l'économie des données et de l'intelligence artificielle.

Le 15 novembre 2017, le gouvernement chinois a lancé un projet titanique de développement de l'industrie de l'IA, se tournant ainsi tout entier vers ce qui pourrait être la quatrième révolution industrielle<sup>210</sup>. Le grand plan IA de la Chine prévoit d'y consacrer 20 milliards d'euros par an et jusqu'à 59 milliards d'euros d'ici 2025. Une nouvelle organisation a ainsi été créée afin d'organiser le secteur, associant plusieurs institutions gouvernementales, y compris le ministère des Sciences et de la Technologie, et les géants du secteur que sont les BATX. Une telle initiative « pan-chinoise » mêlant les secteurs public et privé serait difficilement concevable en Europe ou en Amérique du Nord. Et il est remarquable qu'Eric Schmidt, président exécutif de la société mère de Google, Alphabet Inc., lors d'une conférence en novembre 2017, exprimait son sentiment de crise face à la croissance rapide de l'IA en Chine en le comparant au choc ressenti par les américains en 1957 lorsque l'Union Soviétique a lancé avec succès Sputnik, le premier satellite

---

<sup>207</sup> Selon l'Institut national des politiques scientifiques et technologiques (NISTEP), le budget chinois consacré à la science et à la technologie a continué de croître, atteignant six fois le budget du Japon en la matière. Et les sociétés privées investissent également des montants colossaux dans la recherche et le développement.

<sup>208</sup> Une enquête de l'Agence japonaise de science et de technologie (JST) relative aux articles scientifiques relatifs à l'IA les plus cités dans le monde indique que la recherche chinoise représente 21 % d'entre eux, soit un niveau plus élevé que celui des travaux de chercheurs américains.

<sup>209</sup> Selon les statistiques officielles fournies par le gouvernement, il y avait plus de 750 millions d'internautes en Chine début 2018. Ce nombre est plus de deux fois supérieur à la population américaine (320 millions). 96 % de ces internautes possèdent des ordinateurs-téléphones portables, ce qui signifie qu'ils sont en ligne quotidiennement. Outre les informations concernant le nom, l'âge et le sexe, peuvent également être collectées et analysées par l'IA les données de géolocalisation, celles sur les produits et services achetés et celles relatives aux objets des recherches en ligne, cela pour permettre aux entreprises d'offrir plus facilement des services optimaux.

<sup>210</sup> L'objectif de la Chine est d'augmenter l'échelle du marché de l'IA à 150 milliards de yuans d'ici 2020, 400 milliards de yuans d'ici 2025 et 1 000 milliards de yuans d'ici 2030.

artificial. Mais les puissances les plus mises en difficulté par la révolution des intelligences artificielles sont le Japon et les pays d'Europe occidentale, dont la France.

## VI. La stratégie IA de la France

De plus en plus de voix s'élèvent pour critiquer le fait que, quand les technologies et l'économie de l'IA avancent à grands pas sur les autres continents, la France en serait toujours à constituer des comités, à commander des commissions et à publier des rapports et des recommandations d'éthique. Au moment où a été confié au mathématicien et député Cédric Villani une mission sur l'intelligence artificielle, après le rapport FranceIA et celui de l'office parlementaire d'évaluation des choix scientifiques et technologiques, il semblait effectivement que la France risque ne pas entrer aussi pleinement que les États-Unis ou la Chine dans la quatrième révolution industrielle. En Europe, les limites éthiques et, notamment, la protection des données personnelles, avec le *Règlement général sur la protection des données* (RGPD), limitent la capacité des entreprises de suivre leurs concurrentes qui, ailleurs dans le monde, progressent vite, sans achopper sur des barrières relevant des droits et libertés fondamentaux — dont il ne s'agit pas de nier la légitimité ni même la nécessité mais qui sont de réels freins à l'ère de l'économie des données.

Les opportunités économiques relevant de l'intelligence artificielle sont pourtant, en France en particulier, nombreuses. Quelques 80 entreprises de taille intermédiaire (ETI) et plus de 270 start-up spécialisées dans le secteur ont été créées depuis 2010, avec une croissance soutenue à plus de 30 %. Surtout, après le temps des comités et des rapports, le gouvernement français a décidé d'investir 1,5 milliard d'euros d'ici 2022 afin de développer rapidement l'économie de l'intelligence artificielle. La France se lance ainsi pleinement dans la course, peut-être avec un temps de retard, souhaitant densifier les investissements dans les start-up, soutenir totalement les plus prometteuses et attirer des investisseurs étrangers. BPI France a recensé 552 sociétés françaises ayant levé plus de 2 millions d'euros et offrant des solutions d'IA. Elles semblent ainsi de plus en plus convaincues que l'IA les concerne toutes, qu'elle touche tous les secteurs, tous les acteurs et tous les métiers.

Si les pouvoirs publics souhaitent que beaucoup de rapports leur soient remis, la mise en œuvre de l'ensemble des préconisations est bien sûr impossible car trop coûteuse. Ils doivent donc procéder à des décisions stratégiques adaptées à leurs capacités d'investissement. Le 3 juillet 2019, un

an après la présentation de la stratégie nationale pour l'intelligence artificielle par le Président de la République Emmanuel Macron, et quelques mois après l'exposé des volets recherche, défense et éthique — ce qui est peut-être significatif d'un certain axe des priorités —, le Premier Ministre, Édouard Philippe, a présenté le volet industriel et économique de cette stratégie nationale. Les objectifs, a-t-il expliqué, sont de faire émerger les futurs « champions français de l'IA ». La priorité est donc donnée au développement et à la structuration d'une véritable offre française en la matière<sup>211</sup>. Bruno Le Maire, Ministre de l'Économie et des Finances, a pour sa part désigné les leviers sur lesquels le Gouvernement entend agir afin d'encourager cette offre : le financement, les talents et l'accès aux données — soit exactement les mêmes que ceux visés par les pouvoirs publics chinois deux ans plus tôt ; or, deux ans de décalage, à la vitesse à laquelle l'IA se déploie, est peut-être un retard trop important.

Concernant le financement, plusieurs services sont proposés par la Direction Générale des Entreprises (DGE), en lien avec BPI France et le Secrétariat Général pour l'Investissement (SGPI). Il s'agit, par exemple, des Challenges IA, financés à hauteur de 5 millions d'euros par le Programme d'investissements d'avenir (PIA) et qui visent à mettre en relation offreurs et utilisateurs d'IA pour résoudre des problèmes concrets dans les quatre secteurs mis en avant par la mission Villani : santé, transport-mobilité, environnement et défense-sécurité. Mais, pour Bruno Le Maire, « la bataille la plus essentielle est celle des talents. Il faut former et disposer des compétences clés. Or, aujourd'hui, nous manquons de *data scientists*, d'ingénieurs, de développeurs, de chercheurs en IA. Nous allons mettre le paquet sur ce sujet-là. C'est notre priorité ». Il est vrai que, si l'excellence française est reconnue, le pays manque de jeunes diplômés et qualifiés pour répondre aux besoins des entreprises.

La France est confrontée au défi de la « fuite des cerveaux ». C'est l'un des principaux freins qui empêchent le développement optimal de l'intelligence artificielle dans l'hexagone. Dans son rapport de mars 2018, Cédric Villani a ainsi insisté sur le fait que, bien que le niveau des formations dispensées en France compte parmi les meilleurs au monde, de nombreux mathématiciens, informaticiens et ingénieurs choisissent de s'expatrier en raison des conditions plus avantageuses offertes aux chercheurs à l'étranger. Cédric Villani place donc au cœur des enjeux prioritaires le besoin de renforcer l'attractivité des centres de recherche français, non seulement en augmentant les salaires des

---

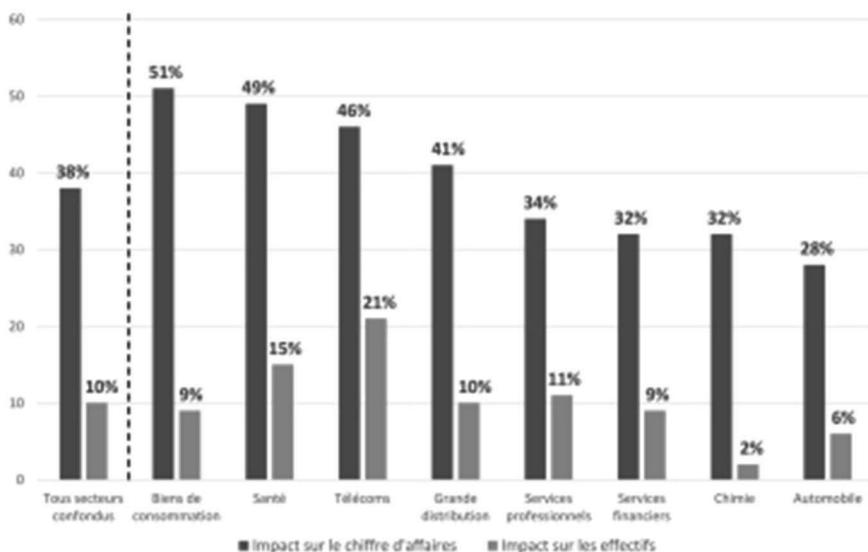
<sup>211</sup> Huit entreprises (Thales, Valeo, Air Liquide, Dassault Aviation, EDF, Renault, Safran et Total) étaient présentes lors de ces annonces et ont signé avec le gouvernement le « Manifeste pour l'intelligence artificielle au service de l'industrie ». Le but de cette initiative ouverte est de construire une base stratégique commune en matière d'IA et de données coordonnée avec la stratégie nationale.

chercheurs, mais aussi en améliorant leur qualité de vie au sein des laboratoires. Selon lui, la pression administrative, les difficultés de recrutement et d'évolution ou encore la mise à disposition limitée de matériel seraient les principaux motifs des départs à l'étranger.

Les pouvoirs publics français souhaitent « développer et structurer une offre française en IA, diffuser une IA accessible à toutes les entreprises et développer une économie française et européenne de la donnée ». Les obstacles éthiques, juridiques et même systémiques, en raison de la taille du marché français et de son mode de fonctionnement, sont cependant lourds et délicats à contourner.

## VII. Souplesse, flexibilité, adaptabilité

Dans de multiples secteurs, l'intelligence artificielle peut impacter positivement le chiffre d'affaires et les effectifs :



Pour en profiter, les entreprises doivent miser sur ces nouvelles technologies et s'y adapter, loin de tout conservatisme et de toute passivité. En faisant évoluer leurs organisations afin de mieux préparer leurs salariés à la collaboration homme-machine, elles pourraient augmenter leurs chiffres

d'affaires de 38 % et leurs effectifs de 10 % d'ici à 2022<sup>212</sup>. Les opportunités de croissance des entreprises dépendraient ainsi fondamentalement de leur capacité à évoluer et à s'adapter. Et un même constat peut sans doute être dressé concernant l'action des pouvoirs publics afin de porter l'économie des données et de l'IA.

Les dirigeants et leurs collaborateurs semblent plutôt optimistes à l'égard de la capacité de l'IA à engendrer une forte croissance économique et à créer plus d'emplois qu'elle n'en détruirait. 72 % d'entre eux estiment ainsi que les technologies intelligentes seront le facteur déterminant de la capacité de leurs entreprises à se différencier ; et 61 % jugent que le nombre d'emplois requérant une collaboration entre homme et machine augmentera au cours des trois prochaines années ; tandis que 69 % des salariés s'accordent sur l'importance pour eux d'acquérir de nouvelles compétences afin de pouvoir travailler aux côtés des IA<sup>213</sup>.

Peut-être faut-il protéger non les emplois mais les employés ; et alors les mots d'ordre seraient bien souplesse, flexibilité, adaptabilité. Il importeraît de réinventer le travail, à l'échelle macroéconomique et, plus encore, à l'échelle microéconomique de chaque entreprise. Celle-ci peut en premier lieu répartir les tâches — et non les emplois — entre hommes et machines en cherchant le meilleur équilibre entre le besoin d'automatisation des calculs et de la gestion des données et le besoin de valeur humaine ajoutée. Plus généralement, peut-être les mutations technologiques actuelles et à venir appellent-elles un état d'esprit ou une philosophie adapté, pour que chaque collaborateur soit en mesure de concevoir de nouvelles expériences clients et de s'impliquer dans la croissance future de son entreprise ; et pour que les dirigeants soient ouverts au besoin de souplesse, à tous les niveaux, permettant d'identifier et saisir les nouveaux relais et opportunités de développement.

Ce qui frappe avec la quatrième révolution industrielle, dont on vit seulement les prémisses, est la vitesse à laquelle les compétences des salariés et les positionnements, les innovations et les organisations d'ensemble des entreprises doivent évoluer, pour suivre l'accélération des mutations technologiques et des bouleversements économiques qui les accompagnent. L'économie des données et de l'IA qui semble devoir s'imposer à court et moyen termes devrait faire reposer la compétitivité et la croissance sur la capacité à combiner intelligence artificielle et intelligence humaine. Mieux vaut s'équiper et se mettre à niveau au plus vite, car les luttes sur les marchés traditionnels et sur les nouveaux marchés seront, comme toujours,

---

<sup>212</sup> Étude d'Accenture publiée à l'occasion du Forum Economique Mondial de Davos de 2018.

<sup>213</sup> *Ibid.* Selon cette étude, 63 % des dirigeants pensent que l'IA permettra une création nette d'emplois dans leurs entreprises au cours des trois prochaines années. Et 62 % des employés estiment que l'IA aura un impact positif sur leur travail.

impitoyables. La différence entre concurrents se fera sur leur capacité à tirer le meilleur parti des compétences humaines et à les associer à la puissance des nouvelles technologies.

## Conclusion

L'entrée dans l'ère de l'économie des données et des IA pourrait rendre les entreprises, dans l'ensemble, plus productives et plus compétitives. En premier lieu, les salariés sont supposés pouvoir travailler plus efficacement. Par exemple, il est aujourd'hui déjà devenu courant de travailler depuis son domicile sans perte d'efficacité et même en gagnant — loin de l'époque des chaînes de travail du type Taylor. Cela pourrait par exemple entraîner des conséquences sur l'aménagement du territoire. Pouvoir travailler virtuellement à Paris tout en restant au fond de la campagne franc-comtoise ou bretonne serait une réponse au problème de la saturation du territoire, tant en termes de logements que du point de vue des transports. Mais cela supposerait une connectivité performante, un développement suffisant des réseaux de télécommunications. De mêmes effets bénéfiques pourraient exister s'agissant des accidents de travail, de la pénibilité au travail ou du coût écologique du travail. L'IA permettrait de telles ruptures aux avantages macro-économiques divers et importants.

Ce qui est nouveau fait pourtant toujours peur, y compris aux acteurs économiques — et pas nécessairement pour de mauvaises raisons. La résistance au changement peut être tenace, nourrie par des fantasmes et des discours sensationnalistes plus ou moins détachés des réalités. Il importe donc, sans les minimiser, de dédramatiser les changements pour pouvoir les affronter lucidement et prendre les bonnes décisions, du point de vue microéconomique d'une entreprise ou sous un angle macroéconomique s'agissant des États, des gouvernants ou encore des autorités de régulation. Il importe notamment de rappeler que l'IA n'est qu'un outil créé par l'homme et au service de l'homme, neutre en soi et dépendant entièrement des usages que celui-ci en fait. L'humain reste et restera donc au centre du jeu, maître de son activité.

Chaque acteur, du salarié à l'entreprise en passant par l'État et par l'autorité de régulation, doit préparer l'avenir et se préparer à l'avenir. Il s'agit déjà de songer aux meilleures manières de coordonner, associer et même fusionner l'intelligence humaine et l'intelligence artificielle, cela afin qu'elles apprennent l'une de l'autre et se fassent progresser mutuellement. L'IA et l'IH ne serviront à rien si elles sont envisagées séparément, comme ayant chacune son domaine de compétences et ses tâches dédiées.

Du point de vue public et étatique, il faudra produire de nouvelles lois et de nouveaux cadres juridiques afin de diminuer l'écart entre l'évolution des technologies et celle des régulations associées. Il conviendra aussi de ne pas perdre de vue les défis éthiques que l'IA pose et auxquels il importe de répondre de manière équilibrée : en consacrant des normes protectrices des droits et libertés fondamentaux sans sacrifier pour autant toute une économie et, étant donnée l'importance de cette économie, tout un pays. Au-delà, les acteurs devraient promouvoir et respecter des bonnes pratiques relatives au développement et à l'utilisation des IA — il n'est pas insensé de penser que l'autorégulation pourrait être efficace en la matière. Enfin, les pouvoirs publics doivent insister sur les bénéfices économiques tangibles et durables auxquels l'IA peut donner lieu et gérer de manière préventive et de façon réactive ses conséquences, notamment en aidant les acteurs en difficulté, car positionnés sur les marchés les moins porteurs, à se reconvertis.

Comme lors de chaque révolution industrielle, un temps de transformation est indispensable afin de permettre à tous les marchés, y compris le marché du travail, de s'adapter aux nouveaux besoins. C'est pourquoi la formation semble bien être l'une des clés d'une mutation réussie. En France, on prévoit ainsi de doubler le nombre d'étudiants formés à l'IA — et à l'éthique de l'IA — au cours des prochaines années.

Reste que l'intelligence artificielle, spécialement à travers les progrès de la robotique qu'elle permet, pourrait bel et bien être le pilier essentiel d'un bouleversement de l'économie analogue en amplitude à celui qui a vu l'emploi industriel l'emporter sur l'emploi agricole, ou à celui qui a vu l'emploi dans les services prendre le pas sur l'emploi dans les usines.

C'est donc une nouvelle guerre économique qui s'annonce autour des immenses gains de productivité et nouvelles sources de profits promis par l'intelligence artificielle ; une guerre économique qui ne cache même pas son nom au vu des discours et des investissements du gouvernement chinois, qui a érigé l'IA et les données en grande cause nationale. La question fondamentale qui se pose alors est celle de la gouvernance globale de cette nouvelle révolution industrielle et des luttes entre États et entre multinationales qui vont l'accompagner. Quelles institutions et quelles formes de puissance permettront à l'humanité et à la société mondiale de profiter le plus possible de l'entrée dans l'ère des IA et des *big data* ?

# L'intelligence artificielle

*Dimension anthropologique*



# Une intelligence artificielle anthropocentréée

**Magalie Ochs**

*« There's nothing artificial about AI. It's inspired by people, and — most importantly — it impacts people »*

Li Fei-Fei, co-director of Stanford University's Human-Centered AI Institute, 2018

Le biomimétisme est inhérent à l'intelligence artificielle. L'essor des algorithmes d'apprentissage automatique et en particulier des réseaux de neurones profonds, au cœur de l'intelligence artificielle, s'inspire du cerveau humain. Cependant, l'intelligence humaine et l'intelligence artificielle des machines restent divergentes, en particulier relativement à leur capacité de calcul, aujourd'hui incomparable, qui permet à une machine de traiter d'immenses volumes d'informations. L'intelligence artificielle permet aujourd'hui de résoudre un certain nombre de tâches dans des domaines aussi variés que la reconnaissance d'image, le traitement automatique de la langue naturelle, la détection de fraude ou le ciblage marketing, et cela principalement grâce au développement récent d'algorithmes nécessitant à la fois beaucoup de données et une grande puissance de calcul. Toutefois, l'intelligence d'une machine ne repose pas seulement sur une puissance de calcul mais aussi, à l'image de l'humain, sur une intelligence sociale et émotionnelle.

La plupart des communications humain-machine sont en effet socialement situées dans un contexte dans lequel le système interactif, représenté par exemple sous la forme d'Agent Conversationnel Animé (ACA) ou de robot humanoïde, incarne un rôle social particulier, tel que celui de tuteur virtuel, de

compagnon ou de coach<sup>214</sup>. En effet, les changements sociaux profonds ont amené de nouveaux usages et de nouvelles pratiques de l'informatique pour répondre aux besoins émergents tels que l'e-éducation, la formation ou la santé. Dans ces contextes d'usages, les systèmes interactifs doivent être dotés d'une certaine forme d'intelligence sociale et émotionnelle pour incarner ces rôles sociaux avec succès. L'intelligence sociale se définit comme la capacité de reconnaître et d'exprimer des comportements sociaux, tels que la politesse ou la dominance, et la capacité de gérer ces comportements pour construire une relation sociale avec un autre individu<sup>215</sup> et l'amener à coopérer<sup>216</sup>. L'intelligence émotionnelle se définit quant à elle comme la capacité de reconnaître, exprimer et gérer ses propres émotions et celles d'autrui<sup>217</sup>.

Du point de vue de la machine, cette intelligence socio-émotionnelle est d'autant plus importante que l'utilisateur lui-même a une propension à adopter un comportement social face à ces entités artificielles : c'est le paradigme CASA (« *Computer Are Social Actors* »). Beaucoup de travaux sur cette thématique ont été notamment conduits par Reeves, Nass et leurs collègues montrant, par exemple, la tendance naturelle des individus à adopter des comportements de politesse envers la machine. Les individus ont, de plus, tendance à anthropomorphiser les systèmes interactifs, même caricaturaux, en leur attribuant un genre, une ethnie et une personnalité. Les comportements sociaux des utilisateurs, qui savent qu'ils interagissent avec une machine, mais qui se comportent comme si celle-ci était humaine, sont déclenchés automatiquement et de manière inconsciente<sup>218</sup>. D'autres recherches ont confirmé ces comportements sociaux des utilisateurs dans le contexte d'interactions avec des agents conversationnels animés<sup>219</sup>.

Dans la perspective de doter les systèmes interactifs d'une intelligence sociale et émotionnelle sont nés deux courants de recherche en Intelligence Artificielle : l'*Affective Computing*<sup>220</sup> et le *Social Signal Processing* (SSP)<sup>221</sup> (ou *Socially Aware Computing*<sup>222</sup>). Ces courants de recherche ont comme objectifs de développer des systèmes interactifs capables de reconnaître, comprendre et exprimer des émotions et des comportements sociaux. L'intelligence socio-émotionnelle des machines est construite dans une approche méthodologique anthropocentrique reposant sur l'analyse de

---

<sup>214</sup> Albaina et al., 2009, Hoque et al., 2013, Anderson et al., 2013.

<sup>215</sup> Kihlstrom and Cantor, 2000.

<sup>216</sup> Albrecht, 2006.

<sup>217</sup> Mayer et al., 2004.

<sup>218</sup> Nass and Moon, 2000.

<sup>219</sup> Hoffmann et al., 2009, Kopp et al., 2005, Krämer, 2008.

<sup>220</sup> Picard, 1997.

<sup>221</sup> Vinciarelli et Pentland, 2015.

<sup>222</sup> Pentland, 2005.

comportements humains à partir de méthodes d'Intelligence Artificielle (I). Le développement de telles machines a permis l'émergence d'environnements virtuels d'apprentissage permettant aux humains d'améliorer leur propres compétences sociales et émotionnelles (II). L'envers du décor d'une intelligence artificielle anthropocentré est aussi celui de l'intégration au sein des mondes virtuels des stéréotypes de genre et des discriminations omniprésents dans nos sociétés et donc dans nos données (III).

## I. Une intelligence socio-émotionnelle artificielle

Pour créer des personnages virtuels ou des robots humanoïdes dotés d'une intelligence socio-émotionnelle artificielle, ces derniers doivent être enrichis de modèles informatiques leur permettant de raisonner sur les émotions et les comportements sociaux. En particulier, l'enjeu scientifique est de concevoir des :

- Modèles de perception pour reconnaître les comportements socio-émotionnels des utilisateurs afin de reconnaître en temps réel et automatiquement les signaux sociaux et les émotions exprimés par l'utilisateur suivant le contexte de l'interaction ;
- Modèles d'expression pour exprimer des comportements sociaux et des émotions, *i.e.* exprimer à travers un comportement multimodal (verbal, non-verbal et para-verbal) des signaux sociaux et émotionnels afin de rendre compte de comportements sociaux ou d'émotions des systèmes interactifs ;
- Modèles de raisonnement pour raisonner sur les comportements sociaux et émotionnels à exprimer pour optimiser l'interaction, *i.e.* être capable de déterminer le comportement social à adopter ou l'émotion à exprimer étant donné le contexte de l'interaction afin de préserver et d'optimiser la crédibilité du système, la satisfaction de l'utilisateur et les performances du système et de l'utilisateur dans la réalisation d'une tâche.

Pour implanter ces capacités d'intelligence sociale dans des systèmes interactifs, l'un des enjeux est de modéliser les interactions sociales humain-humain, *i.e.* d'identifier les paramètres des comportements socio-émotionnels à la fois intra-personnels mais aussi interpersonnels, comprendre leurs interactions, leurs causes et leurs effets.

L'approche méthodologique pour doter les machines d'une intelligence socio-émotionnelle est profondément anthropocentré. Elle nécessite en effet de mieux comprendre l'humain pour pouvoir ensuite simuler des comportements socio-émotionnels sur une machine.

Pour construire les modèles informatiques socio-émotionnels (de perception, d'expression ou de raisonnement), plusieurs approches méthodologiques peuvent être explorées. Une première approche consiste à se fonder sur les travaux théoriques et empiriques sur l'humain mettant en évidence les liens entre les signaux socio-émotionnels, les comportements exprimés et les états mentaux en considérant l'influence du contexte social. Il s'agit en d'autres termes de construire un corpus réunissant les études empiriques et théoriques disponibles relatives aux émotions et aux comportements sociaux des individus. La problématique consiste alors à traduire des données souvent qualitatives, incomplètes et non-unifiées en modèles computationnels complets, automatisables, symboliques et/ou quantitatifs. Les premiers travaux en Informatique Affective se sont principalement appuyés sur une approche que l'on pourrait qualifier de « *top-down* ». Cette approche consiste à explorer les travaux sur l'humain pour en extraire manuellement des règles permettant de lier des états mentaux à des états socio-émotionnels, eux-mêmes liés à des comportements expressifs. En d'autres termes, il s'agit de comprendre ce qui conduit un individu à ressentir ou exprimer un état socio-émotionnel particulier et comment ce dernier est exprimé, pour ensuite pouvoir simuler ces comportements au sein d'un système interactif. Plusieurs travaux de recherche se sont appuyés sur cette approche pour apporter une dimension socio-émotionnel aux systèmes interactifs humanoïdes. En particulier, cette approche est particulièrement adaptée au développement de modèles de raisonnement intégrant des règles socio-émotionnelles issues de la littérature en Psychologie Cognitive. Par exemple, le module ALMA permet de calculer automatiquement et en temps réel les émotions d'un personnage virtuel ou d'un robot suivant sa personnalité et son humeur<sup>223</sup>. Ce module informatique a été développé à partir des théories en psychologie cognitive sur les émotions. Un module permettant de calculer les relations sociales des personnages virtuels dans les jeux à partir d'un calcul automatique de leurs émotions est proposé<sup>224</sup>. Il vise à éviter des comportements scriptés et répétitifs des personnages des jeux et permettre une adaptation automatique suivant le comportement du joueur.

Une deuxième approche pour intégrer des compétences socio-émotionnelles dans les systèmes interactifs humanoïdes consiste à exploiter des données réelles d'interaction pour construire ces modèles. Il s'agit alors de recueillir un corpus d'interactions humain-humain ou humain-machine. Dans cette approche, la problématique repose à la fois sur la collecte de données réelles et sur l'extraction automatique des informations permettant la modélisation socio-émotionnelle des comportements de l'agent. Cette approche est particulièrement utilisée aujourd'hui pour construire des modèles de perception et d'expression des robots et des personnages virtuels. Elle repose sur des

---

<sup>223</sup> Gebhard, 2005.

<sup>224</sup> Ochs, et al., 2009.

techniques récentes d'intelligence artificielle de fouille de données et d'apprentissage automatique. Par exemple, des algorithmes de fouille de données ont été utilisés pour identifier les séquences de signaux non-verbaux exprimés par un individu caractérisant les attitudes sociales de dominance et d'appréciation<sup>225</sup>. À partir des fichiers vidéo d'interaction humain-humain annotés (annotations discrètes des signaux non-verbaux et continues des attitudes sociales), des trames de signaux non-verbaux ont été créées. Ces séquences de signaux ont été corrélées à des variations d'attitudes sociales (*e.g.* augmentation de dominance, diminution de l'appréciation) à partir d'algorithmes d'apprentissage automatique. Ces séquences extraites ont été utilisées pour permettre à un recruteur virtuel de déterminer quels signaux sociaux exprimer en fonction de l'attitude sociale visée (*Figure 1*).



*Figure 1 : Image d'écran du recruteur virtuel capable d'exprimer différentes attitudes sociales suivant le niveau de difficulté de l'entretien*

Un autre exemple de développement est l'instrument *open source* OpenFace<sup>226</sup>, construit par apprentissage sur de grandes masses de données, permettant de détecter en temps réel les émotions des utilisateurs à partir de leurs expressions faciales. La reconnaissance des émotions est cependant limitée par la problématique de l'universalité des expressions. Seules les

---

<sup>225</sup> Chollet et al., 2017.

<sup>226</sup> Baltrušaitis, 2016.

émotions dites basiques (joie, peur, colère, surprise, dégoût, regret et tristesse) sont reconnues efficacement. Concernant les comportements sociaux, leur reconnaissance automatique et leur expression à travers un agent artificiel restent des problématiques de recherche ouvertes sur lesquelles travaillent aujourd’hui plusieurs équipes, en France et à l’international. Les problématiques de recherche pour doter les systèmes interactifs humanoïdes ne sont pas une question de puissance de calcul ou de performance d’algorithmes d’apprentissage propres à l’Intelligence Artificielle. Créer une intelligence socio-émotionnelle artificielle nécessite en effet de surtout mieux comprendre l’humain dans une approche interdisciplinaire.

## II. Améliorer l’humain à travers les machines : l’informatique prosociale

Les systèmes interactifs humanoïdes peuvent être utilisés pour promouvoir les coopérations entre les individus et plus généralement les comportements prosociaux<sup>227</sup>. Les questionnements sur l’utilisation de l’Intelligence Artificielle pour le bien de la société sont récents. Différents événements sont organisés en ce sens au niveau mondial (par exemple, le workshop AAAI WS : « *AI and Operations Research for Social Good* », en 2017, ou le « *AI for Good Global Summit* » de Genève, en 2017 également). Ainsi émerge un nouveau courant de recherche, l’informatique prosociale, qui se définit comme « l’informatique visant à soutenir et à promouvoir des actions qui profitent à la société et aux autres »<sup>228</sup>. Le comportement des systèmes interactifs humanoïdes peut en effet influencer le comportement humain. Par exemple, des travaux montrent que le simple regard d’un robot peut déclencher un comportement altruiste chez l’humain<sup>229</sup>. L’empathie, un élément essentiel au comportement prosocial, a été largement explorée par plusieurs chercheurs en Informatique afin d’intégrer cette composante dans les systèmes interactifs humanoïdes<sup>230</sup>.

Dans le domaine de l’e-éducation et de l’e-formation, un intérêt croissant est apparu autour de la « formation par la simulation » dans des environnements virtuels. De nombreux environnements de réalité virtuelle ou mixtes ont été développés afin d’améliorer certaines compétences techniques. Plus récemment, un certain nombre de travaux de recherche ont été menés

---

<sup>227</sup> Paiva et al., 2018.

<sup>228</sup> Paiva et al., 2018.

<sup>229</sup> Burnham et Hare, 2007.

<sup>230</sup> Ochs et al., 2008 ; Paiva et al., 2017.

autour de dispositifs permettant de simuler une interaction sociale avec un agent conversationnel animé (ACA) et ainsi entraîner ses propres compétences sociales (« *Virtual Agents for Social Skills Training (VASST)* »<sup>231</sup>). Les compétences sociales, partie intégrante de l'intelligence sociale, se définissent comme la capacité de gérer son comportement, verbal et non-verbal, pour construire une relation sociale avec autrui<sup>232</sup>.

Les recherches montrent que les systèmes interactifs humanoïdes peuvent permettre d'améliorer ces compétences interactionnelles<sup>233</sup>. Par exemple, dans le projet européen Tardis<sup>234</sup>, un personnage virtuel jouant le rôle d'un recruteur est utilisé pour former les jeunes adultes aux entretiens d'embauche. Le système permet de détecter en temps réel le comportement verbal et non-verbal du candidat. Cette détection permet de calculer des indices comportementaux reflétant la qualité de l'entretien (*e.g.* disfluences, direction du regard, posture) et d'obtenir un retour sur l'entretien à la fois du point de vue du contenu verbal et du point de vue du contenu non-verbal.

Dans le projet eCUTE<sup>235</sup>, les enfants et les jeunes adultes sont confrontés à différents scénarios d'interactions avec des personnages virtuels pour les sensibiliser aux différences culturelles et développer leur empathie. La plateforme FearNot!<sup>236</sup> a été développée pour lutter contre les comportements de harcèlement scolaire à l'école. Elle repose sur la propension des individus à ressentir de l'empathie envers les systèmes interactifs humanoïdes. En jouant le rôle de la victime à travers un personnage virtuel et en développant ainsi de l'empathie pour les victimes d'harcèlement scolaire, les harceleurs changent leurs comportements. Une audience virtuelle permet de s'entraîner à la prise de parole en public<sup>237</sup>. L'audience virtuelle s'adapte au comportement verbal et non-verbal de l'orateur. Un comportement inadapté de l'orateur engendrera ainsi une audience virtuelle simulant l'ennui à travers un ensemble de postures. À l'inverse, un comportement engageant de l'orateur sera reflété par le comportement simulé de l'audience virtuelle. Dans le projet VICTEAMS<sup>238</sup>, un environnement de réalité virtuelle peuplé d'ACAs est développé afin de former les leaders d'équipes médicales dans des situations de crise.

Dans nos travaux, nous avons développé un environnement de réalité virtuelle pour permettre aux médecins de s'entraîner à l'annonce

---

<sup>231</sup> Bruijnes, et al., 2019.

<sup>232</sup> Albrecht, 2006.

<sup>233</sup> Aylett, 2007.

<sup>234</sup> Anderson et al., 2013.

<sup>235</sup> Hall et al., 2011.

<sup>236</sup> Aylett, 2007.

<sup>237</sup> Chollet et al., 2018.

<sup>238</sup> Huguet et al., 2016.

d'événements indésirables graves, cela avec un patient virtuel<sup>239</sup>. La façon dont les médecins livrent des mauvaises nouvelles à un patient a en effet un impact significatif sur le processus thérapeutique (*e.g.* évolution de la maladie, effets secondaires des traitements, litiges). Cependant, les cliniciens expérimentés comme les étudiants en médecine restent très peu formés en la matière (coût, manque de temps) — une formation pourtant fortement recommandée par la Haute Autorité de la Santé (HAS). Pour répondre à cette demande, nous avons développé une plateforme dans laquelle, à travers un casque de réalité virtuelle, le médecin peut dialoguer en langage naturel avec une patiente virtuelle humanoïde autonome. Un système de reconnaissance vocale permet de comprendre en temps réel le discours du médecin. Le moteur de dialogue traite automatiquement le message et identifie la réponse de la patiente virtuelle ; une réponse exprimée à la fois verbalement par une synthèse vocale expressive mais aussi à travers un ensemble de comportements non-verbaux traduisant son état émotionnel (*e.g.* expressions faciales d'émotions, gestes de douleurs).

Les travaux de recherche en informatique prosociale impliquant des systèmes interactifs humanoïdes demeurent peu nombreux. Chaque domaine d'application nécessite plusieurs années de recherche pour modéliser les compétences sociales et les environnements de simulation correspondants. Aujourd'hui, de nombreux domaines d'applications restent à explorer.

### **III. Stéréotypes et discriminations du monde réel au monde virtuel**

Les systèmes interactifs humanoïdes et, plus globalement, les algorithmes en Intelligence Artificielle sont créés et optimisés pour résoudre des tâches particulières — par exemple la reconnaissance automatique de la parole, la détection d'objets ou la classification de documents. Cependant, les comportements de ces algorithmes peuvent induire des effets sociétaux inattendus à la fois positifs et négatifs<sup>240</sup> ; des conséquences non anticipées par leurs créateurs.

La question de l'impartialité des algorithmes d'intelligence artificielle (« *Fairness AI* ») et, plus généralement, de leurs effets sur la société doit être au centre des préoccupations avant l'intégration de ces outils dans notre quotidien. Ces derniers peuvent, de manière insidieuse, renforcer des stéréotypes et des discriminations au sein de notre société. Un exemple

---

<sup>239</sup> Ochs et al., 2019.

<sup>240</sup> Rahwan et al., 2019.

frappant que nous proposons de développer en ces lignes est celui des stéréotypes et discriminations liés au genre.

Les assistants vocaux, des systèmes interactifs, comme l'assistant Siri d'Apple ou Alexa d'Amazon, ont pris une place grandissante dans notre quotidien. Les utilisateurs interagissent en langage naturel avec ces assistants à travers un large éventail de requêtes possibles (recherche sur internet, changer de chanson etc.). Ces assistants vocaux sont souvent non incarnés (dénoués de corps virtuel ou physique). Force est de constater qu'ils sont genrés essentiellement avec des voix de jeunes femmes. Ce choix est motivé par des préférences sociétales selon lesquelles les voix de femmes seraient plus agréables que celles des hommes. Les résultats des recherches menées sur ce sujet sont pourtant loin d'être tranchés : les préférences de voix dépendent de la tâche et du genre de l'utilisateur<sup>241</sup>. Il n'en reste pas moins que les voix de femmes sont perçues comme plus aidantes et coopératives que les voix d'hommes, plus autoritaires<sup>242</sup>. Il semble en outre que les voix de femmes soient perçues comme moins intelligentes et amènent à plus de tolérance en cas d'erreur de reconnaissance vocale<sup>243</sup>. Elles apparaissent donc effectivement plus adaptées au rôle d'assistant vocal dans lequel la reconnaissance de la parole peut échouer régulièrement.

Ces assistantes vocales reflètent, renforcent et perpétuent les stéréotypes sociaux, retracés dans des produits issus des nouvelles technologies. Ainsi, comme le soulignent des chercheurs de l'Université d'Harvard travaillant sur les biais inconscients<sup>244</sup>, plus la société reflètera une association entre femme et assistante, même dans un monde virtuel, plus les femmes, dans le monde réel, seront associées à des rôles d'assistantes et pénalisées si elles n'adoptent pas ce rôle. Le biais genré des systèmes interactifs peut donc non seulement perpétuer les stéréotypes discriminants mais aussi les renforcer et les étendre. À l'inverse, ces systèmes interactifs humanoïdes genrés peuvent aussi être utilisés afin de briser des stéréotypes, par exemple en utilisant des assistantes personnelles féminines pour accomplir des tâches difficiles.

Outre la persistance de stéréotypes, une autre problématique est le comportement dialogique des assistantes virtuelles personnelles face au harcèlement sexuel des utilisateurs, représentant plus de 5 % des interactions<sup>245</sup>. La plupart d'entre elles répondent de manière évasive ou positive, montrant une grande tolérance à ce genre de comportement. Comme le souligne les auteurs du rapport de l'Unesco pour promouvoir la parité dans

---

<sup>241</sup> Mitchell et al., 2011 ; Stromberg, 2013.

<sup>242</sup> Nass et Brave, 2005.

<sup>243</sup> Nass, 2010.

<sup>244</sup> Lai et Mahzarin, 2018.

<sup>245</sup> Coren, 2016.

le numérique<sup>246</sup>, les comportements de ces assistantes virtuelles renforcent les stéréotypes d'assistantes serviles et l'idée selon laquelle ces réponses ambiguës, voir positives, seraient appropriées aux harcèlements sexuels. Ces éléments sont d'autant plus problématiques que les assistantes virtuelles vocales ou incarnées (*i.e.* dotées d'une représentation physique) sont de plus en plus réalisistes en termes de voix et d'apparence, donnant de plus en plus l'impression aux utilisateurs qu'ils interagissent avec une humaine, mais stéréotypée, docile, offrant des réponses simplistes, et parfois incapable de réaliser des tâches très simples<sup>247</sup>. Un certain nombre de recherches montrent, de plus, que la représentation sexualisée des systèmes interactifs induit des comportements sexistes des utilisateurs qui perdurent dans le monde réel. Les conséquences d'une interaction avec un personnage virtuel adoptant un comportement de soumission (par exemple à travers le regard) vont jusqu'à un renforcement du mythe de l'acceptation du viol (mythe selon lequel beaucoup de femmes auraient le désir inconscient d'être violées — *Rape Mythe Acceptance*, RMA)<sup>248</sup>.

Les stéréotypes modélisés à travers les systèmes interactifs, quel que soit leur type, engendrent donc des comportements qui dépassent la sphère du virtuel, en véhiculant une image dégradante et nuisible de la femme. Pour répondre à ces problématiques, un certain nombre d'entreprises développant de tels systèmes ont ajouté des voix masculines en option, ou supprimé la voix féminine par défaut, ou proposé des voix non-genrées, ou encore opté pour une incarnation non-humanoïde (*e.g.* des animaux) ; et elles ont supprimé les comportements d'excuse et de flirt en réponse au harcèlement sexuel des utilisateurs.

Le choix des assistantes vocales peut s'expliquer par la prédominance masculine dans les équipes de conception<sup>249</sup>. Outre la parité dans les équipes de conception, des algorithmes permettant d'implémenter des valeurs morales doivent être intégrés dans les assistants pour éviter des dérives telles que celles rencontrées lors de la mise en service du chatbot de Microsoft qui, après une journée d'apprentissage sur Twitter, a associé l'égalité des genres au féminisme et le féminisme à un cancer<sup>250</sup>. Dans cette perspective, Amazon, dans une version pour enfant d'assistante vocale, impose que les requêtes soient exécutées uniquement si elles ont été posées avec politesse.

Les femmes représentent aujourd'hui 12 % des chercheurs travaillant sur les algorithmes d'apprentissage automatique<sup>251</sup>. L'absence de parité dans ce

---

<sup>246</sup> West et al., 2019.

<sup>247</sup> Clark, 2018.

<sup>248</sup> Fox et al., 2009.

<sup>249</sup> West et al., 2019, Campolo et al., 2017.

<sup>250</sup> Caliskan, 2017.

<sup>251</sup> Mantha et al., 2018 ; West et al., 2019.

domaine est un réel problème étant donnée la place que prennent les algorithmes d'intelligence artificielle dans notre quotidien. En particulier, la plupart de ces algorithmes sont fondés sur des masses de données qui contiennent des stéréotypes liés au genre. Entraînés sur ces données, les algorithmes répliquent ces stéréotypes<sup>252</sup>. Par exemple, l'algorithme d'Amazon permettant d'attribuer une note automatiquement entre 0 et 5 aux CV des candidats discrimine les CV contenant le mot « femme » (comme « président du club d'échecs féminin ») car l'algorithme a été entraîné sur des données de candidats principalement masculins<sup>253</sup>. D'autres travaux sur les algorithmes d'analyse automatique des textes du site d'actualités de Google montrent que ces algorithmes construisent un modèle sexiste concernant les carrières des individus, associant les hommes à des métiers de programmeurs et les femmes à des activités de femmes au foyer<sup>254</sup>. De la même manière, le module de reconnaissance vocale de Google, principalement entraîné sur des voix masculines, est nettement plus performant pour reconnaître une voix d'homme que pour reconnaître une voix de femme<sup>255</sup>.

Des algorithmes d'intelligence artificielle ne véhiculant pas de stéréotypes ou de discriminations de genre sont possibles. Des réflexions sont aujourd'hui menées pour veiller, à travers un ensemble de dispositifs d'évaluation de ces algorithmes et à travers des équipes paritaires de développement d'algorithmes d'apprentissage automatique, à éviter les biais sexistes de ces derniers<sup>256</sup>.

## Conclusion

La dimension humaine de l'intelligence artificielle est multifacette. Tout d'abord, un certain nombre d'algorithmes d'intelligence artificielle et de systèmes interactifs sont conçus en s'inspirant de l'humain et en se nourrissant de données humaines. L'humain façonne la machine et, en retour, la machine façonne l'humain<sup>257</sup>. Ainsi, dans une vision optimiste de l'intelligence artificielle, sont apparus de nouveaux courants de recherche, comme l'informatique prosociale, visant à utiliser le potentiel de l'intelligence artificielle pour améliorer les comportements humains. Diverses applications

---

<sup>252</sup> West et al., 2019 ; Bernheim et al., 2019.

<sup>253</sup> Dastin, 2018.

<sup>254</sup> Bolukbasi, 2016.

<sup>255</sup> Tatman, 2016.

<sup>256</sup> West et al., 2019 ; Bernheim et al., 2019.

<sup>257</sup> Rahwan et al., 2019.

sont développées dans les laboratoires français et internationaux afin d'aider les individus et améliorer le bien-être sociétal.

Dans une vision plus pessimiste de l'intelligence artificielle, il n'en reste pas moins que les effets sociétaux de l'utilisation de certains algorithmes ou de systèmes interactifs peuvent être complètement inattendus et non souhaitables. Le comportement de la machine doit être étudié au-delà de la tâche pour laquelle elle a été conçue afin d'évaluer les implications sociétales de l'introduction de l'intelligence artificielle dans notre société. L'étude scientifique et interdisciplinaire du comportement des machines est essentielle pour que l'intelligence artificielle demeure bénéfique à l'humain et à la société<sup>258</sup>.

## Références

- Albaina, I. M., Visser, T., van der Mast, C. A., and Vastenburg, M. H. (2009). *Flowie: A persuasive virtual coach to motivate elderly individuals to walk*. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1-7. IEEE.
- Albrecht, K. (2006). *Social intelligence : The new science of success*. John Wiley & Sons.
- Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssadou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., et al. (2013). *The tardis framework : intelligent virtual agents for social coaching in job interviews*. In *Advances in computer entertainment*, pages 476-491. Springer.
- Aylett, R., Vala, M., Sequeira, P., & Paiva, A. (2007). *Fearnot!—an emergent narrative approach to virtual dramas for anti-bullying education*. In *International Conference on Virtual Storytelling* (pp. 202-205). Springer, Berlin, Heidelberg.
- Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016). *Openface: an open source facial behavior analysis toolkit*. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-10). IEEE.
- Bernheim, A et Vincent F. , *L'Intelligence artificielle, pas sans elles !*, Belin, collection « Egale à égal » du Laboratoire de l'égalité, 2019.
- Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V. and Kalai, A. 2016. *Man is to computer programmer as woman is to homemaker ? Debiasing word*

---

<sup>258</sup> Rahwan et al., 2019.

*embeddings*. Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 4356–6.

Bruijnes, M., Linssen, J., & Heylen, D. (2019). Special issue editorial: *Virtual Agents for Social Skills Training*. Journal on Multimodal User Interfaces, Volume 13, Issue 1.

Burnham, T. C., & Hare, B. (2007). *Engineering human cooperation*. Human nature, 18(2), 88-108.

Caliskan, A., Bryson, J., and Narayanan, A. 2017. *Semantics derived automatically from language corpora contain human-like biases*. Science, Vol. 365, No. 6334, pp. 183–6.

Campolo, A. et al. 2017. *AI Now 2017 Report*. New York, AI Now Institute, New York University.

Cassell, J. (2000). *More than just another pretty face: Embodied conversational interface agents*. Communications of the ACM, 43:70-78.

Chollet, M., Ghate, P., Neubauer, C., & Scherer, S. (2018, November). *Influence of Individual Differences when Training Public Speaking with Virtual Audiences*. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (pp. 1-7). ACM.

Clark, P. 2018. *The digital future is female – but not in a good way*. Financial Times, 17 June 2018.

Coren, M. J. 2016. *Virtual assistants spend much of their time fending off sexual harassment*. Quartz, Oct, 2016.

Dastin, J. 2018. *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters, 9 October 2018.

Fox, J., & Bailenson, J. N. (2009). *Virtual virgins and vamps: The effects of exposure to female characters' sexualized appearance and gaze in an immersive virtual environment*. Sex roles, 61(3-4), 147-157.

Gebhard, P. (2005, July). *ALMA : a layered model of affect*. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems* (pp. 29-36). ACM.

Hall, Lynne, et al. *Fostering empathic behaviour in children and young people: interaction with intelligent characters embodying culturally specific behaviour in virtual world simulations*. INTED2011 Proceedings (2011) : 2804-2814.

Hempel, J. 2018. *Fei-Fei Li's quest to make AI better for humanity*. Wired, 13 November 2018.

Hoffmann, L., Krämer, N. C., Lam-Chi, A., and Kopp, S. (2009). *Media equation revisited: do users show polite reactions towards an embodied*

agent? In *International Workshop on Intelligent Virtual Agents*, pages 159-165. Springer.

Hoque, M. E., Courgeon, M., Martin, J.-C., Mutlu, B., and Picard, R. W. (2013). *Mach : My automated conversation coach*. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 697-706. ACM.

L. Huguet, D. Lourdeaux, N. Sabouret, M.-H. Ferrer. *Perturbed Communication in a Virtual Environment to Train Medical Team Leaders*. Stud Health Technol Inform. 220:146-9. 2016.

Kihlstrom, J. F. and Cantor, N. (2000). *Social intelligence. Handbook of intelligence*, 2:359-379.

Kopp, S., Gesellensetter, L., Kramer, N. C., and Wachsmuth, I. (2005). *A conversational agent as museum guide: design and evaluation of a real-world application*. In *International Workshop on Intelligent Virtual Agents*, pages 329-343. Springer.

Krämer, N. C. (2008). *Social effects of virtual assistants. A review of empirical results with regard to communication*. In *Proceedings of the international conference on Intelligent Virtual Agents (IVA)*, pages 507-508, Berlin, Heidelberg. Springer-Verlag.

Lai, C. and Mahzarin, B. 2018. *The Psychology of Implicit Bias and the Prospect of Change*. 31 January 2018. Cambridge, Mass., Harvard University.

Mantha, Y. and Hudson, S. 2018. *Estimating the gender ratio of AI researchers around the world*. Medium, 17 August 2018.

Mayer, J. D., Salovey, P., and Caruso, D. R. (2004). *Target articles: emotional intelligence: Theory, findings, and implications*. Psychological inquiry, 15(3):197-215.

Mitchell W. et al. 2011. *Does social desirability bias favour humans ? Explicit-implicit evaluations of synthesized speech support a new HCI model of impression management*. Computers in Human Behavior, Vol. 27, No. 1. pages 402–12.

Nass, C. and Moon, Y. (2000). *Machines and mindlessness: Social responses to computers*. Journal of social issues, 56(1):81-103.

Nass, C. and Brave, S. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, Mass., MIT Press.

Nass, C. I., & Yen, C. (2010). *The man who lied to his laptop: What machines teach us about human relationships*. New York : Current.

Ochs, M., Pelachaud, C., & Sadek, D. (2008, May). *An empathic virtual dialog agent to improve human-machine interaction*. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent*

*systems-Volume 1* (pp. 89-96). International Foundation for Autonomous Agents and Multiagent Systems.

Ochs, M., Sabouret, N., & Corruble, V. (2009). *Simulation of the dynamics of nonplayer characters' emotions and social relations in games*. IEEE Transactions on Computational Intelligence and AI in Games, 1(4), 281-297.

Ochs, M., Mestre, D., De Montcheuil, G., Pergandi, J. M., Saubesty, J., Lombardo, E., ... & Blache, P. (2019). *Training doctors' social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence*. Journal on Multimodal User Interfaces, 13(1), 41-51.

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). *Empathy in virtual agents and robots: A survey*. ACM Transactions on Interactive Intelligent Systems (TiiS), 7(3), 11.

Paiva, A., Santos, F. P., & Santos, F. C. (2018, April). *Engineering prosociality with autonomous agents*. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Pentland, A. (2005). *Socially aware, computation and communication*. Computer, 38(3):33-40.

Picard, R. (1997). *Affective computing*, volume 252. MIT press Cambridge.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., & Jennings, N. R. (2019). *Machine behaviour*. Nature, 568(7753), 477.

Sinclair, J. (2004). *Corpus creation. Corpus linguistics : Readings in a widening discipline*, 78-84.

Stromberg, J. 2013. *Why women like deep voices and men prefer higher ones*. Smithsonian Magazine, 24 April 2013.

Tatman, R. 2016. *Google's speech recognition has a gender bias*. Making Noise and Hearing Things, 12 July 2016.

Vinciarelli, A. and Pentland, A. S. (2015). *New social signals in a new interaction world: the next frontier for social signal processing*. IEEE Systems, Man, and Cybernetics Magazine, 1(2):10-17.

West, M., Kraut, R., & Ei Chew, H. (2019). *I'd blush if I could: closing gender divides in digital skills through education*.



# L'intelligence artificielle

*Dimension juridique*



# **Les intelligences artificielles du droit**

**Boris Barraud**

Le droit ne saurait échapper aux mutations induites par l'intelligence artificielle et décrites dans ce livre, qu'il s'agisse du droit compris comme ensemble de normes ou lois positives ou du droit compris comme monde des juristes. Dans toutes ses dimensions, le droit est perturbé et renouvelé par les nouvelles technologies informatiques et, en particulier, par l'utilisation d'algorithmes permettant, à partir de données (les entrées), d'aboutir à des résultats pertinents (les sorties) en suivant différentes étapes qui requièrent des calculs, des opérations logiques, des comparaisons statistiques ou des analogies. Progressivement mais rapidement, l'intelligence artificielle envahit le quotidien des juristes, leur posant des problématiques inédites et largement porteuses d'avenir et leur proposant des solutions originales et efficaces.

Les relations entre intelligence artificielle et droit peuvent être étudiées sous deux angles radicalement différents. Tout d'abord, le droit saisit la technologie. Se pose donc la question du droit des algorithmes, du droit des services automatiques ou encore du droit des robots ; ou bien encore, pour prendre des exemples plus précis, la question du régime juridique des voitures autonomes ou des moteurs de recherche en ligne. Il s'agit là de problématiques de droit privé, s'attachant notamment aux enjeux de responsabilités en cas de préjudice causé par une IA ou à cause d'une IA. Il faut alors répondre à des interrogations de pur droit positif mais aussi et peut-être surtout à des hésitations éthiques et morales. D'un autre point de vue, la technologie saisit le droit. L'intelligence artificielle, de différentes façons, impacte le monde juridique. Cela soulève des problématiques relevant de la philosophie du droit et du droit public fondamental, mais aussi de la science administrative ou encore de la science politique.

Les technologies numériques sont l'objet d'un droit et, en même temps, le droit est l'objet de ces technologies. Seul ce second aspect sera traité en ces pages. Le droit de l'IA sera délaissé afin de se concentrer sur l'IA du droit, sur l'impact qualitatif et quantitatif de l'intelligence artificielle sur le milieu juridique. De ce point de vue, les algorithmes et les réseaux de neurones sont à la fois une nouvelle source du droit (*I*) et un nouvel outil d'application du droit (*II*). Ils font le droit en produisant une normativité très particulière ; et ils sont au service du droit dès lors qu'ils sont créés et utilisés par des *legaltechs* et des *legal start-up*. Mais ils ne profitent pas nécessairement à la qualité du droit, à la qualité du travail des juristes, des juges ou des législateurs. C'est pourquoi il importe de questionner les limites des algorithmes juridiques et même les dangers qui risquent de les accompagner.

## I. Des intelligences artificielles à l'origine du droit

La normativité jaillissant de l'intelligence artificielle étonne et détonne dans le monde du droit pour plusieurs raisons. Tout d'abord, elle accompagne des mouvements de personnalisation et de factualisation qui ne peuvent s'analyser autrement qu'en tant que régressions normatives. Les nouvelles technologies révolutionnent les modes de création du droit. Et cela autant en matière de normativités publiques qu'en matière de normativités privées. Les nouvelles technologies numériques induisent une transformation majeure des modes de régulation des conduites individuelles et sociales. La normativité étatique et sa nature délibérée dans le cadre d'un Parlement est mise en ballotage par la normativité algorithmique, par des normes issues de traitements automatisés de données. Or il y a dans la délibération une humanité qui est peut-être indispensable à la création des normes encadrant les comportements et les relations sociales. Celle-ci assure le débat et la confrontation des arguments, permet d'obtenir des compromis politiques et est censée défendre l'intérêt général.

L'algorithme « auto-apprenant » par rapprochements successifs, en dégageant des corrélations, fait émerger des normes. Il produit donc des règles à partir des régularités et des coïncidences qu'il identifie dans un corpus de données. C'est pourquoi le droit tend à se résumer à un ensemble de données : l'heure est au « droit en datas » et à la « loi des algorithmes ». Cette dernière ne partage que peu de choses — *i.e.* seulement son effet normatif — avec la loi au sens strict, c'est-à-dire la loi produite par les parlementaires en suivant une procédure constitutionnelle. Notamment, la loi parlementaire a une portée générale et impersonnelle qui garantit l'absence de toute discrimination ; et, par souci d'impartialité, elle ne prend que très peu en compte les situations

individuelles. Il n'existe en revanche guère d'égalité devant la « loi des algorithmes ». Celle-ci, au contraire, va de pair avec un mouvement d'individualisation des règles de droit.

La Révolution française de 1789 a affirmé le légalisme, imposant aux citoyens des normes générales et abstraites pour qu'ils ne soient plus enfermés dans des dépendances personnelles comme sous l'Ancien Régime. Mais l'hyper-personnalisation par l'intelligence artificielle en vient à recréer de telles dépendances. La « dictature des algorithmes » donne fatallement lieu à des discriminations. L'objet même de ces outils technologiques est de traiter différemment chaque utilisateur, par exemple en lui proposant des contenus personnalisés. L'effet performatif de ce qui peut apparaître comme une simple suggestion fait le reste.

L'automaticité de cette nouvelle forme de droit est donc synonyme de personnalisation en plus d'être synonyme de factualisation. Or, en France, l'article 225-1 du Code pénal incrimine toute forme de discrimination fondée notamment sur l'origine, le sexe, la situation de famille, l'apparence physique, le patronyme ou le lieu de résidence. Mais les algorithmes ne cessent de discriminer sur la base de critères illicites.

C'est avant tout en matière d'activités sur les plateformes numériques et sur le web que des intelligences artificielles interviennent. Or, si les données personnelles sont le nouveau pétrole, leur exploitation par des algorithmes non désintéressés idéologiquement peut entraîner de nouvelles formes d'esclavagisme ou, du moins, de téléguidage des comportements collectifs et individuels. La régulation technologique malmène le libre arbitre. La force de la normativité de l'IA repose sur les résultats individualisés qui sont produits. Les prophéties auto-réalisatrices des algorithmes sont telles que, de fait, la technologie choisit pour l'humain ; et l'humain s'en remet assez aveuglément à elle.

De plus en plus, les conduites sont orientées, les esprits formatés, les désirs définis et les besoins standardisés par des intelligences artificielles. Et ces dernières tendent à favoriser le suivisme, le panurgisme, et à limiter au maximum les envies et les modes de vie alternatifs, car ceux-ci sont trop peu monétisables. On se retrouve ainsi positionné dans un silo de comportement et de consommation. La normalité devient normativité.

Loin de l'ouverture que le web était supposé offrir à chacun, il enferme, il réduit en données exploitables commercialement, il limite la liberté d'opinion et diminue le pluralisme des courants d'idées. On peut par conséquent s'inquiéter de la capacité de la société numérique à être une société démocratique et ouverte plutôt qu'une société tyrannique et fermée. La démocratie, partout dans le monde, affronte de graves crises. Le web et l'intelligence artificielle sont au cœur de ces crises, les entretenant peut-être plus qu'ils contribuent à les éteindre.

Le gouvernement par l'intelligence artificielle — derrière laquelle se cachent des intelligences humaines non désintéressées idéologiquement — a pour finalité d'anticiper les comportements et de les orienter dans une direction prédéterminée, comme toute forme de gouvernement. Ainsi les algorithmes redessinent-ils le visage du monde, affirment-ils des légitimités nouvelles, décident-ils des succès et des échecs, y compris au moment de scrutins démocratiques visant à élire un président ou à décider de sortir ou non d'une organisation économique et politique interétatique. Ils déterminent ce qui est « bien » et ce qui est « mal ». Ils écrivent le projet de société de demain.

La puissance des réseaux sociaux et des grandes plateformes de partage de contenus est devenue immense, en même temps que celle des quelques hommes qui les font fonctionner en élaborant leurs algorithmes. Cela soulève de nombreuses interrogations au regard de la démocratie et des libertés fondamentales. Car il s'agit bien de puissance : une puissance concurrente à la puissance étatique, tandis qu'une autre souveraineté, une souveraineté technologique et économique, s'élève face à la souveraineté politique. On imagine déjà que les jours de l'État, du droit et de la justice modernes seraient comptés, qu'ils ne seraient plus que des pis-allers historiques voués à être progressivement remplacés par des modes de régulation scientifiques, mathématiques. Cette évolution est peut-être réelle, mais elle n'est pas irrésistible ni irréversible. Mais encore faut-il en prendre conscience pour pouvoir proposer et mettre en œuvre quelques solutions afin de préserver « le pouvoir du peuple, par le peuple et pour le peuple » et lui éviter de laisser la place à un « pouvoir du peuple, par quelques puissances privées et pour quelques puissances privées ».

Toujours est-il que l'intelligence artificielle est une nouvelle source de droit dès lors qu'on définit le droit tel ce qui régule effectivement les conduites des hommes en société. Cette intelligence artificielle engendre des effets normatifs tout aussi forts que ceux de beaucoup de lois et règlements publics. Cela incite les juristes à reconstruire leurs objets d'étude et leurs activités loin du droit moderne, là où « droit » était synonyme d'« État » à la fois en théorie et en pratique.

La normativité algorithmique illustre la tendance de la normativité descendante, provenant d'une autorité en surplomb, à être débordée par une normativité immanente venant de la technologie. La normativité publique, faite de normes générales, abstraites et impersonnelles, est concurrencée par une normativité paradoxalement à la fois individualisée et à prétention universelle, une normativité rendue possible par la collecte des données personnelles et leur analyse par des IA, autour desquelles s'ordonne le monde actuel. C'est ainsi que les règles de droit publiques, transparentes et conscientes, sont remplacées progressivement par des normes privées, opaques et inconscientes.

L'automaticité du traitement des données et donc de la « loi des algorithmes » témoigne de grands changements dans la manière de créer et d'appliquer les règles de droit. Elle pourrait même indiquer une tendance du droit à se passer de l'humain, à se robotiser. Et ce n'est pas qu'en tant que source inédite de normativité que l'intelligence artificielle redessine le paysage juridique ; elle le fait également en donnant aux juristes, ainsi qu'aux non-juristes, des outils nouveaux pour comprendre, respecter et appliquer les normes.

## II. Des intelligences artificielles au service du droit

L'intelligence artificielle se met au service du droit de différentes façons et, en premier lieu, avec les *legaltechs* et *legal start-up* qui proposent des services juridiques originaux, cela dans un cadre privé. Elle se met également au service du droit à l'initiative des institutions publiques qui entendent mettre à profit les nouvelles opportunités technologiques afin de réduire les coûts tout en maintenant ou améliorant la qualité de l'administration et peut-être même de la justice. On pourrait également évoquer les « *smart contracts* », nouvelles modalités d'exécution automatique des contrats s'appuyant sur la technologie des *blockchains*, qui constituent un autre exemple d'apport au droit de l'intelligence artificielle. Toutes ces initiatives redessinent le visage du droit, remodèlent ses rouages et son fonctionnement et obligent ses acteurs à s'adapter.

Des *legal start-up*, saisissant l'opportunité de nouveaux marchés juridiques encore peu exploités, élaborent des technologies remplaçant les services humains par des services automatiques. Cela permet de réduire les coûts et, par suite, les prix. Les modèles économiques s'avèrent cependant encore hésitants. Mais la demande est réelle et les *legaltechs* se développent — plus ou moins rapidement et massivement selon les pays et les cultures juridiques et technologiques, les États-Unis étant les plus avancés.

Ces nouveaux acteurs du monde juridique que sont les *legaltechs* provoquent des changements nombreux et profonds. Ils s'adressent déjà aux particuliers avec des services dématérialisés et automatisés, notamment de conseil ou de production de documents, à des prix très compétitifs mais à la pertinence et à la précision plus ou moins limitées. Ainsi une problématique importante apparaît-elle : celle du développement d'un secteur (immense) du discount du droit. Il reste que ces services favorisent l'accès au droit et à la justice de toute une partie de la population qui, dans tous les cas, ne consulterait pas un juriste professionnel.

Les entreprises sont d'autres cibles des *legal start-up*. Certaines souscrivent des abonnements à des prix fort élevés pour accéder à des technologies plus élaborées. Des *legal start-up* permettent de générer des contrats divers et personnalisés à partir de formulaires intelligents et dynamiques. En dialoguant avec un *legal chatbot*, donc avec une intelligence artificielle capable de comprendre les besoins et d'y répondre, on peut créer divers documents ou mener à bien diverses procédures. Immatriculation d'une société en quelques clics, élaboration des statuts, édition de baux au cas par cas, rédaction de contrats, gestion de contentieux etc. De plus en plus d'aspects des vies juridiques des personnes physiques et morales sont concernés.

L'intelligence artificielle est également utilisée par les juristes, spécialement les avocats. Ceux-ci profitent des gains et économies que la technologie permet, en premier lieu lorsqu'elle prend la place des collaborateurs humains. Certaines activités sont aujourd'hui confiées à des collaborateurs électroniques et non plus à des collaborateurs humains. Plus encore, certaines activités irréalisables par des hommes ou réalisables seulement dans de mauvaises conditions et/ou à un coût très élevé par des hommes peuvent être effectuées facilement et rapidement par des IA, qu'il s'agisse de recherche d'informations, de tri, de classement, de comparaison de cas semblables au sein d'immenses bases de données ou encore de mise en relation de parties adverses à des fins de conciliation et de transaction.

Bénéficier de statistiques fiables quant à l'issue des litiges s'ils venaient à être présentés devant la justice (la « justice prédictive ») est peut-être le premier apport des *legaltechs*. Les informations contingentes obtenues, relatives au seul cas concerné, sont pour cette raison très pertinentes et donc utiles. Elles permettent de déterminer s'il serait judicieux ou vain de porter le litige devant les tribunaux ou bien s'il serait préférable de rechercher un accord amiable. Par exemple, lorsqu'un avocat est informé par l'algorithme que son client n'a que 5 % de chances de remporter un éventuel procès dans une affaire donnée, il va nécessairement l'inciter à tout mettre en œuvre afin d'éviter ce procès, quitte à accepter une transaction plutôt défavorable.

De plus en plus, les cabinets d'avocats s'appuyant sur de telles intelligences artificielles profitent de grands avantages concurrentiels par rapport à ceux qui font le choix de ne pas y recourir. Ces technologies sont la source de gains de productivité et d'une forte valeur ajoutée.

Toutefois, la révolution numérique n'est pas uniquement une opportunité pour les professions juridiques. Les *legaltechs* transforment profondément et durablement les marchés des services juridiques. Ces professions sont dans l'obligation de repenser leurs activités à l'image des notaires avec les *blockchains* ou des avocats avec la justice prédictive. Se pose par exemple la question d'une possible « ubérisation » des métiers de la justice, le cadre

public réglementé laissant la place à un cadre privé autorégulé. Les tribunaux engorgés ne peuvent qu'y voir un bienfait, mais il n'en va pas de même des professionnels du conseil juridique et de la défense en justice.

Le phénomène de l'autojuridication interroge de la même manière dès lors que de plus en plus de personnes (physiques ou morales) se contentent des informations trouvées en ligne et des documents produits automatiquement en quelques clics. Dès lors, elles préfèrent ne pas s'attacher les services d'un juriste professionnel. Sitôt que, par exemple, une intelligence artificielle donne « clés en mains », pour un prix faible ou nul, les arguments à mettre en avant devant un juge, pourquoi rémunérer un juriste de métier, si ce n'est parce qu'il s'agit d'une obligation légale — dont la légitimité pourrait alors être remise en question — ?

En outre, se développent les plateformes de résolution amiable des litiges. La pratique, appelée *Online Dispute Resolution*, est très à la mode aux États-Unis et au Canada. Or, tandis qu'on craint que la résolution amiable soit abandonnée à des algorithmes, le projet de loi de programmation 2018-2022 et de réforme pour la justice encourage le recours à des médiateurs et conciliateurs algorithmiques.

Ces diverses évolutions dans les pratiques et les marchés juridiques conduisent les avocats et la plupart des professions du droit à se questionner au sujet de l'évolution de la demande et des nouveaux moyens d'y répondre, afin d'être acteurs du changement plutôt que de le subir.

Reste que les *legaltechs* et les *legal start-up* participent au phénomène de factualisation du droit. Produisant des statistiques quant aux solutions retenues face à des éléments juridiques et factuels donnés, elles ne séparent pas les uns des autres et les abordent tous tels des faits. Autrement dit, les règles de droit ne sont jugées ni plus ni moins importantes que les situations particulières de chaque situation. Du point de vue des *legaltechs*, les lois, règlements et autres jurisprudences sont de purs faits.

Par ailleurs, l'intelligence artificielle permet à l'administration et aux administrés de réaliser des économies grâce à l'automatisation de certains services, certaines procédures et certaines démarches. L'algorithmitation et l'automatisation de l'État et de l'administration prennent toute leur place dans la modernisation, la rationalisation et la dématérialisation des services publics. Ces phénomènes constituent une rupture à la fois qualitative et quantitative au sein de la modernisation technologique des institutions. L'intelligence artificielle va jusqu'à subroger des décisions informatiques à des décisions humaines ou, du moins, mettre la décision informatique au service de la décision humaine. Les algorithmes deviennent progressivement — et assez insidieusement — les nouveaux déterminants, cadres et vecteurs des politiques publiques et du droit qui les accompagne.

L'appareil d'État a ainsi tendance à s'automatiser progressivement. Or cette automatisation n'est guère neutre ni apolitique. Une décision prise par un algorithme ou avec le soutien d'un algorithme semble au premier abord plus juste qu'une décision humaine risquant d'être orientée politiquement et en tout cas soumise à une subjectivité. La rigueur mathématique et la logique de l'algorithme jouent pour lui. Néanmoins, il reste un simple outil produit par et pour l'homme. Aux côtés de la technologie, il se trouve toujours des individus qui l'utilisent et l'orientent en fonction d'objectifs précis. Les critères, les paramètres et, surtout, les données qui font qu'un algorithme produit certains résultats plutôt que d'autres dépendent des hommes. L'objectivité, la neutralité et la logique mathématique des algorithmes ne sont que des façades derrière lesquelles se cachent des intentions et des arbitrages humains. Si « *Code is law* », « *Code is politic* ».

Plus largement, l'architecture et l'environnement que l'intelligence artificielle dessine dépendent de visions politiques. Elles changent le monde à des fins bien déterminées. L'automatisation du droit n'est pas une simple évolution technique, une simple adaptation conjoncturelle. Cela explique d'ailleurs pourquoi certains États ont, dans ce domaine, de l'avance et d'autres du retard.

## Conclusion

Le secteur privé a été le premier à exploiter les grands ensembles de données (*big data*) et les nouvelles technologies de gestion de ces données (*data management*). À présent, néanmoins, le secteur public produit lui-aussi ses propres jeux de données et ses propres algorithmes afin d'automatiser certaines de ses activités, qu'il s'agisse d'administration ou, mais avec davantage de précautions, de justice et même de législation. La « loi des algorithmes » régira de plus en plus les activités individuelles et sociales ; et cette loi ne sera plus seulement une loi privée, elle sera également issue des autorités publiques. Ainsi observe-t-on l'affirmation du « droit en datas », ainsi pénètre-t-on dans l'ère du « cyber-droit », un droit ordonné autour de la collecte massive et de l'exploitation automatique de données.

La bonne gestion des affaires d'une société est trop humaine pour que l'on puisse l'abandonner à des robots ; et les robots sont trop intelligents pour que l'on puisse les ignorer. Comme souvent, tout n'est ici qu'une question d'équilibre : équilibre entre technophobie et technophilie, donc technolucidité. Il ne s'agit ni de s'opposer en bloc aux nouvelles opportunités issues de l'intelligence artificielle ni de la laisser décider de tout. Mais il est important que l'administration, la justice et toutes les formes de prise de

décision publique soient les plus ouvertes et transparentes possible, tandis qu'il faudrait que les algorithmes — et donc ceux qui les conçoivent — fassent preuve de loyauté. Les administrés et les justiciables ont besoin de comprendre comment les décisions qui les concernent sont prises, surtout lorsqu'elles le sont par une intelligence artificielle ou à l'aide d'une intelligence artificielle.

La révolution de l'intelligence artificielle s'accompagne de lourdes mutations politiques, économiques et sociales. Elle appelle un nouveau contrat social, de nouveaux liens entre l'État et les citoyens. L'algorithisation est beaucoup plus qu'un projet technique. C'est un projet politique visant à réinventer, dans un contexte de crise de la démocratie représentative, la relation entre les administrés et les administrateurs. C'est pourquoi il semble nécessaire de susciter le plus large débat public autour de ces questions, afin de permettre à chacun de prendre position. Faire vivre en permanence la réflexion permettrait d'anticiper certaines difficultés et d'évoluer progressivement mais avec précautions vers les meilleures solutions, suivant la méthode des petits pas.

Cette révolution de l'intelligence artificielle conduisant notamment à réinventer le droit, les juristes doivent au moins autant que les autres participer aux réflexions et aux discussions.



# L'intelligence artificielle

*Dimension chinoise*



# L'intelligence artificielle en Chine : un état des lieux

**Aifang Ma<sup>259</sup>**

L'intelligence artificielle est devenue l'une des priorités politiques de la Chine à partir de 2013, année où les *Opinion directrices* du Conseil d'État sur la promotion du développement coordonné et sain des objets connectés ont été rendues publiques. Afin de doper le développement de l'IA en Chine et de pouvoir concurrencer les États-Unis au niveau international, la Chine a élaboré des stratégies quantifiables à long terme, a consacré un budget public colossal au secteur de l'IA, et a soutenu l'expansion des champions numériques nationaux. Des progrès importants ont été réalisés en termes d'applications industrielles de l'IA. Aujourd'hui, l'IA est utilisée en Chine dans des secteurs aussi divers que la santé digitale, la protection environnementale, la sécurité publique ou la défense nationale. Qui plus est, la Chine est en train de prendre le dessus sur les États-Unis dans ce domaine. Face aux nombreux doutes et réserves sur les intentions réelles du gouvernement chinois, cet article entend montrer que, pour la Chine, mobiliser l'IA afin de résoudre des problèmes intérieurs est un objectif non moins important que de se disputer le leadership technologique dans le monde. Il sera plus constructif d'identifier des opportunités de coopération, afin d'exploiter le riche potentiel de l'IA au profit du plus grand nombre de personnes.

En Chine, le 11 novembre est une date spéciale, non pas parce qu'elle célèbre l'armistice de la Première guerre mondiale, mais parce qu'elle marque la Fête des célibataires, laquelle gagne en popularité parmi les jeunes. L'importance de cette fête provient d'une pure imagination commerciale,

---

<sup>259</sup> Cet article est basé sur une note publiée par la Fondation pour l'innovation politique en novembre 2018.

selon laquelle la combinaison de quatre chiffres identiques (« 1 ») incarnerait le rassemblement des célibataires.

La création de la Fête des Célibataires sert avant tout les profits des commerçants électroniques chinois tels que JD, Tmall, Taobao et Alibaba. Les consommateurs surfent sur Internet et achètent en grandes quantités pour profiter des remises qui sont particulièrement importantes ce jour-là. Le chiffre d'affaires réalisé en cette seule journée est souvent très impressionnant. Par exemple, le 11 novembre 2018, Alibaba, géant chinois de l'e-commerce, a réalisé un chiffre d'affaires de 30,71 milliards de dollars, soit une augmentation de 27 % par rapport à 2017 (25,3 milliards de dollars), de 72,5 % par rapport à 2016 (17,8 milliards de dollars), et de 127,8 % par rapport à 2015 (13,485 milliards de dollars).

Comment expliquer la victoire triomphante des commerçants électroniques ? La réponse réside dans l'application de l'intelligence artificielle qui a contribué au développement galopant du e-commerce en Chine. Cette technologie de pointe a permis aux commerçants électroniques d'envoyer des publicités personnalisées, de cibler des consommateurs et de prévoir quels types de marchandises peuvent être le mieux vendues auprès de quelles catégories de personnes et dans quelles régions du pays.

Cet exemple montre combien la Chine est devenue l'un des pays les plus avancés dans le domaine de l'intelligence artificielle. Les ingénieurs et les chercheurs chinois se concentrent sur la recherche et le développement d'une panoplie de technologies liées à l'IA : reconnaissance faciale, reconnaissance vocale, drones, voitures autonomes, robots, assistant domestique, objets connectés, réalité augmentée, apprentissage profond etc. Le gouvernement chinois a donné son feu vert au développement de l'IA et lui accorde même une importance stratégique. Quand le Président chinois Xi Jinping présentait ses vœux de nouvelle année en 2018 à la télévision, les deux livres affichés derrière lui traitaient de l'IA : *The Master Algorithm*, de Pedro Domingos, et *Augmented Life in the Smart Lane*, de Brett King. Un geste assez révélateur.

## I. Les politiques chinoises pour le développement de l'intelligence artificielle

Quand la Chine s'est lancée dans le développement de l'IA au début des années 2010, elle s'est fixée au moins deux objectifs. Premièrement, elle envisage de mettre pleinement à profit l'IA pour résoudre des problèmes intérieurs, y compris le processus d'urbanisation accélérée, la réduction de la pauvreté, la pollution environnementale, la raréfaction des ressources

naturelles, la mise à niveau des structures économiques ou encore l'entrée dans une société du vieillissement. La présence de ces défis rend impératif de trouver une « *elegant solution* » permettant de répondre à tous ces problèmes en même temps.

En second lieu, et dans une certaine mesure, accorder une importance à l'IA a pour objectif de faire face à la concurrence américaine. En mai 2016, la Maison Blanche a créé un sous-comité spécifique au sein du National Science and Technology Council (NSTC), organisme chargé de suivre de près l'évolution du secteur de l'IA et de coordonner les activités fédérales concernées. Entre mai et juillet 2016, quatre sessions de travail publiques ont eu lieu, pour engager la discussion avec le public et évaluer les opportunités, les risques et les implications réglementaires et sociales de l'IA. Trois rapports stratégiques ont été rendus publics par l'administration Obama entre octobre et décembre 2016 : *Preparing for the Future of Artificial Intelligence*, *The National Artificial Intelligence Research and Development Strategic Plan* et *Artificial Intelligence, Automation, and the Economy*. De plus, l'embargo que les États-Unis ont déclaré en 2015 sur les ventes à la Chine de circuits intégrés venant des entreprises Intel, AMD et NVIDIA (leaders dans le domaine des microprocesseurs) rend impératif pour le pays de devenir autonome afin de répondre au blocage américain.

Pour atteindre ces deux objectifs, le gouvernement chinois a placé l'IA au cœur de ses priorités, ce qui se traduit par la prise de mesures tous azimuts à partir de 2013 avec la proclamation des *Opinions directrices* du Conseil d'État sur la promotion du développement coordonné et sain des objets connectés (ci-après *Opinions*).

## A. Proclamation de stratégies précises et mesurables

L'intelligence artificielle est une technologie d'avenir et met en jeu les intérêts économiques et la sécurité nationale d'un pays. C'est pour cette raison que la Chine a mis en place des stratégies précises et mesurables. En juillet 2017, le Conseil des affaires d'État a publié le *Plan de développement de la prochaine génération d'intelligence artificielle* pour la période de 2016 à 2030 (ci-après le *Plan*). Dans ce document, la Chine se fixe l'objectif de devenir le leader mondial de l'IA d'ici à 2030. Elle envisage de réaliser son but en trois étapes (voir les encadrés 1, 2 et 3, qui sont des traductions du *Plan* concernant les objectifs précis de chaque étape).

D'ici 2020, l'objectif principal est d'élever le niveau technologique de l'IA et celui des applications de l'IA. L'industrie de l'intelligence artificielle sera devenue une nouvelle dynamique importante au cœur de la croissance

économique. Les applications de l'IA s'imposeront comme un nouveau moyen d'améliorer le bien-être des chinois.

**Encadré 1 : objectifs précis d'ici 2020**

Des progrès importants doivent être réalisés en matière de théories et de technologies de la nouvelle génération d'IA. Les théories fondamentales et les technologies clé concernant Big Data, l'intelligence cross-média, l'intelligence de groupe, l'intelligence hybride renforcée, et l'automation doivent avoir avancé à grands pas. Des résultats importants doivent avoir été obtenus dans les domaines des méthodes de modélisation de l'intelligence artificielle, des périphériques centraux, des équipements haut de gamme et des logiciels de base, etc.

La compétitivité de l'industrie de l'intelligence artificielle chinoise aura figuré au premier rang international. Les normes technologiques, le système de services et la chaîne écologique industrielle de l'intelligence artificielle vont prendre forme. La Chine aura formé plusieurs entreprises phares de l'IA. Le secteur pèsera alors en Chine plus de 150 milliards de yuans (environ 23 milliards de dollars), et les secteurs liés, plus de 1000 milliards de yuans (environ 156 milliards de dollars).

L'environnement de développement de l'intelligence artificielle sera optimisé encore davantage. L'IA sera appliquée à fond dans certains domaines clés. Des équipes d'innovation et des talents de haut niveau en IA vont être mobilisés. Les normes éthiques, les politiques et les réglementations, quoique préliminaires à ce stade-là et qui ne couvrent que partiellement le secteur de l'IA, vont se mettre en place.

D'ici 2025, des percées significatives seront réalisées concernant la théorie fondamentale de l'IA. Le développement de certaines technologies et applications de l'IA atteindra le niveau le plus avancé du monde. L'IA sera le vecteur principal de la mise à niveau industrielle et économique de la Chine. Des progrès positifs seront faits dans l'édification d'une société intelligente.

**Encadré 2 : objectifs précis d'ici 2025**

Quoique préliminaires en 2025, les théories et le système technologique de la nouvelle génération d'IA seront mis en place. Des progrès importants seront obtenus en matière d'apprentissage profond. Des résultats de recherche avant-gardistes dans plusieurs domaines de l'IA seront disponibles

L'industrie chinoise de l'intelligence artificielle sera entrée dans le high-end de la chaîne mondiale de la valeur. La nouvelle génération de l'IA sera largement utilisée dans la production intelligente, la santé digitale, les villes intelligentes, l'agriculture intelligente, et la défense nationale. Le secteur pèsera alors en Chine plus de 400 milliards de yuans (environ 63 milliards de dollars), et les secteurs liés, plus de 5000 milliards de yuans (environ 781 milliards de dollars)

Le système de lois, de réglementations, de normes éthiques et de politiques de l'IA, quoique préliminaire en 2025, se mettra en place, pour être capable d'évaluer, de gérer et de contrôler la sécurité de l'IA.

D'ici 2030, la Chine deviendra le centre mondial de l'innovation en IA et figurera au premier rang mondial concernant le développement des théories, des technologies et des applications. Des progrès tangibles auront été faits dans l'économie intelligente et dans la société intelligente, ce qui contribuera à la croissance économique et au renforcement de la capacité d'innovation du pays.

## *L'intelligence artificielle*

### Encadré 3 : objectifs précis d'ici 2030

Les théories et le système technologique de la nouvelle génération de l'IA seront mis en place et auront atteint une certaine maturité. Des percées significatives auront été obtenus en matière d'intelligence cérébrale, d'automation, d'intelligence hybride, et d'intelligence de groupe. La Chine exercera une influence importante dans la recherche internationale de l'IA et atteindra un niveau technologique très avancé.

La compétitivité de l'industrie chinoise de l'IA aura atteint un très haut niveau à l'échelle internationale. Des avancées extrêmement importantes auront été faites concernant la largeur et la profondeur de l'application de l'IA dans la production, la vie quotidienne, la gouvernance sociale et la défense. Une chaîne industrielle complète et un cluster industriel haut de gamme qui couvrent les technologies essentielles, les systèmes clés, les plateformes de support et les applications intelligentes seront prêts. Le secteur de l'IA pèsera alors plus 1000 milliards de yuans (environ 156 milliards de dollars), et les secteurs liés, plus de 10000 milliards de yuans (environ 1562 milliards de dollars).

La Chine aura établi plusieurs bases de première classe mondiale pour pousser l'innovation technologique de l'IA et la formation de talents. Elle aura mis en place un système plus élaboré de lois, de réglementations, de normes éthiques et de politiques sur l'intelligence artificielle.

En outre, le ministère de l'Industrie et des Technologies informatiques (MIIT) et le ministère des Finances ont publié, le 7 décembre 2016, un document officiel intitulé *Intelligent Manufacturing Plan (2016-2020)* pour promouvoir l'application industrielle de l'IA, notamment dans les dix domaines clés indiqués dans le plan *Made in China 2025*, y compris les nouvelles technologies informatiques, les outils de contrôle numériques, les équipements aérospatiaux, les bateaux high-tech, les équipements ferroviaires, les économies d'énergie, les appareils médicaux, les machines agricoles et les équipements utilisés pour la production d'électricité.

Enfin, le 18 mai 2016, le ministère de la Science et des Technologies, la Commission nationale du développement et de la réforme (NDRC), le MIIT ainsi que le Bureau de régulation de l'Internet et des télécommunications ont publié le *Plan d'action de trois ans pour la promotion de l'Internet Plus et de l'intelligence artificielle (2016-2018)*. Selon ce document officiel, fin 2018 la Chine devait achever l'édification de la structure industrielle du projet *Internet Plus*, à base d'intelligence artificielle. Un système de services d'innovation et de standardisation devait être mis en place. Le niveau technologique et industriel de l'IA chinoise devait alors rattraper la moyenne internationale.

## B. Investissements colossaux et soutien public constant

Le gouvernement chinois considère l'intelligence artificielle comme l'un des secteurs clés qui décideront de la place du pays sur la scène internationale de demain. Des investissements à hauteur de 150 milliards de yuans (environ 23,15 milliards de dollars) ont été prévus pour la période 2017-2020 afin d'aider les universités, les incubateurs et les start-up à développer leur expertise dans l'IA. Cet engagement pourrait même atteindre 400 milliards de

yuans (environ 60 milliards de dollars). Deux autres chiffres contribuent à illustrer l'importance des investissements chinois dans l'IA : en 2017, la progression dans ce secteur a été de 141 % par rapport à 2016, ce qui a permis l'émergence de 1 100 nouvelles start-up.

Par rapport à l'enthousiasme manifeste du gouvernement chinois à l'égard de l'IA, le soutien public d'autres pays semble moins spectaculaire. Bien que les États-Unis puissent s'appuyer sur les performances jusqu'à présent inégalées de son secteur privé dans l'IA, les investissements publics sont un peu timides. En avril 2018, le Pentagone a obtenu un budget de 9,3 milliards de dollars pour le développement des drone militaires, selon un rapport publié par le Centre d'étude du drone du Bard College. Les États-Unis ont alloué en 2019 un montant de 93 millions de dollars au projet *Maven*, dont Google est un partenaire important. Selon le calcul effectué par Bloomberg Government, le gouvernement fédéral investira en 2020 un montant de 4,9 milliards de dollars pour soutenir la R&D liée à l'intelligence artificielle et à l'apprentissage automatique, ce qui représente une augmentation significative par rapport à l'année 2019. La Maison Blanche a en outre révélé en mai 2018 sa décision de créer un comité spécial consacré à l'IA. Alors que cette décision pourrait enchanter les milliardaires de la Silicon Valley, le fait que l'administration Trump n'ait ni prévu les financements pour le comité ni changé sa politique migratoire — pénalisante pour l'arrivée de talents étrangers — n'est guère rassurant.

Quant aux pays européens, leurs investissements en IA sont plutôt timides, même s'il faut bien évidemment tenir compte de la taille du PIB de chaque État. En mars 2018, le Président français Emmanuel Macron a promis de dédier 1,5 milliards d'euros d'argent public durant son quinquennat pour accompagner l'émergence d'un grand pôle mondial de l'intelligence artificielle. Dans le *Industrial Strategy White Paper*, le gouvernement britannique considère l'IA et les données comme l'un des quatre domaines importants où le Royaume-Uni devra être capable d'assumer le leadership mondial dans les années qui viennent. Une somme de £0,95 milliards est prévue pour soutenir le secteur de l'IA et une somme de £250 millions est dédiée au développement des voitures autonomes. Quant à l'Allemagne, selon la *Stratégie nationale en IA* publiée en novembre 2018, le pays investira 3 milliards d'euros afin de permettre à l'Allemagne de concurrencer la Chine et les États-Unis. Combiné avec 3 milliards d'euros issus du secteur privé, la totalité des investissements allemands atteint 6 milliards d'euros, soit le niveau le plus élevé en Europe.

## C. Achat de technologies étrangères

La fabrication d'équipements clés, tels que les puces spécialisées, est un domaine où la Chine reste encore faible. Alors qu'elle s'efforce de former ses champions technologiques nationaux, elle redouble d'efforts pour acheter des technologies clés venues de pays étrangers. Selon Forrester Research, la Chine (groupes publics et privés confondus) aurait augmenté en 2018 ses achats en technologie de 8 %, pour un montant total de 234 milliards de dollars.

D'ailleurs, la Chine applique au domaine de l'intelligence artificielle la même méthode de raccourcis utilisée depuis la politique de réforme et d'ouverture initiée en 1978. L'accès des entreprises étrangères au marché chinois se fait souvent sous conditions de transferts technologiques. Vu l'immense potentiel du marché chinois, les entreprises étrangères adoptent souvent une attitude coopérative.

La situation a changé avec l'adoption de la nouvelle loi sur les investissements étrangers le 15 mars 2019. Contenant en son sein 42 articles, elle entre en vigueur le 1er janvier 2020. Dans ce nouveau cadre, les investisseurs étrangers bénéficient du même traitement que les entreprises chinoises. Il est désormais interdit de recourir à des mesures administratives pour forcer les entreprises étrangères à opérer des transferts technologiques au profit de leurs homologues chinoises. La coopération technologique entre la Chine et les investisseurs étrangers doit se dérouler en conformité avec les principes d'égalité et de volonté. La Chine doit donc de plus en plus compter sur sa propre capacité d'innovation afin de soutenir le développement de l'IA à long terme.

## D. Crédit d'un environnement social favorable

Dans le *Plan*, il est indiqué qu'« il faut utiliser tous les moyens de communication traditionnels ou émergents pour informer le peuple chinois des avancées importantes réalisées dans le domaine de l'IA. Le développement de cette dernière doit s'ériger en consensus social. En outre, il convient de mobiliser la participation active de toute la société au développement de l'IA ». Ce n'est donc pas une surprise si 65 % des citoyens chinois se déclarent confiants en cette technologie, contre 29 % seulement dans d'autres pays — à en croire un chiffre du bureau britannique Dentsu Aegis, multinationale basée à Londres et spécialisée dans les médias, la communication et le marketing numérique.

En plus de politiques nationales favorables au développement de l'IA, le contraste saisissant entre l'optimisme de la société chinoise et l'angoisse de la société occidentale s'explique aussi par la formation et le profil des dirigeants politiques. Alors que la grande majorité des dirigeants chinois ont suivi un parcours universitaire en sciences dures, leurs homologues occidentaux ont plutôt fait leurs études en sciences humaines et sociales. Le résultat est que, par rapport aux occidentaux, le public chinois se montre dans l'ensemble plus confiant et plus ouvert face aux nouvelles technologies, ce qui explique le soutien plus élevé des chinois envers l'intelligence artificielle.

## **II. Un état des lieux du développement de l'intelligence artificielle en Chine**

Il est peut-être un peu tôt pour dresser un bilan sur le développement de l'IA dans l'Empire du Milieu. Néanmoins, depuis 2016, année où le programme d'intelligence artificielle Alpha Go a battu le troisième joueur mondial de go, Lee Sedol, des progrès spectaculaires ont été constatés en Chine. Bien que le pays soit toujours confronté à de nombreuses difficultés telles que l'innovation dans les puces ou le recrutement de talents, des changements positifs en matière d'applications de l'IA ont eu lieu dans les domaines de la santé digitale, de la protection environnementale, du maintien de l'ordre public et de la défense nationale, et ce grâce à plusieurs dynamiques économiques et sociales importantes.

### **A. Dynamique des start-up et des BATX**

En Chine, le marché de l'intelligence artificielle est en constante expansion. Prenons l'exemple du paiement en ligne, étape à laquelle la Chine a directement accédé en sautant celle des cartes de crédit. Le réseau social WeChat, développé par Tencent en janvier 2011 et équipé de la fonctionnalité de paiement en ligne, enregistre plus de 600 millions de transactions par jour. Selon la *MIT Technology Review*, en 2017, la valeur des transactions effectuées a dépassé 15 400 milliards de dollars, soit 41 fois plus qu'aux États-Unis. Le potentiel considérable du marché chinois de l'IA s'explique avant tout par l'essor des start-up et des géants des télécommunications et de l'informatique. Ces entreprises font déjà preuve d'une expertise très avancée, concurrente de celle des GAFAM américains.

iFlytek, société pékinoise créée en 1999 et cotée à la Bourse de Shenzhen, est un leader national en matière de traitement de la voix et de reconnaissance vocale. En plus de traductions dans toutes les langues, les logiciels développés par iFlytek peuvent distinguer une voix parmi vingt autres et la retranscrire en texte, même si cette voix s'exprime en dialecte et dans une salle bruyante. La société aide aussi la police à créer des bases de données biométriques pour permettre l'identification instantanée des individus.

La société Baidu, le « Google chinois », est en train de prendre une longueur d'avance sur son homologue américain. Avec une précision de 95,4 %, la technologie de reconnaissance d'images de Baidu dépasse celle de Google, dont l'exactitude se stabilise à 95,2 %. Le programme d'intelligence artificielle de la société est aussi parvenu à battre un champion de go, faisant ainsi jeu égal avec l'IA de Google. Avec son avantage inégalé dans le stockage de données, Baidu mise sur le développement de véhicules autonomes. La société a notamment commercialisé des minibus autonomes. En collaboration avec JAC Motors, BAIC et le groupe automobile Chery, Baidu a également lancé plusieurs modèles de voitures autonomes en 2019 et 2020.

La coopération sino-française est également très poussée. À l'occasion de la visite du Président Emmanuel Macron en janvier 2018, la France et la Chine ont officialisé un grand nombre de partenariats, dont la création d'un fonds commun d'investissement dans l'intelligence artificielle d'un milliard d'euros visant la recherche fondamentale et des applications industrielles. En outre, les deux pays se sont mis d'accord sur un échange annuel de vingt talents de chaque pays, créant ainsi un mécanisme régulier de partage de savoir-faire.

Quant aux États-Unis, la forte compétition sino-américaine dans l'IA n'empêche pas que les deux pays établissent des relations de coopération au niveau de leurs entreprises. Aux États-Unis, Baidu a créé de nombreux centres de recherche en IA. Après avoir installé un premier centre de recherche en reconnaissance vocale dans la Silicon Valley dès 2011, un second est devenu opérationnel le 9 octobre 2017. Son objet est de développer les technologies de conduite autonome. Le 11 octobre 2017, lors du sommet GeekWire, Baidu a annoncé vouloir ouvrir un troisième centre de recherche à Seattle, qui se focaliserait sur les technologies d'informatique en nuages.

Le dynamisme des entreprises chinoises s'explique par plusieurs facteurs. Premièrement, elles ont la possibilité de collecter un grand nombre de données grâce à l'utilisation généralisée des réseaux sociaux WeChat et Weibo. Comme il est indiqué dans la *Figure 1*, WeChat a conquis 1097,6 millions d'utilisateurs en Chine pendant le quatrième trimestre de 2018. En Chine, la moitié des smartphones sont équipés de l'application de paiement mobile, ce qui permet aux consommateurs d'effectuer leurs achats et leurs paiements en ligne. Avec un nombre gigantesque de données collectées, les entreprises chinoises peuvent ainsi entraîner et améliorer leurs algorithmes, prédire les

actions des consommateurs et fixer des stratégies commerciales avec une précision que beaucoup d'entreprises étrangères ne parviennent pas à obtenir.

*Figure 1 : Évolution du nombre d'utilisateurs WeChat en Chine de 2011 à 2018*



Par ailleurs, se focalisant dans un premier temps sur l'impact de l'IA sur l'emploi, les autorités chinoises n'imposent pas beaucoup de contraintes à leurs entreprises en ce qui concerne la dimension éthique de l'IA. Par conséquent, les conditions dans lesquelles les entreprises chinoises d'IA se développent sont relativement souples. Ce n'est pas le cas en Europe, où les préoccupations en matière d'éthique, de sécurité et de protection des données personnelles l'emportent, à tel point que les innovations en IA ont parfois du mal à avancer. L'équation entre l'innovation technologique, d'une part, et la préservation des données personnelles, d'autre part, reste toujours difficile à résoudre.

En plus d'une opinion sociale favorable à l'IA, les entreprises chinoises tirent profit de puissantes capacités de calcul des superordinateurs du pays. La Chine maintient depuis 2013 sa position de leader dans le classement Top 500 des superordinateurs les plus puissants du monde. Cette position a été encore consolidée avec la naissance du superordinateur Sunway Taihu Light en 2018. Tout en étant très économique en énergie, ce dernier est cinq fois plus puissant que le premier superordinateur états-unien (lui-même troisième du classement). Une grande quantité de données et une excellente capacité de calcul constituent les deux conditions dont l'intelligence artificielle ne peut se passer pour être « intelligente ». Réunissant ces deux conditions, la Chine pourra donc voir ses entreprises se développer sans entraves.

Il existe en Chine une véritable synergie entre les acteurs publics et privés afin de promouvoir le développement de l'IA. Selon le *Plan*, le gouvernement

chinois va prendre d'autres mesures pour soutenir le rôle décisif du secteur privé dans le financement de l'industrie de l'IA. Répondant aux exigences du gouvernement central, les autorités locales sont particulièrement actives et apportent un soutien décisif aux entreprises du secteur. En plus des BATX qui travaillent main dans la main avec les gouvernements à tous les niveaux, un autre exemple est fourni par SenseTime, société créée en 2014 et dont la valeur dépasse déjà le milliard de dollars grâce au soutien d'une quarantaine de gouvernements locaux. Ces derniers coopèrent avec elle pour la vérification des cartes bancaires et le développement de systèmes de sécurité. Quant à CloudWalk, société spécialisée en reconnaissance faciale, elle a reçu une subvention de 301 millions de dollars de la part du gouvernement local de Guangzhou afin d'installer un centre de traitement d'images dans la ville.

## B. E-santé

En Chine, la santé est un domaine prioritaire des applications de l'IA et des progrès spectaculaires sont constatés. Le marché chinois de la santé digitale s'est rapidement élargi à partir de 2014 et a généré plus de 10 milliards de yuans (1,3 milliards d'euros) en 2017. La Chine est depuis 2014 le deuxième pays le plus avancé du monde en ce qui concerne l'e-santé. Elle représente 37 % des parts du marché Asie-Pacifique. L'essor de la santé digitale dans l'Empire du Milieu se justifie avant tout par deux facteurs : la faible couverture médicale et les gains de temps et de budget promis par l'IA. Avec une population de 1,4 milliard d'habitants, la Chine n'offre qu'1,5 médecin pour 1000 personnes. Les investissements dans les applications sanitaires de l'IA pourraient aider à considérablement améliorer l'accès des citoyens à des services médicaux de qualité. En conséquence, les médecins, aidés par l'IA, pourraient s'occuper d'un plus grand nombre de patients par jour, tout en maintenant la même qualité de service.

Ayant ressenti le potentiel commercial du secteur de l'e-santé, les entreprises chinoises s'y lancent les unes après les autres. En tant que plate-forme phare de la stratégie « *Double H* » (*Health and Happiness*) du groupe Alibaba, AliHealth est devenue une référence dans l'e-pharmacie et les services pharmaceutiques *O2O* (*Online to Offline*) en Chine. Des habitants de plus de cent villes chinoises bénéficient déjà des services d'AliHealth. L'application Weiyi, développée par le groupe Weiyi du Zhejiang et réunissant plus de 200 000 médecins, se spécialise quant à elle dans la prise de rendez-vous médicaux en ligne. Les patients peuvent utiliser leurs portables et fixer gratuitement un rendez-vous avec un médecin de leur choix. Le groupe a conclu des partenariats avec plus de 1 900 hôpitaux chinois, permettant ainsi

à de nombreux patients — notamment ceux habitant à la campagne et dans des régions reculées — de se faire soigner sans avoir à effectuer des déplacements coûteux et parfois inutiles.

Le groupe d'assurances Ping An est un autre exemple à travers lequel on peut se faire une idée de l'état d'avancement des applications d'IA dans la santé. Fondé en 1988 à Shenzhen, et avec un chiffre d'affaires en 2017 de 975 milliards de yuans (environ 125 milliards d'euros), Ping An s'impose aujourd'hui comme numéro un mondial de l'assurance, dépassant les géants Allianz et Axa. Sa capacité d'innovation en matière d'IA contribue au succès commercial que le groupe a rencontré. L'application Good Doctor, branche médicale de Ping An et cotée à la Bourse de Hong-Kong, attire déjà 193 millions d'usagers grâce aux services intégrés qu'elle propose. En analysant les photos et les radiographies, les algorithmes de Ping An parviennent à diagnostiquer vingt-trois affections simples et à détecter les risques de cancer pulmonaire. Fort d'un réseau de plus de 1 000 médecins, de 3 100 hôpitaux et de 1 100 cliniques, l'application Good Doctor rend possible la prise de rendez-vous médicaux rapide et les consultations à distance, ce qui optimise la circulation des ressources médicales en réduisant l'écart de qualité des prestations entre les habitants des villes et des campagnes. Sans avoir à faire des déplacements chronophages, les patients peuvent directement acheter des médicaments ou des compléments alimentaires grâce à l'application Good Doctor.

## C. Préservation de l'ordre public

En Chine, l'urbanisation, poussée principalement par l'exode rural, se révèle être un processus assez brutal. Entre 1978 et 2016, le taux d'urbanisation chinois a plus que triplé, passant de 17,92 % à 57,35 %. Or le fait que la Chine s'urbanise avant que les villes développent une capacité d'accueil correspondante a généré de nombreux problèmes sociaux.

Bien que la répartition géographique des travailleurs migrants entre l'est, le centre et l'ouest du pays s'équilibre et que leur circulation interprovinciale ralentisse, les quatre métropoles situées à l'est du pays, Beijing, Shanghai, Guangzhou et Shenzhen, absorbent depuis très longtemps une grande quantité de migrants. En conséquence, ces villes souffrent d'embouteillages, d'instabilité et de raréfaction des ressources, notamment sociales. Cette situation risque de se perpétuer si les villes concernées ne trouvent pas les moyens d'améliorer leur capacité d'accueil. Il en résulte un certain nombre de troubles et de tensions sociales.

En ce qui concerne le maintien de l'ordre public, l'IA possède trois avantages forts. Tout d'abord, l'algorithme, qui s'entraîne de manière répétitive et se nourrit d'une quantité gigantesque de données, peut atteindre une précision impressionnante que même un personnel extrêmement qualifié et chevronné ne peut obtenir. Les résultats acquis par l'IA jusqu'à présent en matière d'analyse d'imagerie médicale révèlent déjà son potentiel. Il est possible d'affiner encore davantage l'exactitude des évaluations de l'algorithme, puisque l'IA est capable d'apprendre par elle-même. Ensuite, contrairement aux policiers ou aux gardiens, l'IA ne se fatigue jamais. Elle peut travailler 24 heures sur 24, sans demander d'augmentation de salaire ni d'améliorations de ses conditions de travail. Enfin, en comparaison avec les personnes chargées de la sécurité qui peuvent être attaquées en cas de conflit avec des malfaiteurs, l'IA est imbattable. Elle surveille les délinquants, enregistre leurs actions et offre des informations précieuses aux policiers qui peuvent les arrêter sans prendre de risques majeurs. À l'aide de l'IA, le maintien de l'ordre public peut grandement s'améliorer, et ce à un prix très abordable. La Chine a ainsi développé des technologies de reconnaissance faciale et vocale très avancées, grâce auxquelles l'identification et l'arrestation des contrevenants peuvent s'effectuer avec efficacité.

À Shanghai, à Shenzhen, ainsi que dans de nombreuses autres villes chinoises, l'IA est aussi mise à profit par les autorités locales dans la gestion du trafic. Que des agents de police soient présents ou non, les piétons qui traversent au rouge ou qui n'empruntent pas les passages réservés voient leur visage apparaître sur des écrans installés dans les arrêts de bus ou dans d'autres endroits moins discrets. Le visage ne disparaît que lorsqu'une amende est payée au commissariat du quartier concerné. À Pékin, les distributeurs de papier de certaines toilettes publiques sont équipés de technologies de reconnaissance faciale pour lutter contre les abus : une même personne ne peut pas utiliser plus de soixante centimètres de papier toilette en neuf minutes. Étant donnée la dimension écologique de cette mesure, l'exemple de la capitale est suivi par d'autres villes du pays.

Concernant les voitures, la société iFlytek est parmi les premières sociétés chinoises à avoir appliqué l'IA dans l'industrie automobile. En 2017, elle a mis sur le marché son nouveau produit, Xiaofeiyu, une application vocale polyvalente installée à l'intérieur de la voiture. Il suffit au conducteur de donner des ordres vocaux pour connaître le meilleur trajet, obtenir des informations en temps réel, communiquer à distance ou écouter de la musique. Forte d'une technologie de reconnaissance vocale de premier niveau, l'application Xiaofeiyu peut même fonctionner dans des environnements complexes, avec un taux de reconnaissance de 90 %.

L'IA a également fait son apparition dans deux autres domaines touchant au maintien de la sécurité publique : la prédition des actes criminels et l'intervention de la police avant que les crimes ne soient commis. Cette

nouvelle direction confirme ce que Li Meng, vice-ministre chinois des Sciences et des Technologies, a annoncé en juillet 2017 : « Si l'on utilise correctement nos systèmes intelligents et nos équipements intelligents, on peut savoir à l'avance [...] qui pourrait être un terroriste ou, plus généralement, qui pourrait faire quelque chose de mal ». CloudWalk, société pionnière dans la technologie de reconnaissance faciale, est ainsi un partenaire privilégié de la police chinoise. L'algorithme de CloudWalk peut effectuer des comparaisons entre les images recueillies et celles stockées dans la base de données de la police. Après une analyse des comportements de personnes suspectes, la société identifie les risques de passage à l'acte criminel des individus concernés. Les forces de l'ordre sont alors prévenues et sollicitées afin d'empêcher ces crimes présupposés d'être commis.

La presse occidentale ainsi que différents chercheurs critiquent l'usage que la Chine fait de l'intelligence artificielle dans le domaine du maintien de l'ordre public. Ils se posent des questions sur les intentions du gouvernement chinois. Un journaliste commence ainsi, ironique, un article : « *Big brother is watching you*, enfin non, *is carrément arresting you* ».

Face aux commentaires uniformément négatifs de la presse occidentale sur l'intelligence artificielle chinoise, il convient peut-être d'apporter quelques nuances car, en mettant trop l'accent sur la dimension répressive de l'IA en Chine et en diabolisant outre mesure les relations entre les autorités chinoises et les citoyens, les journalistes occidentaux oublient les bénéfices que cette technologie a apportés et pourra encore apporter aux habitants du pays. Prenons l'exemple de Xiaofeiyu, accusée d'espionnage au profit du gouvernement chinois. L'avantage de cette application par rapport au GPS traditionnel consiste à permettre aux conducteurs de commander leurs voitures vocalement, de garder toujours les mains sur le volant et donc de rester en sécurité. Concernant la collecte des informations personnelles, il n'y a pas de réelles différences entre Xiaofeiyu et un GPS classique. La seule différence est qu'avec Xiaofeiyu le chauffeur est en sécurité au moment où ses informations personnelles sont collectées et qu'avec un GPS ordinaire il risque sa vie en tapant le numéro de la rue sur le cadran — numéro qui, bien sûr, peut de toute façon être aussi collecté avec ses informations personnelles.

## D. Protection de l'environnement

Concernant la protection de l'environnement, le *Plan* indique que la Chine envisage de construire des réseaux intelligents de supervision environnementale et des plateformes de services. Elle ambitionne de mettre sur pieds des méthodes de modélisation intelligente pour suivre de près la

consommation de ressources naturelles et d'énergie, ainsi que l'émission de matières polluantes. Des systèmes de prévention et de contrôle intelligents seront construits dans la bande économique du fleuve Yangtsé et les provinces contiguës du Hebei, de Pékin et de Tianjin, dans le but de prévenir la pollution environnementale et les catastrophes naturelles.

En Chine, environ 70 % de l'électricité est produite à partir du charbon. Selon une étude menée par l'Université de Californie à Berkeley, la pollution atmosphérique est responsable de près de 4 000 morts chaque jour en Chine. Le développement économique au prix de la pollution environnementale n'est pas une voie durable pour la Chine et risque de devenir un goulot d'étranglement qui empêchera la réalisation d'une société de moyenne aisance.

Pour lutter contre la pollution environnementale et mettre à niveau la structure énergétique, il semble judicieux pour la Chine d'intégrer l'IA dans ses stratégies écologiques. Des initiatives issues du secteur privé émergent et sont prolongées par les acteurs publics. Le 10 juin 2017, les ingénieurs d'AliCloud, division informatique du groupe Alibaba, ont ainsi lancé une initiative technologique de protection environnementale : *ET Environment Brain*. Mettant à profit la technologie de perception intelligente de l'environnement et sur la base d'images collectées par des satellites, *ET Environment Brain* est capable d'effectuer des analyses croisées sur la température, le vent, la pression atmosphérique, l'humidité, les précipitations et le rayonnement solaire. Plusieurs provinces littorales l'ont utilisé dans la surveillance de la production, du transfert, du traitement et de la réutilisation des déchets solides. Elles l'ont en outre appliqué pour identifier de manière précoce les fausses déclarations sur les productions de déchets. *ET Environment Brain* a aidé les provinces pilotes à lancer au total 131 alertes environnementales, avec une efficacité de 93 %.

Concernant le volet énergétique de l'économie chinoise, le taux de croissance de la consommation énergétique est passé de 5 % en 2016 à 6,6 % en 2017. La consommation pourrait encore augmenter de 40 % au cours des quinze années à venir. Le gaspillage d'énergie et le manque d'énergies renouvelables s'ajoutant aux problèmes de consommation excessive, la sécurité de l'approvisionnement énergétique pourrait être mise en question en Chine, impactant en retour sa croissance économique et son urbanisation.

Les autorités publiques et les entreprises chinoises font donc appel à l'IA afin de mieux gérer l'enjeu stratégique de sécurité énergétique. En mars 2012, le gouvernement chinois a publié son *Douzième plan quinquennal sur l'industrialisation des sciences et des technologies relatives au réseau électrique intelligent*, reconnaissant l'importance primordiale du réseau électrique intelligent dans la mise en œuvre des stratégies énergétiques nationales et dans l'optimisation de la distribution des énergies. La publication

de ce document officiel a pu favoriser l’innovation scientifique du Laboratoire national de contrôle sur la protection et le fonctionnement du réseau électrique intelligent. Rattaché à la State Grid Corporation of China (SGCC), chargée de la gestion de 80 % du réseau électrique de la Chine, et en partenariat avec des universités prestigieuses telles que l’Université Tsinghua, l’Université du Zhejiang et l’Université de Nankin, le laboratoire a développé une expertise importante dans l’ajustement intelligent et l’automatisation des équipements des stations d’électricité locales.

Vu les défis écologiques auxquels la Chine est confrontée, l’application de l’IA dans le domaine des énergies a de fortes chances de s’élargir à une plus grande échelle dans le pays. À l’heure actuelle, nous vivons dans un monde confronté à un double défi : d’une part, la quantité d’énergies traditionnelles utilisables se réduit de jour en jour ; d’autre part, bien que les énergies renouvelables s’annoncent prometteuses, l’intermittence de la production et les difficultés de stockage posent des problèmes non négligeables. Dans ce cadre, l’IA peut aider à minimiser les conséquences néfastes de ces deux défis. En stockant et en analysant les données sur la consommation énergétique des utilisateurs, elle pourrait aider les entreprises et les acteurs publics à fixer des politiques incitant les utilisateurs à adopter des comportements plus économies en énergie. Elle pourrait également aider à remédier au problème d’intermittence des énergies renouvelables, amplifiant ainsi leurs contributions économiques et sociales.

## E. Défense nationale

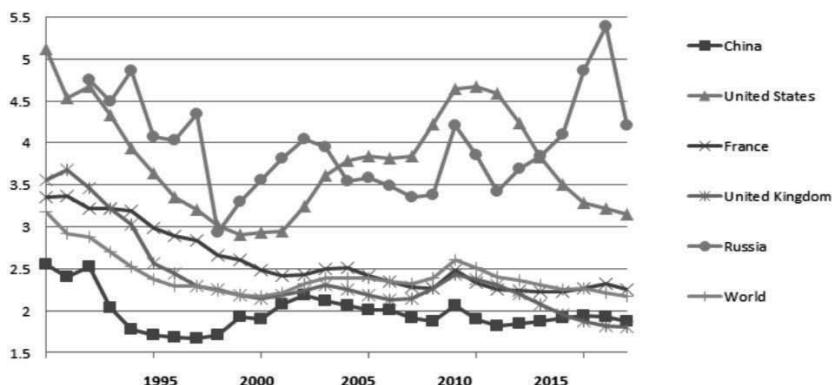
En juillet 2017, dans un discours adressé à l’Académie des sciences militaires, le Président chinois Xi Jinping a déclaré que la Chine devrait construire des instituts de technologie militaire de classe mondiale, l’objectif étant de contenir la supériorité militaire des États-Unis dans les points chauds du secteur Asie-Pacifique. Répondant aux besoins stratégiques du pays, des recherches scientifiques en matière d’applications militaires de l’IA se multiplient afin de relever les défis qui s’annoncent.

Deux raisons peuvent aider à comprendre l’essor des recherches dans ce domaine. D’une part, dans la région Asie-Pacifique, où les pays semblent craintifs vis-à-vis de la montée en puissance de la Chine, cette dernière a besoin de se prémunir contre tous les aléas possibles, dans le but de perpétuer et de sécuriser son développement. Puisque la Chine et les États-Unis figurent tous les deux parmi les pays possesseurs de l’arme nucléaire, il est important pour la Chine de prendre une longueur d’avance par rapport aux États-Unis. Autrement dit, en dehors de la dissuasion nucléaire, se munir d’un outil encore

plus efficace pourrait donner à l'Empire du Milieu un avantage décisif dans les jeux géopolitiques de la région.

D'autre part, les nouvelles technologies à base d'IA pourraient aider la Chine à restreindre l'augmentation de ses dépenses militaires. Les armes nucléaires sont jusqu'à présent les plus puissantes et les plus destructrices. Logiquement, la possession de telles armes doit être assez rassurante pour que les États maintiennent ou réduisent leurs budgets militaires. Ce n'est cependant pas le cas. Selon les données de la Banque mondiale (*Figure 2*), bien que les dépenses militaires américaines se soient réduites après la guerre froide, elles restent toujours au-dessus de 3 % du PIB. Ce chiffre s'est même élevé à 4,67 % en 2010. Ce n'est évidemment pas un exemple que la Chine peut se permettre de suivre, étant donné les moyens financiers dont elle a besoin par ailleurs pour relever les nombreux défis écologiques, démographiques et environnementaux qui se profilent. Le développement de nouvelles méthodes de défense nationale lui permettrait donc de mieux se protéger contre les éventuels risques sécuritaires, cela à un coût moins élevé.

*Figure 2 : Dépenses militaires dans les cinq pays membres permanents du Conseil de sécurité des Nations Unies (% du PIB)*



*Source : Banque Mondiale*

Des progrès spectaculaires ont ainsi été obtenus en matière d'applications militaires de l'IA. Par exemple, grâce à ses avancées dans le domaine des drones moyenne altitude longue endurance (Male), la Chine a pu mettre en avant sa dernière version du drone de combat *Wing Loong*, le *Wing Loong II*. Fabriqué par le groupe d'aéronautique China Aviation Industry Corporation, conglomérat d'État de l'aérospatiale et de la défense, ce drone a fait l'objet de commandes à l'exportation, notamment de la part de l'Arabie Saoudite, pour une quantité totale de 300 appareils. Et, fin 2017, Chengdu Aircraft Industrial Group (CAC), filiale de China Aviation Industry Corporation, a procédé à la

première livraison de drones *Wing Loong II* à ses clients. En comparaison avec la Chine, l'Europe, qui se trouve en période de spécification de son futur Male, semble en retard.

À l'aide de l'IA, la Chine est également en train d'élaborer un système embarqué à bord des sous-marins nucléaires d'attaque ou lanceurs d'engin. L'objectif est d'améliorer la capacité d'interprétation et de décision du commandement, ce dernier étant susceptible d'être impacté par le stress du confinement. Toujours dans le domaine de la marine, la Chine a commencé début 2018 la construction d'un site de test pour les navires sans pilote à Zhuhai. Financée par le gouvernement local de Zhuhai, la Société de classification de Chine, l'Université des technologies de Wuhan et Océanalpha, producteur et fournisseur de drones de première classe mondiale et société spécialisée dans les drones marins et les navires sans pilote, la construction du site s'effectue dans le but de reproduire, tester et vérifier les fonctions des navires sans pilote, notamment la planification des itinéraires, le repérage, l'accostage et l'appareillage.

## F. Rivalité avec les Etats-Unis

À l'heure actuelle, la Chine et les États-Unis sont considérés comme les deux pays les plus importants en matière d'IA. En raison de la méfiance traditionnelle des États-Unis envers les intentions supposées impérialistes de la Chine et de la ferme volonté du gouvernement chinois de se maintenir à un haut niveau dans la zone Asie-Pacifique, il est indispensable de bien prendre la mesure de la compétition entre les deux pays.

La concurrence dans le domaine de l'IA entre les deux puissances s'est intensifiée en 2017, suite à la publication de trois documents officiels par l'administration Obama entre octobre et décembre 2016. Quand on regarde de près le *Plan* publié par le gouvernement chinois, on s'aperçoit qu'il présente beaucoup de similarités avec le *National Artificial Intelligence Research and Development Strategic Plan* rendu public en octobre 2016 par l'administration Obama.

Le rapport *Preparing for the Future of Artificial Intelligence*, également publié en octobre 2017 par le gouvernement américain, faisait remarquer que les États-Unis n'étaient plus la première nation en termes de publications sur l'apprentissage profond. En fait, la Chine était en train de prendre le dessus non seulement en matière de publications scientifiques mais aussi à d'autres niveaux. Entre 2001 et 2010, on a compté 6 046 publications aux États-Unis, contre 554 en Chine. Mais, entre 2011 et 2015, la Chine a dépassé les États-

Unis et s'est placée en première position mondiale en terme de volume de publications relatives aux technologies d'IA.

Bien que les américains dominent toujours le secteur des puces spécialisées, des avancées considérables ont été constatées du côté de la Chine. Cambricon Technologies Corp Ltd, fabricant chinois de puces spécialisées en intelligence artificielle, a lancé sa puce Cambricon-IA en 2016. La puce peut gérer 16 milliards de neurones virtuels par seconde. Sa capacité de pointe peut atteindre les 2 milliards de synapses par seconde. Le 3 mai 2018, la société a présenté à Shanghai ses deux nouveaux produits : une puce intelligente basée sur le cloud Cambricon MLU100 et un Cambricon 1M processor. La nouvelle puce peut être utilisée dans l'apprentissage automatique basé sur le *cloud*, y compris la visualisation, les données audio et le traitement du langage naturel (NLP). Consciente de son retard dans les technologies essentielles de l'IA et désirant vivement briser sa dépendance technologique par rapport aux pays étrangers, la Chine investit massivement dans la R&D des semi-conducteurs. Un fonds spécifique d'un montant total d'au moins 150 milliards de yuans (environ 23,96 milliards de dollars) sera bientôt consacré à cette fin. D'ailleurs, moins freinée par les débats éthiques et les tentatives de régulation, il est probable que la Chine connaisse une ascension plus rapide et plus radicale dans les domaines de la sécurité nationale et des armes autonomes.

En comparaison avec les États-Unis, la Chine attire davantage d'investissements pour soutenir le développement de son industrie de l'IA. En 2017, 15,2 milliards de dollars ont été investis à l'échelle mondiale dans des start-up spécialisées dans le secteur, et près de la moitié de cette somme est allée directement vers la Chine, contre 38 % en direction des États-Unis. C'est la première fois que la Chine dépasse les États-Unis sur ce type d'investissements. Dans son livre *AI Superpowers : China, Silicon Valley and the New World Order*, Lee Kai-Fu confirme la position privilégiée de l'Empire du Milieu dans le domaine de « l'AI Internet » et « l'AI Perception » par rapport aux États-Unis depuis 2010. Grand spécialiste chinois de la question de l'IA, il note que : « *If data is the new oil, then China is the new Saudi Arabia* ». Tout en soulignant le leadership de la Chine face à son concurrent américain, Lee Kai-Fu insiste sur la nécessité pour les deux pays, dotés tous deux d'une puissance technologique exceptionnelle, de prendre leurs responsabilités dans la création de nouvelles valeurs culturelles et d'apporter une réponse forte aux bouleversements que l'intelligence artificielle ne va pas tarder à semer dans le monde du travail et les réglementations internationales.

Cela ne veut pas dire que la Chine deviendra certainement une superpuissance de l'IA, car le pays a des limites importantes à surmonter. En comparaison avec les États-Unis, le niveau de développement général de l'IA est moins élevé. La Chine a notamment des retards concernant la théorie fondamentale, les algorithmes clés, les puces spécialisées et les logiciels. Les

instituts de recherche et les entreprises n'ont pas encore pu former un véritable écosystème numérique ni une chaîne industrielle avec une influence internationale. Les experts et les spécialistes de l'IA manquent cruellement. La Chine a un besoin urgent de mettre sur pieds une infrastructure et un système de lois, de réglementations et de normes éthiques pour mieux encadrer le développement l'IA.

### **III. Les visées stratégiques des investissements colossaux en IA de la Chine**

Quand la Chine investit des sommes très élevées dans l'IA, l'Europe, les États-Unis et, surtout, les pays d'Asie du Sud-Est se demandent si derrière cela ne se cache pas une intention impérialiste de domination géopolitique. Pour un État, l'IA est non seulement importante pour promouvoir sa croissance économique mais aussi pour assurer sa sécurité. Vu l'importance stratégique de l'IA, il n'est pas surprenant que la communauté internationale se pose des questions sur intentions de la Chine. Comme Vladimir Poutine l'a dit à juste titre, « le pays qui sera leader dans le domaine de l'intelligence artificielle dominera le monde ». Faisant écho à cette déclaration franche, Elon Musk a aussitôt tweeté que « la compétition pour la supériorité nationale en matière d'IA sera la cause la plus vraisemblable de la troisième guerre mondiale ». Qui plus est, à en croire Jean Mearsheimer, chef de file du réalisme structurel, la meilleure méthode pour s'imposer comme leader international est d'abord de s'imposer comme leader régional. Dans l'esprit des occidentaux, si la Chine, à l'aide de la technologie IA, peut arriver à devenir une puissance incontestable en Asie-Pacifique, elle s'imposera tôt ou tard comme un leader mondial.

Cette façon d'appréhender la Chine est on ne peut mieux exprimée par une chercheuse européenne qui décrit la trajectoire que la Chine empruntera pour le développement de son IA en ces termes : « Le Conseil des affaires d'État a présenté en juillet 2017 sa vision globale en trois étapes pour le secteur de l'IA : rattraper les États-Unis d'ici 2020, les dépasser en 2025 et devenir leader mondial en 2030 ». Les intentions du gouvernement chinois en matière d'IA, ainsi que les efforts que la Chine a déployés dans le projet « *One Belt One Road* » (OBOR) et la Banque asiatique d'investissement pour les infrastructures (AIIB), sont souvent interprétées de cette façon simplifiée. Mais ce serait un véritable gâchis si la peur et l'inquiétude suscitées par la Chine détournaient les chercheurs occidentaux des vrais enjeux.

Les investissements colossaux que la Chine consacre à l'IA s'expliquent tout d'abord par l'environnement international inamical dans lequel elle se

trouve. Sa montée en puissance sur la scène internationale a favorisé l'alliance entre le Japon et l'Inde, et fait l'objet de nombreuses convoitises. Elle a aussi suscité chez d'autres pays asiatiques un appel à l'intensification de la présence américaine dans la région. L'IA se révèle comme une technologie extrêmement prometteuse. Pour la Chine, il est donc impératif d'être leader mondial dans ce domaine pour éviter d'être prise de court par les vicissitudes internationales. Mais il convient de nuancer la menace que le développement de l'IA chinoise poserait à la sécurité internationale.

Tout d'abord, il faut reconnaître le fait que tous les pays veulent devenir des leaders mondiaux dans le domaine. En France, par exemple, le rapport Villani indique que le pays devrait prendre le leadership afin d'ériger les développements parallèles de l'IA et de la transition écologique en enjeu mondial. Dans *Growing the Artificial Intelligence Industry in the UK*, rapport rédigé par Dame Wendy Hall et Jérôme Pesenti sur la demande des Secrétaire d'État au Business et à la Culture du Royaume-Uni, il est indiqué que le pays est en position forte pour piloter, au lieu de suivre, à la fois le développement de la technologie IA et son déploiement dans tous les secteurs de l'industrie, de l'éducation et du gouvernement. Quant à l'Allemagne, le pays a publié en novembre 2018 sa stratégie nationale de l'IA (en allemand *Nationale Strategie für Künstliche Intelligenz AI Made in Germany*). Il y est écrit que « *Germany is to become a leading centre for AI, especially through the pursuit of a speedy and comprehensive transfer of research findings to applications and the modernisation of administration in the context of the needs of a state based on the rule of law* ». Et ce document de déclarer que « *We want Germany to build on its strong position in Industrie 4.0 and to become a world leader in AI applications in this field* ».

Il n'est donc pas légitime d'accuser un pays d'avoir l'ambition d'être numéro un dans un domaine dans la mesure où les moyens pour y parvenir ne sont pas malhonnêtes vis-à-vis des autres pays. Les accusations et les soupçons lancés à l'encontre de la Chine s'expliquent avant tout par le fait que la Chine est un nouveau challenger sur la scène internationale. La suprématie américaine est tellement omniprésente dans les esprits qu'il paraît toujours plus ou moins anormal d'imaginer qu'ils puissent être tôt ou tard dépassés par un autre pays — même quand ce n'est pas la Chine.

Face à de nombreux défis domestiques, la Chine a un immense besoin d'IA pour les relever. Elle est pour l'instant le plus grand émetteur de gaz à effet de serre dans le monde. Elle doit aussi améliorer les services publics au bénéfice de ses 1,4 milliard d'habitants, d'autant plus qu'elle est déjà entrée dans une société du vieillissement. L'IA pourrait aider la Chine à résoudre ces problèmes, ce qui est dans l'intérêt de tous les États du monde. Sinon, le pays risquerait d'exporter ses problèmes et de susciter une instabilité internationale.

L'IA est pour la Chine un enjeu de croissance économique et de sécurité nationale. Une campagne nationale de réduction de la surcapacité industrielle a été lancée en 2016 : 500 millions de tonnes de charbon et 100 à 150 millions de tonnes d'acier devaient ainsi être produites en moins de 2016 à 2020 (le total de la réduction est de 1,2 milliard de tonnes) et 1,8 million de postes dans le secteur sidérurgique devaient être supprimés. La décision des autorités chinoises entraînera la fermeture de la filière charbon et d'usines d'acier ; elle risque d'engendrer un chômage massif dans ces secteurs. La Chine a besoin de trouver une nouvelle dynamique de croissance plus économique en énergie et qui pourra créer une nouvelle période de prospérité économique continue pendant une dizaine d'années, voire plusieurs décennies. L'IA répond à cette demande.

Si la Chine veut défendre sa souveraineté numérique, elle ne peut pas continuer de compter sur la méthode des raccourcis en raison des changements de cadre juridique applicable aux entreprises étrangères implantées en Chine et des risques d'usurpation de données pouvant émerger dans les processus de transferts technologiques. Si elle reste inactive concernant l'IA, les GAFAM risquent d'usurper et d'exporter les données personnelles des citoyens chinois — ce qui est déjà arrivé aux européens, et c'est l'une des raisons pour lesquelles l'Europe a adopté le RGPD en 2016. D'ailleurs, la Chine a aussi investi dans l'IA dans un but préventif, afin de se protéger contre les attaques extérieures telles que les frappes automatiques par des drones, des attaques informatiques de type APT (*Advanced Persistent Threat*) ou l'utilisation de la désinformation.

Les efforts que la Chine est en train de déployer dans le domaine de la coopération internationale en IA manifestent la volonté du pays de jouer un rôle important dans la réglementation de l'IA de demain. La tenue de la Conférence mondiale de l'intelligence artificielle (*World Artificial Intelligence Conference*) du 17 au 19 septembre 2018 à Shanghai s'inscrit parfaitement dans cette logique de parrainage chinois de la coopération internationale. Cet événement traduit à la fois son leadership en IA et sa ferme volonté de piloter le développement de cette technologie dans le monde de demain. Réunissant plus de 200 entreprises phares de l'IA et des experts d'environ 40 pays, cette conférence mondiale a été pour les autorités chinoises une occasion de rassurer ses partenaires et de leur redonner espoir dans le contexte de guerre commerciale sino-américaine. Ce n'est donc pas une surprise si Microsoft, Amazon et consorts ont annoncé le 18 septembre, soit le deuxième jour de la Conférence, qu'ils allaient ouvrir des centres de recherche en IA à Shanghai.

La Chine souhaite figurer parmi les États qui établissent les règles et les normes de ce secteur, ce qui se traduit par le fait que la Chine encourage ses entreprises à participer à la fixation des standards internationaux. À l'appui de l'initiative OBOR, la Chine va ainsi se lancer dans la création de bases de

coopération internationale en IA et de centres de recherche conjoints. Et elle soutient l'établissement d'organisations internationales de régulation de l'IA afin de fixer ensemble des standards internationaux.

Cette demande est plutôt légitime. Il est grand temps de mettre fin à la situation de monopole des pays développés dans la fixation des normes internationales auxquelles les pays en voie de développement sont obligés d'obéir, sous peine d'isolement ou même d'exclusion.

Enfin, l'IA est pour le régime chinois un moyen de contrôle social. En Chine, on compte aujourd'hui 170 millions de caméras et ce chiffre va probablement tripler dans les trois années qui viennent. Néanmoins, toutes ces caméras ne sont pas utilisées à des fins de surveillance civile. Elles servent aussi à lutter contre la délinquance, le terrorisme et l'incivisme, ainsi qu'à identifier des enfants kidnappés ou des victimes de trafics d'êtres humains. Concernant la surveillance sociale, il s'agit d'un phénomène qu'on trouve partout dans le monde. Le test des portes à reconnaissance faciale dans des grandes chaînes de supermarchés américaines telles que Walmart, Target et Jacksons Food Store est un exemple. Les exagérations de la presse occidentale à propos du cas chinois se vendent certes très bien auprès des lecteurs, mais cela encourage le maintien d'un certain nombre de stéréotypes et, surtout, ne contribue en rien à améliorer la situation de la Chine. Avec l'utilisation généralisée de smartphones, de cookies et de réseaux sociaux, le problème de la surveillance se mondialise. S'il n'existe pas de difficultés liées au recueil de données et à leur utilisation pour le profilage des internautes en Europe, on peut se demander pourquoi tant d'associations et d'ONG militent pour la protection de la vie privée.

## Conclusion

Durant le premier quinquennat du Président chinois XI Jinping, la Chine a porté une attention particulière à l'intelligence artificielle, avec comme point de départ la publication en 2013 des *Opinions*. En plus d'investissements publics considérables dans l'IA, la Chine a fixé des objectifs ambitieux et quantifiables, a créé un environnement social favorable et a promu des partenariats entre les secteurs public et privé pour mieux soutenir le développement des champions numériques nationaux. Avec la suppression de l'obligation de transfert technologique pour les entreprises étrangères implantées en Chine, mesure effective à partir du 1er janvier 2020, il est probable que la Chine investira encore plus dans l'achat de technologies de pointe issues de pays étrangers. Ayant déployé des efforts acharnés, des progrès spectaculaires ne cessent de se produire en Chine, surtout depuis 2016.

Aujourd’hui, l’IA est très sollicitée pour promouvoir la santé digitale, la protection environnementale, la défense nationale et la préservation de l’ordre public.

La Chine est devenue un pays aussi important que les États-Unis dans le secteur de l’IA. Elle s’est même montrée plus performante que ces derniers en termes de volume de publications et les a distancés en 2017 en attirant davantage d’investissements. Elle a donc de fortes chances de devenir le leader mondial de l’IA, vu la disponibilité d’une quantité gigantesque de données, les performances inégalées des supercalculateurs chinois, et surtout l’opinion publique nettement favorable à ces nouvelles technologies. Néanmoins, l’état d’avancement actuel de la Chine ne permet pas de prédire avec certitude son prochain leadership, étant donné ses lacunes importantes en matière de fabrication de puces spécialisées, de recrutement de talents de haut niveau, d’élaboration d’algorithmes et de logiciels clés. En outre, les autorités chinoises doivent œuvrer à la définition de normes éthiques et à l’établissement d’un cadre juridique fiable afin de contenir les effets indésirables de l’IA.

Face aux nombreuses réserves et critiques concernant les intentions dites impérialistes du gouvernement chinois, il convient d’apporter trois clarifications. Premièrement, vu l’environnement international inamical dans lequel la Chine se trouve et les nombreux défis domestiques auxquels le pays est confronté, il est normal que la Chine fasse appel à l’IA qui pourrait l’aider à résoudre différents problèmes sociaux. Pour l’État chinois, l’utilisation de l’IA afin de résoudre ses défis domestiques est un objectif non moins important que celui de devenir le leader mondial du secteur. Deuxièmement, puisque l’intelligence artificielle est pour un État à la fois un enjeu de croissance économique et un enjeu de sécurité nationale, il est inévitable que tous les pays du monde cherchent à occuper une position privilégiée dans la concurrence internationale autour des nouvelles technologies. Troisièmement, la quête chinoise de leadership mondial s’explique aussi par le besoin de briser le monopole des pays développés-occidentaux dans la définition des normes internationales. Ces dernières, dont la plupart ont été négociées entre ces pays, peinent à être respectées par les pays en voie de développement, qui, en revanche, créent leurs propres cadres de coopération, tels que le AIIB ou l’Organisation de coopération de Shanghai.

Le monde dans lequel on vit est en train de connaître des changements extrêmement profonds (et intéressants), avec les États-Unis qui, sous la direction d’un président, Donald Trump, imprévisible, perdent petit à petit leur *hard* et leur *soft powers*. Dans un contexte international marqué par l’instabilité, il est plus que jamais important d’adopter une attitude pragmatique en essayant de mettre provisoirement de côté les discussions idéologiques. Qu’on le veuille ou pas, la Chine a de vraies chances de s’imposer à l’avenir comme le leader mondial de l’IA. Les pays européens

comme la France et l'Allemagne sont très performants dans la formation de mathématiciens de haut niveau, la recherche fondamentale, ainsi que les réflexions sur les impacts éthiques et sociétaux de l'IA. Si, suivant les principes d'égalité et de reciprocité, les différentes parties parviennent à joindre leurs efforts au sein de projets d'IA communs, tels que la formation de nouveaux talents ou le développement de voitures autonomes, les bénéfices que ces coopérations généreront seront immenses.

## Références

### *Articles/ouvrages*

« Chine : 25 milliards de dollars en une seule journée pour Alibaba », Le Point, le 12 novembre 2017.

« L'e-santé en Chine : un marché prometteur malgré les obstacles », Daxue Conseil, le 30 mars 2017.

« La Chine a débuté la construction d'un site pour tester les navires sans pilote », le 14 février 2018, lien : <http://french.peopledaily.com.cn/n3/2018/0214/c31357-9427367.html>

« La Chine, pépinière pour les solutions de smart energy ? », VVR International, le 7 avril 2018, lien : <https://www.vvrinternational.com/la-chine-pepiniere-pour-les-solutions-de-smart-energy/>

« Le Chinois Alibaba bat son record de ventes sur Internet en un jour », BFMTV, le 12 novembre 2018, lien : <https://www.bfmtv.com/economie/le-chinois-alibaba-bat-son-record-de-ventes-sur-internet-en-un-jour-1564364.html>.

« Policy Paper : AI Sector Deal », le 21 mai 2019, gouvernement britannique, lien : <https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>

« Voitures et enceintes : Baidu accélère sur l'intelligence artificielle », Le Point, le 16 novembre 2017.

Braun, Elisa, « Intelligence artificielle : la Chine attire plus d'investissements que les Etats-Unis », Le Figaro, le 16 février 2018.

Braunstein, Juergen, Marion Laboure and Haiyang Zhang, « La Chine a une chance de devenir leader de l'intelligence artificielle », Le Monde, le 18 avril 2018.

Cornillie, Chris, « Finding Artificial Intelligence Money in the Fiscal 2020 Budget », Le Monde, le 18 avril 2018.

Delcker, Janosch, « Germany's €3B plan to become an AI powerhouse », Politico, le 14 novembre 2018.

Deschamps, François, « Global Shopping Festival : l'étonnante performance du géant chinois Alibaba », Isa-conso, le 11 novembre 2017, lien : <https://www.lsa-conso.fr/global-shopping-festival-l-etonnante-performance-du-geant-chinois-alibaba,272007>

Domingos, Pedro (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books.

Duval, Jean-Baptiste, « 25 milliards de dollars, 800 millions de colis... Alibaba et les chiffres hallucinants de sa Journée des célibataires », Huffingtonpost, le 14 novembre 2018.

Felardos, Mathis, « La stratégie de la Chine sur l'intelligence artificielle face aux États-Unis », Centre de ressources et d'information sur l'intelligence économique et stratégique, Centre de ressources et d'information sur l'intelligence économique et stratégique, le 24 janvier 2018.

Fischer, Sophie-Charlotte, « Intelligence artificielle : les ambitions de la Chine », Politique de sécurité : analyses du CSS, n°220, février 2018.

Gerlat, Pierre-Yves, « La Chine utilisera l'IA pour développer des systèmes d'aide à la décision pour ses sous-marins nucléaires », actuia.com, le 5 février 2018.

Gettinger, Dan, « Summary of Drone Spending in the FY 2019 Defense Budget Request », Center for Government, le 28 mars 2018.

Guillemoles, Alain, « La Chine fait des concessions aux investisseurs étrangers », La Croix, le 17 mars 2019.

Hagry, Arthur, « L'intelligence artificielle en Chine : le nouveau “Grand Bond” en Avant ? », Asialyst, le 30 novembre 2017.

Hall, Dame Wendy and Jérôme Pesenti (2017), « Growing the Artificial Intelligence Industry in the UK », publié le 15 octobre 2017, lien : [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/652097/Growing\\_the\\_artificial\\_intelligence\\_industry\\_in\\_the\\_UK.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf)

Haski, Pierre, « Intelligence artificielle : qui sera maître du monde ? », Nouvel Observateur, le 17 septembre 2017.

Hellmann, Melissa, « When convenience meets surveillance : AI at the corner store », Seattle Times, le 30 juin 2019.

Kenmann, Henri, « Drone : Chengdu Livre les premiers lots de Wing Loong II aux clients », East Pendulum, le 2 janvier 2018, lien : <http://www.eastpendulum.com/drone-chengdu-livre-premiers-lots-de-wing-loong-ii-aux-clients>

King, Brett (2016). *Augmented Life in the Smart Lane*. Marshall Cavendish International.

Kirsten Korosec, « Baidu just made its 100th autonomous bus ahead of commercial launch in China », le 30 novembre 2017.

Le Vent de la Chine, « L'intelligence artificielle en Chine : dada chinois #2 », Le Petit Journal, le 19 février 2018

Le Vent de la Chine, « L'intelligence artificielle : dada chinois », Le Petit Journal, le 9 février 2018.

Lee, Kai-Fu, *AI Superpowers : China, Silicon Valley and the New World Order*, Houghton Mifflin Harcourt, Septembre 2018.

Leplâtre, Simon, « En Chine, la reconnaissance faciale envahit le quotidien », Le Monde, le 9 décembre 2017.

Mearsheimer, J., « Structural Realism », in Timothy Dunne, Milja Kurki, and Steve Smith (eds.), *International Relations Theories : Discipline and Diversity*, Oxford : Oxford University Press, 2010, pp. 77-94.

Paehlke, Robert C. (2014). « Chapter 5: Global Citizenship Without Global Government ». Dans Robert C. Paehlke, *Hegemony and Global Citizenship: Transnational Governance for the 21st Century*, Palgrave Macmillan, pp. 139-167.

Prévost, Thibault, « Les supermarchés américains testent des portes à reconnaissance faciale », Slate, le 5 juillet 2019.

Rochereuil, Chloé, « La Chine veut prédire les crimes de ses citoyens grâce à une intelligence artificielle », France 24, le 24 juillet 2017.

Samama, Pascal, « Le Pentagone obtient un budget de 10 milliards de dollars pour créer une armée de TechCrunch », le 3 juillet 2018.

Thibout, Charles, « Chine, surveillance 3.0 », Institut des Relations Internationales et Stratégiques, le 18 janvier 2018.

Timmerman, Matthieu, « Pollution massive : la Chine en pleine évolution verte », Atlantico, le 23 mars 2017.

Trujillo, Elsa, « En Chine, le grand bond en avant de la reconnaissance faciale », Le Figaro, le 13 décembre 2017.

Wu, Yimian, « Chinese AI Chip Maker Cambricon Unveils New Cloud-Based Smart Chip », China Money Network, le 4 mai 2018.

Yang, Yuan, « Why Millennials Are Driving Cashless Revolution in China », Financial Times, le 17 juillet 2018.

Zaugg, Julie, « Ping An, le géant chinois de l'intelligence artificielle », Le Temps, le 4 mai 2018, lien : <https://www.letemps.ch/economie/ping-an-géant-chinois-l'intelligence-artificielle>

*Documents officiels*

Executive Office of the President, National Science and Technology Council, Committee on Technology, *Preparing for the Future of Artificial Intelligence*, octobre 2016, lien : [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf)

National Science and Technology Council, Networking and Information Technology Research and Development Subcommittee, *The National Artificial Intelligence Research and Development Strategic Plan*, octobre 2016, lien : [https://www.nitrd.gov/PUBS/national\\_ai\\_rd\\_strategic\\_plan.pdf](https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf)

Executive Office of the President, *Artificial Intelligence, Automation and the Economy*, décembre 2016, lien : <https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF>

« Discours du Président de la République, Emmanuel Macron, sur l'intelligence artificielle », Paris, Collège de France, le 29 mars 2018.

« Rapport de suivi des travailleurs migrants en 2017 », Bureau national des statistiques de Chine, le 4 mai 2018.

Villani, Cédric, « Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne », rapport de mission confiée par le Premier ministre Edouard Philippe, mars 2018, p.125.

*Nationale Strategie für Künstliche Intelligenz AI Made in Germany*, Novembre 2018, Gouvernement fédéral d'Allemagne, lien : [file:///C:/Users/56510/Downloads/Nationale\\_KI-Strategie\\_engl.pdf](file:///C:/Users/56510/Downloads/Nationale_KI-Strategie_engl.pdf).

# Un regard chinois sur la relation Homme-Machine : le cas du jeu de go

**JianMing Zhou, Yu Li**

Dans la Chine ancienne, à l'époque de la dynastie Tang (618-906), le go était l'un des Quatre Arts nécessairement maîtrisés par toute personne cultivée, avec la poésie, la musique et la peinture. Bien que parfois le champ lexical de la guerre et de la bataille soit utilisé pour décrire l'affrontement de deux joueurs, le go n'est pas considéré comme un affrontement entre deux concurrents, ni comme une compétition. Bien au contraire, le go doit permettre aux joueurs de se détacher, d'entrer en transcendance, d'adopter un état d'esprit d'ermite. Il s'agit là d'une caractéristique typique des chinois : dépasser le simple jeu, aller au-delà des règles et de l'enjeu.

Une légende, l'« histoire du manche pourri » (265-420), traduit fort bien cet état d'esprit : Wang Zhi, un bûcheron, alla couper du bois dans la montagne. Il vit deux ermites qui jouaient au go et resta à leurs côtés, observant la partie. Le temps passait et, de temps à autre, il mangeait une datte. Soudain, il entendit crier : « Pourquoi ne rentres-tu pas ? Regarde le manche de ta hache ! ». Il découvrit alors que son manche avait pourri. Cent ans s'étaient écoulés<sup>260</sup>.

Loin de pareil état d'esprit d'ermite, des intelligences artificielles très terre à terre parviennent aujourd'hui à battre les hommes au jeu de go. Cela est lourd d'enseignements, notamment parmi une culture populaire chinoise qui

---

<sup>260</sup> [https://en.wikipedia.org/wiki/Ranka\\_\(legend\)](https://en.wikipedia.org/wiki/Ranka_(legend)) ; Papineau Elisabeth. La culture arrogante du go [Le weiqi, une façon chinoise de voir le monde]. In: Perspectives chinoises, n°62, 2000. pp. 45-56. www.persee.fr/doc/perch\_1021-9013\_2000\_num\_62\_1\_2569.

perçoit cela tel un tremblement de terre. La rencontre du go et de l'IA est, pour la Chine, un véritable révélateur, notamment du point de vue de la relation Homme-Machine.

## I. Une pierre génère mille vagues

En mars 2016, Lee Sedol, un des meilleurs joueurs mondiaux de go, a affronté AlphaGo, un programme informatique, une IA. Après plus de quatre heures de match, AlphaGo a jeté la dernière pierre sur le plateau au 176e coup, et il a battu Lee Sedol sur le score de 3 à 0. Quel choc !<sup>261</sup>

Les promoteurs de l'intelligence artificielle « forte » voient dans cet événement un signe de l'arrivée imminente de la « singularité » dans la société humaine : comme l'intelligence artificielle peut désormais l'emporter sur l'intelligence humaine, il est possible que la machine en vienne à contrôler l'homme et à gouverner le monde à sa place. D'un autre côté, beaucoup de gens croient qu'il existe une différence fondamentale entre la machine et l'homme, car la machine ne possède pas les capacités innées des êtres humains telles que les vrais sentiments, la conscience de soi, le bon sens, la responsabilité sociale etc.

Les philosophes, les anthropologues, les sociologues et les théoriciens de la culture semblent embarrassés, car les concepts les plus fondamentaux tels que l'intelligence, la connaissance, l'émotion et la conscience de soi suscitent des hésitations depuis des milliers d'années ; et maintenant l'intelligence artificielle semble rendre leurs travaux archaïques. De plus, ajoute à la confusion le fait que, parce que le développement de l'intelligence artificielle est difficile à prévoir, tout le monde se veut philosophe et entend participer à la réflexion sur la relation entre l'homme et la machine.

## II. Les particularités d'AlphaGo parmi les IA joueuses

Dans un jeu d'information complet joué par une machine, tous les coups possibles peuvent être représentés sous la forme d'un arbre de recherche de taille  $b^d$  ( $b$  : largeur de recherche ;  $d$  : profondeur de recherche). Pour les jeux d'une complexité élevée, comme les échecs ( $b \approx 35$  ;  $d \approx 80$ ), et en

---

<sup>261</sup> <https://fr.wikipedia.org/wiki/AlphaGo> ; [https://fr.wikipedia.org/wiki/Go\\_\(jeu\)](https://fr.wikipedia.org/wiki/Go_(jeu))

particulier le go ( $b \approx 250$  ;  $d \approx 150$ ), une recherche exhaustive n'est pas possible. Il convient alors de concevoir une fonction de valeur pour évaluer la configuration du jeu de manière à réduire la profondeur et la largeur de l'arbre de recherche<sup>262</sup>.

Les pions du jeu d'échecs ont une nature « individuelle », et les rôles des pions ainsi que leurs positions sur le plateau déterminent la valeur d'une configuration. Il est alors possible de concevoir une fonction « analytique » afin d'évaluer une configuration particulière du jeu d'échecs, et cette méthode a été une réussite dans le cas de la machine Deep Blue qui a battu le champion du monde Garry Kasparov il y a plus de 20 ans. Cependant, le go a pour sa part une nature « collective » : la configuration est déterminée non seulement par les positions des pierres sur le plateau, mais aussi par la relation entre la localité et la globalité de l'ensemble des pierres. Par conséquent, la structure de la solution optimale du go est très compliquée à obtenir, et il est difficile de concevoir une fonction analytique d'évaluation. C'est pourquoi le go est considéré comme le jeu le plus ambitieux pour une intelligence artificielle.

AlphaGo combine des réseaux de neurones (Agent) et un algorithme (Monte-Carlo) afin d'exploiter l'immense et complexe espace de recherche constitué par le jeu de go. Il entraîne deux réseaux de neurones profonds via un apprentissage supervisé et un apprentissage par renforcement avec gros volumes de données (*Big Data*) issues de parties historiques et de parties jouées contre lui-même : le « réseau de valeur » permet d'évaluer la probabilité de réussite d'une configuration ; le « réseau de stratégie » sert à simuler des parties. Guidé par ces deux réseaux, AlphaGo recherche les coups optimaux, les plus efficaces.

Alors que les programmes traditionnels utilisent une fonction d'évaluation « analytique » pour mesurer la valeur d'une configuration donnée, AlphaGo le fait au moyen d'un apprentissage « automatique » basé sur un réseau de neurones.

---

<sup>262</sup> Mastering the game of Go with deep neural networks and tree search, Nature 529 (7587): 484–489 ; Mastering the game of Go without human knowledge, Nature, vol. 550, 19 octobre 2017, p. 354–359.

## Dans toutes ses dimensions

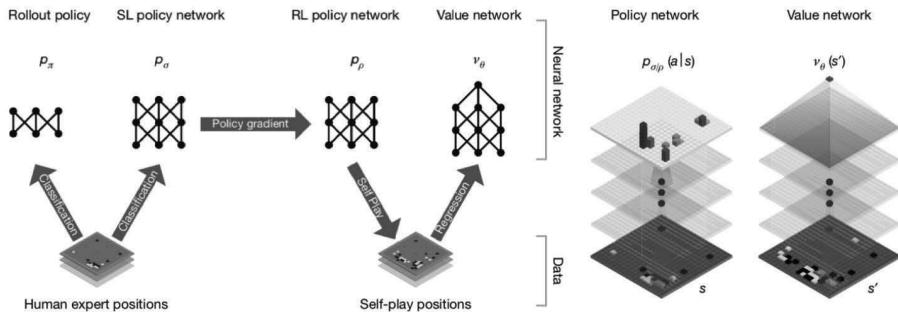


Figure 1<sup>263</sup>

### III. Les décisions d'AlphaGo et de l'Homme en fonction des configurations du go

Quelle est la différence entre les décisions prises par AlphaGo et les jugements des joueurs humains en fonction des diverses configurations du go ? La réponse à cette question peut aider à mieux comprendre la relation entre intelligence humaine et intelligence artificielle.

Le plateau du go est une simple grille orthogonale, un système de coordonnées basé sur le plan géométrique euclidien. L'essence du jeu de go réside donc dans l'utilisation des propriétés de ce plan euclidien. Ensuite, chaque intersection de la grille orthogonale du go possède une signification non géométrique : vide ou plein, libéré ou occupé, mort ou vivant etc., ce qui s'accompagne d'une connotation très riche pour les joueurs de go.

Par rapport à une même configuration, les regards des joueurs humains diffèrent, celle-ci peut présenter des significations très différentes en fonction de chacun. Le go peut refléter l'intuition intellectuelle des individus. C'est pourquoi ce jeu est, dit-on dans le milieu, « facile à apprendre mais difficile à maîtriser » — « dix minutes pour l'apprendre, une vie pour le maîtriser ». Ses exigences élevées en matière d'intuition sont la raison pour laquelle le go est un jeu de haut niveau, exigeant, passionnant.

Par conséquent, les joueurs humains mettent l'accent sur leur état d'esprit au cours du jeu, cela dans le but d'acquérir une compréhension approfondie de son mécanisme. Puisque la configuration du go correspond à la profonde réorganisation de la structure de la grille orthogonale dans l'espace géométrique, chaque coup est une décision relative à l'avenir du jeu, associant

<sup>263</sup> Mastering the game of Go with deep neural networks and tree search, Nature 529 (7587): 484–489.

l'expérience historique et la situation actuelle. C'est pourquoi le joueur doit se concentrer sur son intuition (sa « sensation » du jeu), de sorte que le style de jeu de l'individu puisse être couplé avec une situation non déterministe afin de jouer le meilleur coup. Et alors une nouvelle situation naît en conséquence de ce coup.

En 2017 et 2019, lors de la fête de la science organisée à Amiens, nous avons montré au public comment jouer au go. Nous avons testé deux méthodes d'apprentissage. La première consiste à expliquer les règles du jeu pour ensuite laisser le public jouer. Avec cette méthode, une personne apprenait plus difficilement à jouer, car le positionnement de la première pierre était confus. La seconde méthode consiste à n'expliquer que le but du jeu, puis à laisser le public jouer de façon libre. Plus tard, on lui décrit les règles du jeu au fur et à mesure. Cette dernière méthode permet au public d'apprendre à jouer au go plus vite. Surtout, elle l'amène à s'intéresser davantage au jeu à travers ses propres expériences et une compréhension intuitive du mystère du go. Avec ces expériences d'apprentissage, nous avons compris combien l'intuition du go est intrinsèquement liée à l'intelligence humaine.

Pour leur part, les décisions d'AlphaGo en fonction des configurations du go ne sont pas des traitements bruts de données. En suivant les règles du jeu et sa propre compréhension du go, Deep Mind, le développeur d'AlphaGo, a identifié 48 caractéristiques pour chaque configuration (voir *Extended data Table 2*). Ces caractéristiques concernent principalement le nombre de choix et les changements à chaque intersection. Par conséquent, les entrées des deux réseaux de neurones correspondent à  $19 \times 19 \times 48$  données ; et la sortie est une probabilité de gagner pour chaque position particulière.

**Extended Data Table 2 | Input features for neural networks**

Feature	# of planes	Description
Stone colour	3	Player stone / opponent stone / empty
Ones	1	A constant plane filled with 1
Turns since	8	How many turns since a move was played
Liberties	8	Number of liberties (empty adjacent points)
Capture size	8	How many opponent stones would be captured
Self-atari size	8	How many of own stones would be captured
Liberties after move	8	Number of liberties after this move is played
Ladder capture	1	Whether a move at this point is a successful ladder capture
Ladder escape	1	Whether a move at this point is a successful ladder escape
Sensibleness	1	Whether a move is legal and does not fill its own eyes
Zeros	1	A constant plane filled with 0
Player color	1	Whether current player is black

Feature planes used by the policy network (all but last feature) and value network (all features).

On peut constater, à partir de ces caractéristiques, que l'objectif d'AlphaGo est uniquement de remporter la partie. Il ne poursuit pas de buts esthétiques ni logiques, à l'inverse des humains. Comparées à la richesse et à la « diversité » des jugements intuitifs de l'homme, les décisions d'AlphaGo pour une configuration du jeu donnée sont directes et « monotones ». En ce sens, l'apprentissage automatique d'AlphaGo est « congénital », car donné par l'homme. Il n'a pas la variété, la spontanéité et l'incertitude des jugements des joueurs humains. Il est radicalement différent de l'apprentissage intuitif de l'homme.

« *What magical trick makes us intelligent? The trick is that there is no trick. The power of intelligence stems from our vast diversity, not from any single, perfect principle.* » Marvin Minsky, *The Society of Mind*.

La plus grande différence entre l'apprentissage de l'homme et l'apprentissage automatique réside donc dans le fait que l'homme naît en tant qu'apprenant subjectif, tandis que la machine acquiert sa propre capacité d'apprentissage grâce à une « congénitalité » donnée par l'homme.

## IV. La technique, l'art et le tao du go

L'art du go nécessite une compréhension et une création intuitives de niveau élevé reposant sur la géométrie, et ce genre de créativité s'exerce dans la réciprocité et la compréhension mutuelle des caractères et tempéraments des deux joueurs. En ce sens, le go est parfois appelé « dialogue de mains » (手谈, ShouTan) ou « méditation assise » (坐隐, ZouYin). Le go est une création interactive et abstraite et une communication interne à l'homme.

Le monde du go reconnaît que la personnalité, le tempérament et le caractère du joueur, spécifiquement son éthique et sa culture, sont essentiels du point de vue de sa technique. En général, l'état mental, la maîtrise de la stratégie et la capacité à saisir les opportunités tactiques sont considérés comme des qualités avancées du joueur, appelées « tao du go ». L'art du go repose donc sur les qualités culturelles, intellectuelles et éthiques du joueur. Or il s'agit là d'une dimension forte de la culture et de la pensée chinoises, ce qui était déjà parfaitement expliqué dans un livre de 施定庵 (Shi Dingan, 1710-1770) (弈理指归)<sup>264</sup>.

Le go est devenu le « tao » parce que ses veines sont aussi riches que l'astronomie, sa théorie reflète la doctrine de HeTuo-LuoShu (河图洛书) et le principe du Yin et du Yang, le jugement et le choix, ainsi que leur

---

<sup>264</sup> <https://zh.wikipedia.org/wiki/施襄夏>.

transformation, tout cela est contenu dans le jeu. Le tao du go change à l'infini, il n'existe pas deux parties identiques. Si l'esprit du joueur n'entre pas dans un état harmonieux avec son environnement et ne perçoit pas la logique et le principe profond des choses, il ne pourra pas atteindre un niveau élevé au go.

La machine peut vaincre l'homme parce que la machine n'est pas affectée par les émotions, les sentiments et l'atmosphère de la scène ; elle ne prend en compte que la relation complexe entre le plateau et les pierres, et elle est insensible aux facteurs humains et culturels. Pour la machine, le go est un algorithme, des mathématiques, et on ne peut pas parler d'« art du go ».

Il n'est pas surprenant que les machines soient techniquement supérieures aux hommes, car les outils sont une alternative performante aux capacités limitées des humains. La peur et l'inquiétude concernant la machine ne proviennent pas de sa puissance réelle, potentielle ou fantasmée, mais du fait que nous ne sommes pas parvenus à un consensus sur la compréhension de sa nature et de ses relations avec les humains.

« L'homme n'est qu'un roseau, le plus faible de la nature ; mais c'est un roseau pensant ». La pensée de Pascal n'est pas démodée. La machine n'a pas et n'aura jamais cette essence de l'homme : une faiblesse et une grandeur coexistant en un même individu.

## Conclusion

À partir de cette analyse sommaire du fonctionnement d'AlphaGo par rapport aux jugements de l'homme, on constate que la relation entre l'intelligence artificielle et l'intelligence humaine n'est pas seulement une relation entre l'homme et le monde objectif, mais aussi une relation entre l'homme et la société, une relation entre l'homme et lui-même, c'est-à-dire une relation complexe qui enchevêtre différents niveaux d'analyse. Par conséquent, il faut abandonner la pensée simpliste pour pouvoir explorer véritablement la relation entre l'homme et la machine.

Nous pensons qu'une compréhension fine et profonde de l'intelligence humaine permettrait de cerner plus clairement la relation complexe entre l'intelligence artificielle et l'intelligence humaine. On ne peut pas comprendre l'intelligence humaine du point de vue de l'intelligence artificielle — comme, dans la théorie de la complexité des algorithmes, les problèmes complexes NP

(*Nondeterministic Problems*) ne peuvent être simplement réduits aux problèmes P (*Polynomial time problems*)<sup>265</sup>.

L'expérience du jeu de go et d'AlphaGo nous enseigne que l'intelligence artificielle permet non seulement aux connaissances humaines de se développer, mais aussi à la richesse matérielle et à la civilisation sociale de progresser et à la reconnaissance de notre identité de s'affiner. L'IA — cela devient un lieu commun mais est tellement vrai — est à la fois une opportunité et un défi pour l'humanité. Dans la culture chinoise, le jeu de go en témoigne. AlphaGo a été un véritable électrochoc.

Remerciements :

À M. Boris Barraud pour sa relecture attentive et ses corrections soigneuses.

Aux deux groupes de travail qui ont participé à la fête de la science d'Amiens afin de présenter le jeu de go et AlphaGo :

- En 2019 : Yu Li, Maryline Rosselle, Jean-Baptiste Hoock, Huang Zhi, Nicolas Li, Pierre-Antoine Doutrelot.

- En 2017 : Yu Li, Olivier Baillet, Shixuan Yuan, Yunbo Peng, Shuzhou Wang.

---

<sup>265</sup> JianMing Zhou, Yu Li, Inquiry of P-reduction in Cook's 1971 Paper -from Oracle machine to Turing machine. <http://arxiv.org/abs/1905.06311>.

# L'intelligence artificielle

*Dimension socio-scientifique*



# Quelques repères sur l'apport des sciences sociales pour l'étude de l'intelligence artificielle

**Jean-Sébastien Vayre**

L'intelligence artificielle est une discipline jeune qui est née, à la fin des années 1950, avec le développement de la cybernétique et de l'informatique<sup>266</sup>. Intimement associée aux sciences cognitives<sup>267</sup>, cette discipline est depuis toujours caractérisée par une forte interdisciplinarité. Dès le départ, les spécialistes de l'intelligence artificielle se sont intéressés, de façon variable en fonction de leurs recherches, à l'informatique, aux mathématiques, à la philosophie, à la linguistique, à la psychologie, à la neurologie et/ou à l'anthropologie. Par exemple, Alan M. Turing portait un réel intérêt à la philosophie, comme le montre la qualité des échanges qu'il a pu réaliser avec Ludwig J. J. Wittgenstein<sup>268</sup>. De même, Herbert A. Simon, qui est l'un des fondateurs de l'intelligence artificielle, est largement reconnu pour ses travaux en sciences sociales<sup>269</sup>. Pour autant, il semblerait qu'avec le temps le mouvement de professionnalisation des disciplines scientifiques se soit accompagné d'une division du travail qui a eu pour effet de complexifier et d'invisibiliser progressivement les relations entre intelligence artificielle et sciences sociales. De telle sorte qu'à l'heure actuelle les chercheurs qui étudient l'homme et la société ont souvent du mal à identifier et à comprendre les liens qui peuvent exister entre l'intelligence artificielle et les disciplines qu'ils pratiquent<sup>270</sup>. C'est donc afin de faciliter cette identification et cette

---

<sup>266</sup> Ganascia, 2007.

<sup>267</sup> Andler, 2018.

<sup>268</sup> Cf. Goutefangea, 1999.

<sup>269</sup> Cf. ses recherches sur la rationalité limitée ; Simon, 1955.

<sup>270</sup> Kirtchik, 2019.

compréhension que nous proposons, dans ce texte, de pointer quelques repères sur l'apport des sciences sociales pour l'étude de l'intelligence artificielle. À partir de la lecture de plusieurs travaux issus de la philosophie, de la sociologie et de l'économie, nous dégageons plus exactement quatre formes d'enquête que les sciences sociales peuvent effectuer pour étudier l'intelligence artificielle.

Nous commençons par présenter ce que nous appelons l'enquête épistémologique. Nous verrons que ce type d'enquête consiste à comprendre et/ou à critiquer les fondements et les évolutions théoriques associés à l'hypothèse forte qui est à l'origine de l'intelligence artificielle : celle selon laquelle il serait possible de reproduire matériellement l'intelligence humaine. Dans la deuxième partie, nous ajoutons que cette première perspective peut être prolongée par une autre forme d'enquête, moins abstraite et plus appliquée. Nous qualifierons cette dernière de pragmatique dans la mesure où elle vise moins à comprendre et à critiquer l'hypothèse susmentionnée qu'à pointer les limites des modèles qu'encapsulent les systèmes d'intelligence artificielle et à proposer des alternatives empiriques à ces technologies. Puis nous montrons qu'à ces deux types d'enquête, qui examinent donc l'intelligence artificielle d'un point de vue internaliste, peuvent être associées deux autres formes d'étude qui observent cette même discipline d'un point de vue externaliste. Nous verrons de cette façon que l'enquête interactionniste se distingue des enquêtes épistémologique et pragmatique dans le sens où elle ne cherche pas vraiment, à tout le moins dans un premier temps, à critiquer les fondements théoriques et les développements empiriques de l'intelligence artificielle. Elle a d'abord pour objectif d'étudier les usages individuels et collectifs que les acteurs socio-économiques font des applications de cette science, notamment afin d'identifier et de comprendre les décalages qui peuvent exister entre les usages prescrits par les concepteurs et ceux que réalisent réellement les utilisateurs. Nous ajoutons finalement que ce troisième type d'enquête, qui recouvre *in fine* une perspective micro voire mésosociologique, peut être complétée d'une autre forme d'étude plus macrosociologique. Nous présenterons ainsi comment l'enquête structurelle peut permettre d'approfondir l'enquête interactionniste en cherchant à mieux comprendre comment l'intelligence artificielle, d'une manière générale et par le biais de ses applications socio-économiques, participe à la structuration des rapports sociaux de production et de consommation.

## I. L'enquête épistémologique

Comme nous venons de le signaler, la première forme d'enquête que les sciences sociales peuvent faire sur l'intelligence artificielle est de type épistémologique. Elle est la plus ancienne et la plus classique. Elle regroupe l'ensemble des études qui cherchent à comprendre et à critiquer comment, en tant que discipline scientifique, l'intelligence artificielle et les technologies qu'elle fabrique produisent des connaissances. Ce premier type d'enquête se trouve ainsi à la croisée de la philosophie et de la sociologie des sciences et des techniques. Les travaux de Ludwig J. J. Wittgenstein<sup>271</sup>, de Hubert L. Dreyfus<sup>272</sup> et de John R. Searle<sup>273</sup> en sont le fondement. Rappelons que, d'une manière générale, ces trois philosophes ont consacré une part importante de leurs carrières à réfuter l'hypothèse fondamentale qui est à l'origine de l'intelligence artificielle, c'est-à-dire celle selon laquelle il serait possible de reproduire matériellement l'intelligence humaine. Par exemple, lors des nombreuses discussions qu'il entretient avec son élève Alan M. Turing, Ludwig J. J. Wittgenstein s'oppose à l'idée que la machine que conçoit ce dernier puisse être capable de produire des pensées<sup>274</sup>. Pour Ludwig J. J. Wittgenstein<sup>275</sup>, la pensée est le propre de l'être humain : elle consiste à inventer, à comprendre, à interpréter et à appliquer des règles collectivement partagées ; ce que ne sait pas faire la machine universelle de Turing. Car, en réalité, celle-ci ne fait rien d'autre que matérialiser les règles qu'elle exprime dans un formalisme et un symbolisme particulier. Et c'est précisément pourquoi, à la différence de l'être humain, la machine de Turing est incapable de faire face à des instructions contradictoires.

Un peu plus tard, Hubert L. Dreyfus<sup>276</sup> reprendra à son compte ce type d'argument pour critiquer les simulations cognitives, notamment celles développées par Allen Newell et Herbert A. Simon<sup>277</sup>. Pour cet auteur aussi, l'intelligence humaine ne peut pas être formalisée par le biais d'un système matériel de computation de représentations symboliques dans la mesure où elle dépend de quatre propriétés qui sont le propre des êtres humains. Ces propriétés sont les suivantes. La première est la conscience périphérique, c'est-à-dire cette sorte de *background* qui permet aux hommes de travailler avec des informations non explicitées par le contexte. La deuxième est la

---

<sup>271</sup> Wittgenstein, 1922.

<sup>272</sup> Dreyfus, 1972.

<sup>273</sup> Searle, 1984.

<sup>274</sup> Goutefangea, 1999.

<sup>275</sup> Wittgenstein, 1922.

<sup>276</sup> Dreyfus, 1972.

<sup>277</sup> Cf. Newell & Simon, 1956 ; Newell, Shaw, & Simon, 1959.

tolérance à l'ambiguïté qui met en relation la mémoire et la conscience périphérique. La troisième est la capacité à discriminer ce qui est de l'ordre de l'essentiel et qu'implique la résolution de problèmes complexes. La quatrième est l'association intuitive des éléments pertinents à partir du contexte et que permettent les trois propriétés susmentionnées. Dans une certaine mesure, John R. Searle<sup>278</sup> partage ce point de vue, même s'il comprend et exprime les choses différemment. Pour cet auteur, l'intelligence artificielle comprend l'intelligence humaine comme une sorte de syntaxe, c'est-à-dire un système de règles associant des symboles qui représentent des états de chose. Or l'intelligence humaine comporte également une dimension sémantique : les états de chose sur lesquels portent les états mentaux ont une signification pour les êtres humains. L'auteur soutient alors que cette dimension sémantique peut être saisie à l'aide de quatre singularités qui ne peuvent pas être matérialisées par le biais d'une machine. La première est l'intentionnalité (qui ne doit pas être confondue avec l'intention). C'est en vertu de cette caractéristique que les états mentaux des êtres humains peuvent porter sur leur corps et leur monde. La deuxième est la conscience et est en lien avec l'intentionnalité : elle permet aux êtres humains de connaître leurs états mentaux et de signifier les états de chose sur lesquels ces derniers portent. La troisième est la subjectivité des états mentaux. Elle renvoie à l'idée que les êtres humains ne peuvent accéder aux états mentaux d'autrui pour les décrire objectivement. La quatrième est la causalité du mental. C'est elle qui permet de comprendre comment l'intelligence des humains peut affecter leur corps et leur monde.

Toutefois, l'enquête épistémologique ne se résume pas à pointer les limites philosophiques de l'hypothèse forte qui est au fondement de l'intelligence artificielle. Elle consiste également à mieux comprendre comment cette hypothèse s'exprime concrètement par le biais de conceptions idéelles, matérielles de ce qu'est l'intelligence et qui diffèrent dans l'espace et dans le temps. Précisons alors qu'une telle façon de concevoir l'enquête épistémologique n'est en rien nouvelle puisqu'il y a plus de trente ans, Steve Woolgar<sup>279</sup> soutenait déjà qu'il était possible d'étudier l'intelligence artificielle, non plus du point de vue de la philosophie analytique ou de la phénoménologie, mais de celui de la sociologie des sciences et des techniques. C'est ce que font, par exemple, des auteurs comme Mikel Olazaran<sup>280</sup>, ou encore, Dominique Cardon, Jean-Philippe Cointet et Antoine Mazières<sup>281</sup>. Ces chercheurs nous rappellent en effet que l'histoire de l'intelligence artificielle se structure autour de deux grands courants dits cognitiviste (ou encore symbolique) et connexionniste et qui recouvrent deux façons différentes de

---

<sup>278</sup> Searle, 1984.

<sup>279</sup> Woolgar, 1985.

<sup>280</sup> Olazaran, 1996.

<sup>281</sup> Cardon, Cointet, & Mazières, 2018.

poser et de tester l'hypothèse de la matérialisation de l'esprit humain. Ces auteurs présentent alors comment le succès de l'un de ces courants par rapport à l'autre varie en fonction de la dynamique de l'histoire de l'intelligence artificielle. Un peu comme dans le cas de la sociologie<sup>282</sup>, il existe un balancier de l'intelligence artificielle qui oppose les pôles cognitiviste et connexionniste et dont les penchants évoluent avec la dynamique de la société<sup>283</sup>. Les auteurs précités montrent ainsi qu'en tant que discipline scientifique, l'intelligence artificielle fabrique des machines qui encapsulent les visions que leurs concepteurs se font de ce qu'est l'intelligence. Elle participe par là même à faire exister des philosophies de l'esprit particulières dont les succès sont contingents. Concernant ce dernier point, les travaux de Andrew Pickering<sup>284</sup> sont d'un intérêt particulier, notamment parce qu'ils viennent compléter ceux dont nous venons de parler en y ajoutant une dimension critique. Plus précisément, cet auteur pointe comment les technologies de l'intelligence artificielle recouvrent généralement une vision dualiste et cartésienne selon laquelle les humains sont différents des choses par nature. Il ajoute que cette vision participe à l'institution d'un *telos* du désengagement qui désincarne les humains de leur environnement. En conférant aux machines le soin de prendre en charge leurs activités routinières de façon à leur permettre de se concentrer sur la réflexion et la supervision, les humains se détachent d'une part importante des relations concrètes qu'ils entretiennent avec le monde.

Notons qu'il existe évidemment d'autres types d'enquête épistémologique portant sur l'intelligence artificielle et dont nous ne pouvons pas rendre compte dans cette contribution : cette première forme d'enquête regroupe des perspectives très diverses. Aussi, pour conclure cette première partie, nous souhaitons insister sur le fait que ces perspectives comportent plusieurs contradictions. Harry Collins<sup>285</sup> parle, par exemple, de tension pour exposer comment le principe de symétrie<sup>286</sup>, qui est au cœur de la sociologie des sciences, peut poser problème pour étudier le cas de l'intelligence artificielle. Pour cet auteur, à l'instar des philosophes que sont Ludwig J. J. Wittgenstein<sup>287</sup>, Hubert L. Dreyfus<sup>288</sup> ou encore John R. Searle<sup>289</sup>, le sociologue à quelque chose à dire sur ce qu'est la connaissance. Par conséquent, celui-ci ne peut pas se cantonner à examiner comment l'intelligence artificielle définit l'intelligence : de manière non conforme au

---

<sup>282</sup> Cf. Cuin, 2002.

<sup>283</sup> Vayre, 2016.

<sup>284</sup> Pickering, 2019.

<sup>285</sup> Collins, 2019.

<sup>286</sup> Bloor, 1976.

<sup>287</sup> Wittgenstein, 1922.

<sup>288</sup> Dreyfus, 1972.

<sup>289</sup> Searle, 1984.

principe de symétrie, le sociologue doit également pointer les limites de cette vision au regard de la science qu'il pratique. Partant, l'intérêt des études menées par Harry Collins<sup>290</sup> est d'illustrer comment il est possible de dépasser les contradictions que le chercheur peut éprouver durant l'enquête épistémologique. Il montre plus exactement que, malgré les turbulences que cela peut produire, le chercheur en sciences sociales peut se positionner à équidistance de la philosophie et de la sociologie des sciences et des techniques pour, d'un côté, examiner la manière dont l'intelligence artificielle saisit l'intelligence et, d'un autre côté, pointer les limites de cette même compréhension.

## II. L'enquête pragmatique

Les sciences sociales ne sont cependant pas contraintes d'étudier l'intelligence artificielle selon la seule perspective que constitue l'enquête épistémologique. Elles peuvent aussi le faire d'une façon plus pragmatique. Comme nous l'avons vu, l'enquête épistémologique vise surtout à saisir et à questionner comment l'intelligence artificielle pose l'hypothèse fondamentale selon laquelle il serait possible de matérialiser, par le biais d'une machine, l'intelligence humaine. En référence aux travaux de John R. Searle<sup>291</sup>, l'enquête épistémologique est plutôt orientée vers la compréhension et la critique de l'intelligence artificielle forte ; et cela que ce soit du point de vue des réflexions théoriques que peuvent mener les philosophes ou de celui des études plus empiriques que peuvent réaliser les sociologues des sciences et des techniques.

Ce que nous appelons l'enquête pragmatique renvoie à une autre perspective, moins abstraite, et qui peut être comprise comme une continuité de l'enquête épistémologique. Plus exactement, l'enquête pragmatique recouvre une posture interventionniste qui a pour objet de participer à définir l'orientation de la dynamique de l'intelligence artificielle : l'enquête pragmatique consiste à proposer des manières plus ou moins alternatives de concevoir des machines capables de simuler une voire plusieurs tâches que peut effectuer l'intelligence humaine. Pour reprendre les termes de John R. Searle<sup>292</sup>, cette deuxième forme d'enquête concerne donc moins la version forte que la version faible de l'intelligence artificielle. Autrement dit, si nous qualifions cette enquête de « pragmatique », c'est parce qu'elle vise moins à

---

<sup>290</sup> Collins, 2019.

<sup>291</sup> Searle, 1984.

<sup>292</sup> Searle, 1984.

traiter les problèmes de fond que pose l'intelligence artificielle qu'à critiquer les réalisations pratiques que fabrique cette discipline en proposant des alternatives contribuant concrètement à son développement. Le cas de Francis Chateauraynaud<sup>293</sup> est, à tout le moins en France, un des plus représentatifs de l'enquête pragmatique. Cela fait en effet plus de trente ans que ce sociologue s'intéresse activement à l'intelligence artificielle. Cet intérêt lui a permis d'élaborer, avec son collègue Jean-Pierre Charriaud, un Programme de Sociologie Pragmatique, Expérimentale et Réflexive sur Ordinateur (PROSPERO). Ce programme constitue une sorte de contre intelligence artificielle dans la mesure où il recouvre une vision alternative à celles qui dominent au sein de cette discipline. Car PROSPERO n'est pas un système expert (qui renvoie au courant cognitiviste de l'intelligence artificielle), ni un réseau de neurone (qui renvoie au courant connexionniste<sup>294</sup>). C'est une technologie littéraire conçue pour assister le travail d'analyse des données textuelles que les chercheurs en sciences sociales peuvent effectuer pour décrire les polémiques et les controverses médiatiques<sup>295</sup>. Pour autant, si les travaux de Francis Chateauraynaud<sup>296</sup> sont un exemple typique des formes que peut prendre l'enquête pragmatique, cette dernière ne s'y réduit pas. L'histoire de l'intelligence artificielle nous apprend que l'enquête pragmatique est bien plus ancienne. Si, comme le montre Hubert Dreyfus<sup>297</sup>, l'œuvre de Herbert A. Simon est critiquable, il n'en reste pas moins que les études que ce dernier a menées dans le domaine des sciences sociales<sup>298</sup> lui ont permis de concevoir, avec son collègue Allen Newell, ses fameux *Logic Theorist*<sup>299</sup> et *Global Problem Solver*<sup>300</sup>. Il ne faut effectivement pas oublier que, si cet auteur a pu recevoir le prix Turing en 1975 et le prix Nobel d'économie en 1978, c'est précisément parce que ses travaux sur la rationalité limitée l'ont aidé à développer, avec Allen Newell, une manière alternative (à tout le moins pour l'époque) de concevoir les machines intelligentes.

Plus exactement, l'histoire de l'intelligence artificielle montre que ce que nous appelons l'enquête pragmatique est consubstantiel à cette discipline<sup>301</sup>. Dès le départ, les pionniers de l'intelligence artificielle ont cherché à faire avancer cette discipline en proposant des manières plus ou moins alternatives de concevoir leurs technologies. Et, pour ce faire, ils ont généralement été

---

<sup>293</sup> Chateauraynaud, 2019.

<sup>294</sup> Cf. Vayre, 2016.

<sup>295</sup> Chateauraynaud, 2003

<sup>296</sup> Chateauraynaud, 2019.

<sup>297</sup> Dreyfus, 1972.

<sup>298</sup> Simon, 1955.

<sup>299</sup> Cf. Newell, & Simon, 1956.

<sup>300</sup> Cf. Newell, Shaw, & Simon, 1959.

<sup>301</sup> Vayre, 2016.

conduits à réaliser deux types d'activités : l'observation empirique du comportement intelligent et sa modélisation. La phase d'observation empirique implique, la plupart du temps, la mobilisation d'outils méthodologiques, conceptuels et/ou théoriques issus de la neurologie, de la psychologie, de la linguistique, mais aussi, comme le montrent les travaux de Herbert A. Simon, des sciences sociales. Nous souhaitons alors insister sur le fait que cet auteur n'est pas le seul classique, en intelligence artificielle, à avoir fait usage des sciences sociales pour concevoir ses machines. Robert Wilensky<sup>302</sup> a, par exemple, fait de même. Rappelons que pour cet auteur, et à la différence de beaucoup de ses contemporains informaticiens, les recherches en linguistique de Noam Chomski<sup>303</sup> n'apparaissent pas très pertinentes pour comprendre et modéliser le comportement intelligent. Il en va par contre différemment de celles qu'effectuent ses collègues anthropologues sur la dimension culturelle des échanges langagiers. Pour Robert Wilensky, ces études sont d'une grande utilité pour concevoir son fameux *Plan Applier Mechanism* (PAM)<sup>304</sup>. Car, si Robert Wilensky<sup>305</sup> s'intéresse au langage, c'est moins en tant que compétence linguistique – c'est-à-dire que système symbolique formel – que performance linguistique – c'est-à-dire qu'outils de communication de sens.

Pour autant, si l'anthropologie culturelle a eu une influence sur l'intelligence artificielle, force est de constater que, du point de vue de l'enquête pragmatique, l'impact des sciences sociales sur cette discipline a d'abord été réalisé par le biais de l'économie, notamment à travers l'immense succès qu'a connu la théorie des jeux<sup>306</sup>. Il n'en reste pas moins que, de façon certes plus marginale, la sociologie a également joué un rôle qu'il ne faut pas négliger. À la fin des années 1980, Edward E. Brent et ses collègues<sup>307</sup> ont par exemple développé ERVING, une intelligence artificielle visant à modéliser l'approche sociologique de Erving Goffman de manière à pouvoir tester, sous la forme d'un jeu sérieux, la compréhension qu'en ont ses utilisateurs. Au même moment, Sanjoy Banerjee<sup>308</sup> élaboré, à l'aide du langage PROLOG, une intelligence artificielle capable de simuler les théories sociopolitiques de Theda Skocpol et de Guillermo A. O'Donnell. Cette technologie permet plus exactement d'opérationnaliser la notion de schème proposée par Jean Piaget<sup>309</sup>. Elle constitue une façon de modéliser la dimension culturelle de

---

<sup>302</sup> Wilensky, 1983.

<sup>303</sup> Chomski, 1957.

<sup>304</sup> Wilensky, 1977.

<sup>305</sup> Wilensky, 1983.

<sup>306</sup> Garson, 1987.

<sup>307</sup> Brent, 1989.

<sup>308</sup> Banerjee, 1986.

<sup>309</sup> Piaget, 1967.

l'intelligence comme un système trans-individuel de schèmes qui guident l'action en société. Cette intelligence artificielle est, par là même, une manière de représenter comment l'intelligence humaine est façonnée par le poids des structures sociales.

En outre, le dialogue entre sociologie et intelligence artificielle a plus récemment été favorisé par le développement des sous-domaines de l'intelligence artificielle que forment les recherches sur les systèmes multi-agent (*multi-agent system* ; MAS) et sur l'intelligence artificielle distribuée (*distributed artificial intelligence* ; DAI)<sup>310</sup>. Ajoutons alors qu'à l'heure actuelle, avec la large diffusion des technologies d'apprentissage artificiel (*machine learning*) permettant de traiter les mégadonnées (*big data*), plusieurs formes de collaboration entre intelligence artificielle et sciences sociales ont vu le jour, notamment dans le but de favoriser un développement éthique des machines intelligentes (les spécialistes parlent souvent de machine morale — *moral machine*). Concernant ce dernier point, les travaux de Fiona Berredy, de Gauvain Bourgne et de Jean-Gabriel Ganascia<sup>311</sup> sur la conception d'agents autonomes éthiques, ou encore ceux de Iyad Rahwan<sup>312</sup> sur la modélisation algorithmique du contrat social, constituent des initiatives intéressantes.

### III. L'enquête interactionniste

À la différence des enquêtes épistémologique et pragmatique, l'enquête interactionniste ne consiste pas à étudier, de façon plus ou moins philosophique ou participative, la manière dont l'intelligence artificielle est conçue. Il ne s'agit pas tellement, ici, de comprendre et de critiquer les philosophies de l'esprit qui fondent les grands paradigmes de l'intelligence artificielle ni de chercher à impulser des dynamiques plus ou moins alternatives aux développements technologiques que réalise cette discipline. L'enquête interactionniste a plutôt pour finalité de décrire et de saisir la façon dont les applications concrètes de l'intelligence artificielle sont implémentées au sein des organisations publiques ou privées et sont utilisées par les travailleurs, les consommateurs ou les citoyens. Bien sûr, ces applications encapsulent immanquablement des représentations particulières de l'intelligence, qu'elle soit comprise comme forte ou faible. C'est pourquoi, quelles que soient les perspectives adoptées, les enquêtes épistémologique et pragmatique ne sont jamais inutiles. Pour autant, ces applications cristallisent

---

<sup>310</sup> Cf. Schillo, Fischer, & Klein, 2000.

<sup>311</sup> Ganascia, 2018.

<sup>312</sup> Rahwan, 2018.

également des formes relationnelles particulières dans la mesure où elles sont le nœud d'un réseau d'interactions qui associe les acteurs qui les conçoivent, ceux qui les implémentent et ceux qui les utilisent<sup>313</sup>. L'objectif de l'enquête interactionniste est alors de comprendre comment est fait ce nœud.

Afin de bien comprendre cette troisième forme d'enquête, il est possible de l'aborder selon deux grandes perspectives. La première se focalise sur l'observation de l'environnement de conception, c'est-à-dire celui que compose l'ensemble des relations sociomatérielles qui se jouent entre les acteurs de la conception et ceux de l'implémentation. La seconde se centre sur l'examen de l'environnement d'usage, c'est-à-dire celui que forme l'ensemble des interactions sociomatérielles qui associent les acteurs de l'implémentation et ceux de l'utilisation. Les travaux de Lucy A. Suchman<sup>314</sup>, qui s'inscrivent dans le champ de l'ethnométhodologie, sont une forme classique de l'enquête interactionniste. Rappelons que, dans ses études, l'auteur propose la notion d'action située pour pointer les limites des plans ou, si l'on préfère, des scénarii d'usages que véhiculent les technologies d'intelligence artificielle de type symbolique. Pour ce faire, elle examine le cas d'un système expert d'aide à la photocopie. Cette étude lui permet de montrer le décalage qui existe entre les scénarii qui sont inscrits dans cette intelligence artificielle (*i.e.* les usages prescrits) et l'utilisation qu'en font concrètement ses utilisateurs (*i.e.* les usages réels). Un des principaux arguments de l'auteur est alors que les utilisateurs du système expert n'ont pas accès aux plans qui déterminent son fonctionnement. Ils doivent par conséquent reconstituer ces plans au cours de l'utilisation qu'ils font de la machine intelligente pour comprendre le sens des indications qu'elle leur communique et s'aligner sur son fonctionnement. L'auteur soutient au final qu'il est possible de faciliter ce travail d'alignement en mieux considérant l'ancre social et matériel des interactions hommes-machines. Et c'est précisément en ce sens que Lucy A. Suchman<sup>315</sup> défend sa théorie de l'action située.

À l'heure actuelle, plusieurs auteurs, comme par exemple, Julia Velkovska et Valérie Beaudouin<sup>316</sup>, proposent de continuer cette tradition de recherche. Dans la lignée des travaux de Lucy A. Suchman<sup>317</sup>, ceux que ces deux auteurs ont réalisé sur les services vocaux automatisés montrent que les usages concrets des technologies d'intelligence artificielle produisent de multiples désalignements. Pour éviter ces derniers, les utilisateurs doivent se détacher des formes de conversation ordinaire afin de se conformer au modèle standardisé qui régit le fonctionnement de la machine. Aussi, bien que ces

---

<sup>313</sup> Cf. Vayre, 2016.

<sup>314</sup> Suchman, 1987.

<sup>315</sup> Suchman, 1987.

<sup>316</sup> Velkovska, & Beaudouin, 2014.

<sup>317</sup> Suchman, 1987.

études d'inspiration ethnométhodologique soient d'une grande pertinence pour observer et comprendre les interactions homme-machine, ce que nous appelons l'enquête interactionniste ne s'y réduit pas. Ici aussi, il existe une importante diversité de façons d'examiner et d'interpréter les problèmes organisationnels et humains que posent l'implémentation et l'usage des technologies d'intelligence artificielle.

Par exemple, les études que Michel Freyssenet<sup>318</sup> a réalisées sur l'intelligence artificielle se distinguent nettement de l'ethnométhodologie. Plus directement insérées dans le domaine de la sociologie du travail, ces recherches n'en restent pas moins intéressantes. Elles montrent que, du point de vue de l'environnement de conception, les systèmes experts sont souvent implantés selon un double objectif. Ils doivent, dans un premier temps, assister la prise de décision de leurs utilisateurs. Ils sont par là même supposés, dans un second temps, leur apporter des connaissances sur un domaine d'expertise spécifique. Cependant, du point de vue de l'environnement d'usage, ces machines intelligentes ne remplissent pas correctement ces fonctions d'aide et de formation. Elles permettent en revanche de traiter rapidement et de manière homogène des cas répertoriés. Les systèmes experts sont de cette manière dotés d'un potentiel substitutif que ressentent et craignent les agents qualifiés. Pour autant, Michel Freyssenet<sup>319</sup> pense que ces technologies pourraient réellement effectuer leurs fonctions d'aide et de formation en autorisant l'exploration, par leurs utilisateurs, des chaînes de causalité qu'elles expriment afin de leur permettre de remonter aux causes premières des problèmes à traiter. L'auteur pense que l'intelligence artificielle pourrait de ce fait contribuer à inverser réellement et durablement la division intellectuelle du travail.

L'enquête interactionniste peut faire l'objet de bien d'autres perspectives encore. L'anthropologie des techniques que pratiquent Gérard Dubey<sup>320</sup> ou Alain Gras<sup>321</sup> est tout aussi pertinente que la sociologie du travail de Michel Freyssenet<sup>322</sup> pour étudier les interactions qu'entretiennent les hommes et les machines intelligentes. Les études que ces auteurs ont effectuées sur le développement des pilotes automatiques dans le domaine de l'aviation civile (les *Flight Management Systems* ; FMS) pointent comment l'intelligence artificielle s'immisce toujours un peu plus entre les sens qui permettent aux humains d'appréhender le monde, les connaissances qu'ils développent à son sujet et les actions qu'ils déploient pour agir en son sein, voire le modeler. Ainsi, en véhiculant une représentation cognitiviste de l'intelligence, les

---

<sup>318</sup> Freyssenet, 1990.

<sup>319</sup> Freyssenet, 1990.

<sup>320</sup> Dubey, 2014.

<sup>321</sup> Gras, 1994.

<sup>322</sup> Freyssenet, 1990.

systèmes experts dépossèdent leurs utilisateurs d'une partie des relations qui les attachent à leur environnement et participent à instituer, de manière plus générale, un rapport formel entre le monde et les humains qui l'habitent.

En outre, et de manière moins orthodoxe, les sciences sociales peuvent même, dans certain cas, s'associer à l'ergonomie cognitive, ou encore à la linguistique informatique, pour mettre en avant les limites de l'intelligence artificielle. Par exemple, avec l'avènement des big data et l'augmentation des puissances de calcul informatique, de nombreuses entreprises élaborent des machines intelligentes qui ont pour fonction de générer automatiquement des contenus littéraires sur des sujets qui peuvent être variés. Or, si ces technologies tendent à produire des textes faciles à lire compte tenu de leur caractère standardisé, ils manquent souvent d'une cohérence socio-sémantique qui ne facilite pas leur compréhension<sup>323</sup>. De même, toujours suite au développement des mégadonnées, les acteurs économiques ont cherché à concevoir des machines capables de personnaliser les environnements numériques marchands. Le problème est alors que ces technologies, au moins dans certains cas, cadrent les activités de navigation des consommateurs en favorisant leurs stratégies d'exploration au détriment de celles de recherche d'information. Ces machines intelligentes peuvent en ce sens avoir pour effet de désengager les consommateurs des buts marchands qu'ils se sont donnés de manière à favoriser les finalités des commerçants<sup>324</sup>. Elles peuvent par là même instituer une asymétrie commerciale qui n'est pas un bien pour la communauté.

## IV. L'enquête structurelle

En focalisant l'attention sur les interactions qui se font entre les technologies d'intelligence artificielle, ceux qui les conçoivent, les implémentent et les utilisent, l'enquête interactionniste a l'avantage de rendre compte des rôles relationnels que ces machines performent au niveau micro voire mésosociologique. Pour autant, ce type d'enquête peine à renseigner correctement sur le rôle structurel que remplissent, sur le plan macrosociologique, les applications concrètes de l'intelligence artificielle. Le type d'enquête que nous présentons dans cette dernière partie a pour objectif de combler ce manque. C'est donc en ce sens que nous parlons d'enquête structurelle. D'une manière générale, les auteurs qui effectuent cette quatrième forme d'enquête considèrent l'intelligence artificielle comme

---

<sup>323</sup> Vayre, Delpech, Dufresne, & Lemercier, 2017.

<sup>324</sup> Vayre, Larnaudie, Dufresne, & Lemercier, 2016.

l'aboutissement du processus d'automatisation qui se développe, depuis la révolution industrielle, au sein des sociétés occidentales : les machines intelligentes sont des dispositifs autorisant la réalisation informatique de tâches que l'on croyait, jusque-là, non automatisables compte tenu du travail intellectuel qu'elles nécessitent (*e.g.* l'évaluation d'une radiographie, ou encore le pilotage d'un véhicule).

De ce point de vue, la diffusion des technologies d'intelligence artificielle au sein de la société soulève une question fondamentale sur les plans économique et social : comment ces machines participent-elles à la structuration des rapports sociaux de production et de consommation ? En référence aux travaux de John M. Keynes<sup>325</sup> sur le chômage technologique, certains économistes pensent que les technologies d'intelligence artificielle pourraient avoir pour effet de détériorer la structure de l'emploi. Carl B. Frey et Michael A. Osborne<sup>326</sup> estiment par exemple que 47 % des emplois hautement qualifiés, 19 % des emplois moyennement qualifiés et 33 % des emplois non qualifiés seraient aujourd'hui menacés par l'intelligence artificielle. Cette estimation fait cependant l'objet de nombreuses controverses, notamment parce qu'elle ne tient pas compte du coût des applications de l'intelligence artificielle, des effets de leur menace sur la structure des salaires, mais aussi de la création d'emplois complémentaires qui peuvent émaner de leur implémentation au sein des organisations.

En réponse à ces controverses, d'autres auteurs ont cherché à mieux comprendre comment, par le biais des applications socio-économiques qu'elle développe, l'intelligence artificielle a, jusqu'ici, contribué à structurer l'emploi. C'est le cas de Daron Acemoglu et de Pascual Restrepo<sup>327</sup> qui ont examiné l'influence, durant ces vingt dernières années, des applications d'intelligence artificielle que composent les robots industriels sur le marché du travail états-unien. Les auteurs montrent alors que ces applications ont effectivement un impact négatif sur l'emploi et les salaires. Ils insistent toutefois sur le fait que la magnitude de cet impact est raisonnable au regard des grandes révolutions industrielles passées. En outre, les auteurs soutiennent que cet effet serait plus marqué dans le cas des hommes que dans celui des femmes, mais aussi qu'il diminuerait avec la hausse du niveau d'étude. En conclusion, Daron Acemoglu et Pascual Restrepo<sup>328</sup> nuancent l'estimation de Carl B. Frey et de Michael A. Osborne<sup>329</sup> en soutenant l'idée suivante : si, par le biais des robots industriels, l'intelligence artificielle a eu un impact négatif sur l'emploi, c'est parce que ces machines ont traditionnellement été pensées

---

<sup>325</sup> Keynes, 1931 [1933].

<sup>326</sup> Frey, & Osborne, 2013.

<sup>327</sup> Acemoglu, & Restrepo, 2019.

<sup>328</sup> Acemoglu, & Restrepo, 2019.

<sup>329</sup> Frey, & Osborne, 2013.

comme des substituts à la force humaine de travail. Or, dans l'avenir, il n'est pas impossible qu'avec le développement de l'IoT les robots prennent en charge des tâches complémentaires à celles effectuées par les humains pour finalement impacter positivement l'emploi.

Les travaux des sociologues qui se sont intéressés au processus d'automatisation des activités de travail et de consommation permettent de compléter ceux de leurs collègues économistes. Pour eux, le problème n'est pas tellement que les technologies d'intelligence artificielle sont susceptibles de favoriser le chômage technologique, mais plutôt qu'elles instituent des modifications des activités socio-économiques qui sont problématiques. Comme le montrent les études menées par Antonio Cassili<sup>330</sup>, les nouveaux systèmes intelligents qu'élaborent les grands acteurs du numérique encouragent, par exemple, l'avènement d'une forme inédite de tâcheronisation des activités productives. Afin de pouvoir apprendre à effectuer les activités que les humains souhaitent qu'elles exécutent, ces systèmes doivent disposer d'immenses bases de données conçues à cet effet. Et la conception de ces bases recouvre un gigantesque travail collectif de labellisation d'images et de photographies, de transcription de textes courts, ou encore d'enregistrement de données audios et vidéos. Ces activités peuvent composer deux types de travail. Le premier est qualifié de micro-travail et fait l'objet de micro-rémunérations. Il est caractérisé par une grande flexibilité, mais aussi un haut niveau de précarité. Le deuxième est gratuit et tend à être invisibilisé. Il constitue une nouvelle forme de « mise au travail des clients »<sup>331</sup>. Partant, en renouvelant les sources de création de la plus-value et la division du travail intellectuel, les applications actuelles de l'intelligence artificielle contribuent, plus généralement, à l'institution de nouveaux rapports sociaux entre les propriétaires des appareils de production et de distribution des biens, les responsables qui les conçoivent, les contrôlent, les dirigent, les travailleurs qui assurent leur fonctionnement, mais aussi les consommateurs sans qui ces appareils n'auraient pas lieu d'exister.

En résumé, les auteurs cités dans cette dernière partie insistent sur le fait que, dans la droite ligne de l'histoire des processus d'automatisation industrielle, un des plus importants problèmes des technologies d'intelligence artificielle est que, à l'heure actuelle encore, elles sont souvent conçues comme des substituts au travail humain. Pour ces auteurs, le rôle structurel de l'intelligence artificielle correspond par là même, à tout le moins à l'heure actuelle et de manière générale, à ce que Yann Ferguson<sup>332</sup> appelle le « scénario de l'homme remplacé ». L'avantage des travaux de ce dernier est alors de souligner que ce rôle n'est pas le seul possible. Les humains ont en

---

<sup>330</sup> Cassili, 2019.

<sup>331</sup> Tiffon, 2013.

<sup>332</sup> Ferguson, dans ce volume.

réalité toujours associé aux technologies d'intelligence artificielle une multitude de fonctions qui peuvent potentiellement favoriser l'avènement de divers rapports sociaux entre les hommes et les machines. Yann Ferguson<sup>333</sup> dégage plus précisément cinq scénarios possibles. Le premier est celui de l'homme remplacé dont nous venons de parler. Le second est celui de l'homme augmenté, dans lequel les machines intelligentes et les humains forment une collaboration capacitante : les premières s'occupent de ce qui est de l'ordre de la technique, tandis que les seconds prennent en charge les aspects plus analytiques et relationnels. Le troisième, qui est probablement le plus dystopique, est celui de l'homme dominé. Ce scénario pointe, non pas tellement le risque de retournement des machines contre les humains, mais celui, à la fois plus réaliste et présent, du développement du pouvoir d'injonction que véhiculent ces technologies. Ce pouvoir est effectivement porteur de nombreux risques comme, par exemple, celui de la déqualification des travailleurs, de leur prolétarisation et de leur déresponsabilisation. Le quatrième est celui de l'homme divisé. Il correspond à l'idée que la large diffusion des applications de l'intelligence artificielle institue de nouvelles formes d'inégalité entre ceux qui détiennent ou non les capitaux culturel et économique leur permettant de les utiliser à bon escient. Le cinquième, qui est le plus utopique, est celui de l'homme réhumanisé. Dans ce scénario, les technologies d'intelligence artificielle peuvent permettre aux hommes de se détacher de la dimension machinique des actions qu'ils effectuent afin de favoriser le bon déploiement des singularités qui constituent le propre de l'humanité.

Selon nous, l'intérêt des études menées par Yann Ferguson<sup>334</sup> est ainsi de pointer comment l'intelligence artificielle recouvre, non pas un, mais plusieurs rôles sociaux et économiques. Et, à l'instar du cas du scénario de l'homme remplacé dont nous avons parlé plus haut, chacun de ces rôles peut *in fine* faire l'objet d'une enquête structurelle visant à mettre à l'épreuve la validité empirique des scénarios qui y sont associés.

## Conclusion

Nous souhaitons insister sur le fait que les quatre types d'enquête que nous venons de présenter doivent être discutés, notamment parce qu'ils reposent sur une revue de littérature qui n'est pas exhaustive. En outre, il convient de souligner que ces formes d'enquête sont des représentations abstraites de la

---

<sup>333</sup> Ferguson, dans ce volume.

<sup>334</sup> Ferguson, dans ce volume.

réalité : dans les faits, la plupart des auteurs précités articulent les enquêtes épistémologique, pragmatique, interactionniste et/ou structurelle en fonction de leurs intérêts de recherche. Par exemple, les critiques que Lucy A. Suchman<sup>335</sup> formule à l'égard de l'intelligence artificielle dépassent le périmètre de l'enquête interactionniste pour rejoindre les enquêtes pragmatique et épistémologique.

En outre, il nous semble important de conclure en mettant en avant la nécessité de renouveler la compréhension que les sciences sociales se font de l'intelligence artificielle. Car il n'est pas rare que les critiques que ces dernières expriment à l'égard de l'intelligence artificielle reprennent les arguments de Ludwig J. J. Wittgenstein<sup>336</sup>, de Hubert L. Dreyfus<sup>337</sup> et de John R. Searle<sup>338</sup>. Ces auteurs sont évidemment importants et il n'est pas question de remettre en cause la pertinence de leurs études. Seulement, comme nous l'avons signalé plus haut, ces études constituent moins un ensemble de critiques à l'égard de l'intelligence artificielle en général qu'une fervente mise en question de l'intelligence artificielle symbolique et, surtout, du paradigme cognitiviste qui en est sous-jacent. Rappelons, par exemple, que si Hubert L. Dreyfus s'opposait vivement à l'intelligence artificielle de type cognitiviste, il manifestait également un certain penchant pour le paradigme connexionniste<sup>339</sup>. Or, aujourd'hui, le courant connexionniste de l'intelligence artificielle est devenu dominant. De notre point de vue, ce renversement de situation est intéressant dans la mesure où il constitue une manifestation de la manière dont, un peu à la façon du capitalisme, l'intelligence artificielle a absorbé des résistances qui raisonnent avec la « critique artiste »<sup>340</sup>.

Aussi, si nous pensons nécessaire de rénover les critiques de l'intelligence artificielle, c'est parce que le développement de cette discipline, toujours en référence au nouvel esprit du capitalisme de Luc Boltanski et de Ève Chiapello<sup>341</sup>, doit être contenu dans la critique : cette dernière doit en effet permettre d'éviter des dérives scientifiques et techniques qui pourraient être catastrophiques pour la société. Nous estimons par conséquent que les sciences sociales, compte tenu de leurs domaines d'expertise, ont un rôle important à jouer dans le renouvellement des critiques de l'intelligence artificielle. Et si le champ des possibles est, à ce niveau, très ouvert, il nous apparaît qu'une bonne façon de réinventer les critiques de l'intelligence artificielle serait de chercher à mieux comprendre la réalisation et les

---

<sup>335</sup> Suchman, 1987.

<sup>336</sup> Wittgenstein, 1922.

<sup>337</sup> Dreyfus, 1972.

<sup>338</sup> Searle, 1984.

<sup>339</sup> Cf. Dreyfus, 1992.

<sup>340</sup> Boltanski, & Chiapello, 1999 ; Vayre, 2019.

<sup>341</sup> Boltanski, Chiapello, 1999.

implications du renversement de situation dont nous venons de parler du point de vue des quatre types d'enquête présentés.

## Références

- Acemoglu, D., & Restrepo, P. (2019). *Robots and jobs: evidence from US labor*. Cambridge : MIT Economics.
- Andler, D. (2018). *Les sciences cognitives : un tour d'horizon*. Dans T. Collins, D. Andler, & C. Tallon-Baudry, *La cognition. Du neurone à la société* (pp. 15-70). Paris : Gallimard.
- Banerjee, S. (1986). *Reproduction of social structures : an artificial intelligence model*. Journal of Conflict Resolution(30), 221-252.
- Berreby, F., Bourgne, G., & Ganascia, J.-G. (2018). *Cadre déclaratif modulaire d'évaluation d'actions selon différents principes éthiques*. Revue d'Intelligence Artificielle, 32(4), 479-518.
- Bloor, D. (1976). *Knowledge and social imagery*. Londres : Routledge.
- Boltanski, L., & Chiapello, E. (1999). *Le nouvel esprit du capitalisme*. Paris : Gallimard.
- Brent, E., Glazier, J., Jamtgaard, K., Wetzel, E., Hall, P., Dalecki, M., & Bah, A. (1989). *Erving: a program to teach sociological reasoning from the dramaturgical perspective*. Teaching Sociology(17), 38-48.
- Cardon, D., Cointet, J.-P., & Mazières, A. (2018). *La revanche des neurones : l'invention des machines inductives et la controverse de l'intelligence artificielle*. Réseaux, 5(211), 173-220.
- Cassili, A. (2019). *En attendant les robots : enquête sur le travail du clic*. Paris : Seuil.
- Chateauraynaud, F. (2003). *Prospéro. Une technologie littéraire pour les sciences humaines*. Paris : CNRS.
- Chateauraynaud, F. (2019). *Petit traité de contre intelligence artificielle. Retour sociologique sur des expérimentations numériques*. Zilsel, 5(1), 174-195.
- Chomsky, N. (1957). *Syntactic structures*. Gravenhage : Mouton & Co.
- Colby, B. (1973). *A partial grammar of Eskimo folktales*. American Anthropologist(75), 645-662.
- Collins, H. (2019). *Sociologie métaprogrammée et intelligence artificielle*. Zilsel, 5(1), 161-173.

- Cuin, C.-H. (2002). *Le balancier sociologique français : entre individus et structures*. Revue européenne des sciences sociales, XL(124), en ligne.
- Dreyfus, H. L. (1972). *What computers can't do: a critique of artificial reason*. New York : Harper & Row.
- Dreyfus, H. L. (1992). *La portée philosophique du connexionnisme*. Dans D. Andler, *Introduction aux sciences cognitives* (pp. 352-373). Paris : Gallimard.
- Dubey, G. (2014). *Autonome comme si vous l'étiez*. La nouvelle revue du travail, (4), en ligne.
- Ferguson, Y. (2019). *Une intelligence artificielle au travail – Cinq histoires d'Homme*. Dans *L'intelligence artificielle – Dans toutes ses dimensions*. Paris : L'Harmattan.
- Frey, C. B., & Osborne, M. A. (2013). *The future of employment: how susceptible are jobs to computerisation?* Oxford : Oxford Martin School.
- Freyssenet, M. (1990). *Les techniques productives sont-elles prescriptives ? L'exemple des systèmes experts*. Cahiers de recherche du GIP « Mutations Industrielles » (45), 1-39.
- Ganascia, J.-G. (2007). *L'intelligence artificielle*. Paris : Le Cavalier Bleu.
- Garson, G. (1987). *Computer simulation in social science*. Dans G. Garson, *Academic microcomputing : a ressource guide* (pp. 110-138). Newbury Park : Sage.
- Goutefangea, P. (1999). *Alan Turing : la « pensée » de la machine et l'idée de pratique*. Nantes : Université de Nantes.
- Gras, A. (1994). *Les sciences de la communication en pratique : le paradigme cognitif dans le cas de l'aviation civile*. Quaderni, (23), 97-118.
- Keynes, J. M. (1931 [1933]). *Essais de persuasion*. Paris : Gallimard.
- Kirtchik, O. (2019). *STS et intelligence artificielle : une rencontre manquée ?* Zilsel, 1(5), 149-160.
- Newell, A., & Simon, H. A. (1956). *The logic theory machine : a complex information processing system*. IRE Transactions on Information Theory, (2), 61-79.
- Newell, A., Shaw, J. C., & Simon, H. A. (1959). *Report on a general problem-solving program*. Proceedings of the International Conference on Information Processing, 256–264.
- Olazaran, M. (1996). *A sociological study of the official history of the perceptrons controversy*. Social Studies of Science, 26(3), 611-659.
- Pharo, P. (2001). *Le sens logique des actes civils*. Dans J.-M. Baudouin, & J. Friedrich, *Théories de l'action et éducation* (pp. 45-66). Louvain-la-Neuve : De Boeck Supérieur.

- Piaget, J. (1967). *Biologie et connaissance. Essai sur les relations entre les régulations organiques et les processus cognitifs*. Paris : Gallimard.
- Pickering, A. (2019). *Techniques de l'engagement : la cybernétique et l'Internet of Things*. Zilsel, 1(5), 208-225.
- Rahwan, I. (2018). *Society-in-the-loop: programming the algorithmic social contract*. Ethics and Information Technology, 20(1), 5-14.
- Schillo, M., Fischer, K., & Klein, C. (2000). *The micro-macro link in DAI and sociology*. Multi-Agent-Based Simulation, 133-148.
- Searle, J. R. (1984). *Minds, brains, and science*. Cambridge : Harvard University Press.
- Simon, H. A. (1955). *A behavioral model of rational choice*. Quarterly journal of economics(69), 99-118.
- Suchman, L. (1987). *Plans and situated actions: the problem of human/machine communication*. Cambridge : Cambridge University Press.
- Sweetser, E. E. (1987). *The definition of lie: an examination of the folk theories underlying a semantic prototype*. Dans D. Holland, & N. Quinn, *Cultural models in language and thought* (pp. 44-64). Cambridge : Cambridge University Press.
- Tiffon, G. (2013). *La mise au travail des clients*. Paris : Economica.
- Vayre, J.-S. (2016). *Des machines à produire des futurs économiques : sociologie des intelligences artificielles marchandes à l'ère du big data*. Toulouse : Université Toulouse Jean Jaurès.
- Vayre, J.-S. (2019). *L'intelligence des machines et l'esprit du capitalisme : une histoire de congruence*. Communication, 36(1), en ligne.
- Vayre, J.-S., Delpech, E., Dufresne, A., & Lemercier, C. (2017). *Communication mediated through natural language generation in big data environments: the case of nomao*. Journal of Computer and Communications(5), 125-148.
- Vayre, J.-S., Larnaudie, L., Dufresne, A., & Lemercier, C. (2016). *Effet distracteur des agents de recommandation et stratégies de navigation*. Revue des Interactions Humaines Médiatisées, 17(1), 77-107.
- Velkovska, J., & Beaudouin, V. (2014). *Parler aux machines, coproduire un service : intelligence artificielle et travail du client dans les services vocaux automatisés*. Dans E. Kessous, & A. Mallard, *La fabrique de la vente : le travail commercial dans les télécommunications* (pp. 97-128). Paris : Presses des Mines.
- Wilensky, R. (1977, august 22-25). *PAM: A program that infers intentions*. Proceedings of the Fifth International Joint Conference on Artificial

Intelligence, 1, en ligne. Récupéré sur <https://www.ijcai.org/Proceedings/77-1/Papers/003A.pdf>

Wilensky, R. (1983). *Planning and understanding: a computational approach to human reasoning*. Boston : Addison-Wesley.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London : Routledge & Kegan Paul.

Woolgar, S. (1985). *Why not a sociology of machines? The case of sociology and artificial intelligence*. Sociology, 19(4), 557-572.

# L'intelligence artificielle

*Dimension philosophique*



# La triade animalité, humanité, *machinité*

## De la Sauvagerie animale à la *Sur-sophistication* de la machine

**Anne-Laure Thessard**

*La « sur-sophistication » exprime la crainte de la « machinité ». On peut faire un parallèle entre la notion de machinité et celle d'animalité. Tout comme la notion d'animalité provient de la projection de comportements humains sur les animaux, dans le but de rejeter certains comportements humains, la machinité correspond à la projection de caractéristiques imputées aux machines qui, appliquées à l'humain, donneraient lieu à des comportements froids, calculateurs, et distanciés émotionnellement. Dans le premier cas, on parlera d'« animalisation projective », dans le second de « machinisation projective ».*

*Prendre conscience de cette projection en termes de machinité permettrait à l'être humain de ne pas demeurer captif de ses représentations afin de mieux évaluer les enjeux des technologies nouvelles.*

Être humain consiste avant tout à projeter une image positive de l'humanité. Cette singulière capacité de projection permet à l'être humain de s'auto-constituer. En ce sens, il est difficile d'affirmer qu'il existe une « nature humaine ». Il serait plus juste d'affirmer que l'être humain s'auto-constitue en ayant recours à des projections symboliques. Comme l'expliquait le

philosophe et anthropologue Gilbert Durand<sup>342</sup>, le processus de l'imaginaire est consubstantiel à l'être humain. Il lui permet d'échapper à son angoisse existentielle de la mort en élaborant des similitudes et en se référant à des identités auxquelles il se réfère et d'autres qu'il exclue.

Tour à tour, l'être humain s'est identifié aux dieux, à un dieu, à la nature, et a repoussé des entités, telles les démons, les animaux et, plus tard, les machines dans le but s'auto-constituer. Dans la vision commune, il existe différentes façons de caractériser l'humanité. La première se réfère directement à l'espèce biologique : pour être humain il faut appartenir à l'espèce *homo sapiens* ; mais cela ne suffit pas. Il faut également être doué de langage, de conscience et d'une culture. Par ailleurs, la capacité de comportement moral, la considération éthique pour sa propre espèce et son environnement est une des caractéristiques qui placent l'homme au-dessus des autres espèces. En dehors de ces catégorisations positives, le fait de définir certains comportements comme inhumain contribue à l'auto-constitution de l'humain. Ainsi, l'homme qui ne considère pas l'autre humain comme un autre lui-même en le traitant tel un objet qu'il est possible d'annihiler de la façon la plus sauvage et la plus froide sera relégué au rang d'animal, voire de machine vide.

Pourquoi dire qu'être humain consiste en une projection de qualités considérées comme humaines ?

Une réponse possible serait la fragilité de l'état d'humain en tant qu'il est le produit d'une construction. Cette vulnérabilité ontologique est potentiellement une force créatrice, un vecteur de liberté, parce que l'humain est le seul à disposer d'une capacité d'auto-production consciente et délibérée. Être humain ne se réduit pas au fait d'appartenir à une espèce biologique ; c'est aussi le produit d'une construction conceptuelle, de projections symboliques. À ce sujet, Ernst Cassirer écrira :

« Ce qui caractérise l'homme est la richesse et la subtilité, la diversité et la souplesse de sa nature. [...] Toutes les prétendues définitions de l'homme ne sont que vaine spéculation tant qu'elles ne s'appuient pas sur notre expérience de l'homme et ne sont pas confirmées par elle. Il n'y a pas d'autres façon de connaître l'homme que de comprendre sa vie et son comportement. [...] L'homme n'a pas de "nature" — pas d'être simple et homogène. »<sup>343</sup>

Ainsi, l'homme est davantage un être en construction dont une détermination conceptuelle définitive ne saurait rendre compte. En parlant du

---

<sup>342</sup> G. Durand. 1960. *Les Structures anthropologiques de l'imaginaire, introduction à l'archétypologie générale*. Grenoble : impr. Allier.

<sup>343</sup> E. Cassirer. 1975. *Essai sur l'homme*. Translated by Norbert Massa. 1 vol. Le sens commun. Paris : Ed. de Minuit, p. 26-27.

« système symbolique » humain, Cassirer explique qu'il s'agit d'une singularité humaine qui lui permet de s'adapter à son milieu :

« L'homme a pour ainsi dire, découvert une nouvelle méthode d'adaptation à son milieu. »<sup>344</sup>

En outre, le fait de projeter des symboles permet de se référer au monde en l'investissant de sens :

« Le principe du symbolisme, par son universalité, sa force et son applicabilité générale, est le mot magique, le sésame du monde spécifiquement humain, celui de la culture. Une fois en possession de cette clé magique, l'homme est assuré d'aller plus avant. »<sup>345</sup>

Ces constructions conceptuelles, qui prennent la forme de projections de symboles, sont pour Ernst Cassirer la possibilité de structurer nos représentations et d'avoir une perception toujours plus précise de soi-même, d'autrui et de notre environnement.

Il existe une analogie entre le fait de construire le monde en se référant à un système de symboles et le fait de projeter des représentations au sujet de l'être humain afin de l'investir d'un sens : celui de construire l'humanité comme valeur positive.

Cette construction repose, du point de vue éthique, sur la projection de qualités considérées comme positives et sur la condamnation de comportements qui sont jugés nuisibles, dans un projet d'épanouissement de l'humanité.

Le repérage des comportements « nuisibles », pour servir de garde-fou, de limite à l'action, revêt une importance particulière. En effet, dans l'opinion commune, la principale menace pour l'humanité est l'animalité, non pas l'animalité en tant que catégorie qui renvoie à la vie animale réelle, mais l'animalité comme représentation d'une puissance active en l'homme, qui peut conduire à la perte de son humanité. L'inhumain est stigmatisé comme « animal » ; on dit couramment devant un acte ignoble que l'humain qui en est responsable n'est qu'un animal, soumis à ses instincts les plus vils, les plus malfaisants. Pourtant, lorsque nous pensons cela, nous référons-nous réellement à l'animalité animale ? Il semble qu'il s'agisse davantage d'une projection de ce que l'homme rejette en lui pour pouvoir s'auto-constituer. Cet aspect nuisible chez l'homme — qualifions-le de « chaud » — renvoie aux passions, à l'impulsivité, à l'inconscience, à la sauvagerie.

Pourtant, nous allons voir qu'il existe aussi une dimension « froide » dans la notion d'humanité, que nous proposons de nommer *machinité*. On aura reconnu la racine du mot « machine ». La machine renvoie à une

---

<sup>344</sup> *Ibid.*, p. 43.

<sup>345</sup> *Ibid.*, p. 58.

catégorisation des objets réels, mais elle renvoie aussi à des représentations projectives, soit de l'être humain sur les machines, soit de la machine sur les comportements humains. À un niveau plus global, la littérature contemporaine, du milieu du XXe siècle, est plutôt hostile à l'égard des machines. Que ce soit pour Arendt, Ellul ou Simondon, la machine est pensée dans un système technocratique qui s'intensifie et qui aurait pour conséquence la machinisation humaine. Dans les représentations communes, la machine est perçue comme froide, sans vie ; elle peut avoir une puissance de calcul sans conscience, elle singe parfois sans intentionnalité des comportements humains, elle est « mécanique ». Qualifions de *machinité* ce miroir froid de l'animalité en les considérant l'une et l'autre du point de vue de l'humanité.

#### Pourquoi définir cette catégorie de *machinité* ?

Tout d'abord, parce que l'homme pendant des siècles a tenté de détourner, de contenir, de sublimer son versant animal : la nouvelle menace intérieure serait la *machinité* en tant qu'elle représente comme l'animalité une menace pour l'humanité. Cette menace a probablement toujours existé en l'homme, mais l'identification d'un comportement machinal s'est opérée plus ou moins confusément en même temps que l'homme utilisait le paradigme mécaniste pour accéder à un savoir empirique. Plus précisément, depuis que la pensée mécaniste de la Renaissance a remplacé une vision déiste du monde (dont l'anti-modèle était le naturel) par une vision mécaniste du monde (dont l'anti-modèle est la subjectivité). La machine a été particulièrement identifiée comme menace pour l'humanité de l'humain à partir du moment où elle s'est développée avec une telle intensité qu'elle a imposé une interrogation sur son emploi et ses conséquences.

À noter que la *machinité* ne renvoie pas à proprement parler à la machine, mais elle fait partie du potentiel humain, tout comme l'animalité est l'expression d'un comportement bien humain qui ne reflète pas nécessairement une particularité animale. L'idéologie nazie est un exemple de comportement qu'il ne faut pas se hâter de qualifier de brutal, bestial, animal, car il pourrait relever de la catégorie de *machinité*. L'expression de l'idéologie nazie s'est révélée être une rationalité *machinale* (rappelant le mot « animale ») qui a abouti à l'élimination systématique, mécanique, rationalisée d'individus prétendus indésirables. Cet exemple compte parmi les plus représentatifs de ce que la *machinité* humaine est en mesure de concevoir et réaliser, c'est la concrétisation d'un des potentiels humains à prendre sérieusement en considération.

En outre, la question contemporaine de la *machinité* ne va pas sans la considération d'un système de valeurs centré sur les technologies. Le développement, l'usage et la réflexion sur l'usage des machines constituent une question qui est devenue significative et importante au sens éthique du terme. Il s'agit de ne pas confondre les objets, qui ne sont en eux-mêmes ni

bons ni mauvais, avec l'usage qu'on en fait. Le risque est en effet de condamner les nouvelles technologies, en se référant par exemple à une idéologie systématique pour laquelle l'homme serait prisonnier d'un *déterminisme technologique*, alors que ce comportement machinal est une caractéristique humaine intrinsèque.

L'humain peut être conçu comme en équilibre entre un pôle « chaud » qui serait l'*animalité* et un pôle « froid » qui serait la *machinité*. Ces deux pôles sont à la fois attractifs et répulsifs selon les circonstances. Pendant longtemps, le pôle animalité a été répulsif dans l'histoire de la constitution de l'humanité comme valeur positive. Nous sommes depuis peu devenus capables d'envisager une relation positive avec notre pôle animal. Le processus équivalent concernant la *machinité* est encore largement dans sa phase répulsive. L'humain « tempéré » doit donc se situer à juste distance de ces deux possibles : l'*animalité* et la *machinité*. On peut parler d'un triangle *animalité, humanité, machinité* qui en termes de projections humaines s'exprime par l'association de valeurs à chacune de ces entités.

Pour illustrer cette image, prenons un exemple cinématographique mettant en jeu les représentations de l'*animalité* et de la *machinité*. L'une des images iconiques de la saga Star Wars illustre la triade *animalité, humanité, machinité* : le pilote Han Solo, aux commandes de son vaisseau, est entouré de l'*animal* évolué, Chewbacca, et du robot, la machine C-3PO. On peut y voir les deux polarités du potentiel humain. D'une part, Chewbacca, l'*animal* très poilu, soumis à ses émotions ; ses sentiments semblent peut subtils, il exprime sa colère, sa tristesse et sa joie par des sons non articulés, et ses mouvements sont animés par la seule force. D'autre part, le robot C-3PO, très délicat, raffiné intellectuellement, dans une maîtrise totale (extrême) de ses mouvements, désincarné à tel point qu'il n'est pas étonnant que sa chair soit de métal. La triade *animalité, humanité, machinité*, loin de correspondre à une tripartition du réel, au sein de laquelle l'être humain devrait interagir avec des dynamiques qui lui sont extérieures, représente finalement la projection à la fois de ce que l'être humain souhaite être (un humain) et de ce qu'il est (un humain animal et *machinal*).

Comment s'exprime la menace de la *machinité* en termes de civilisation ?

Si la *sauvagerie* est condamnable à cause de son caractère animal et donc inhumain, si l'effort de civilisation a longtemps consisté à éloigner les comportements violents, irraisonnés, qui mettent à mal le lien interindividuel et la possibilité d'évoluer ensemble, la *sur-sophistication machinale* constitue paradoxalement une menace tout aussi sérieuse. Par *sur-sophistication machinale*, il faut comprendre toute action, logique ou comportement qui est soumis à une complexité excessive et artificielle. En cela, la *sur-sophistication* est l'inverse du simple, du spontané et du naturel. Si, d'un côté, on souhaite toujours mettre en avant le fait que l'être humain se distingue de l'*animal* par

sa complexité (intellectuelle et émotionnelle), d'un autre côté, on est aussi enclin à rappeler sans cesse que l'homme n'est pas une machine, qu'il est spontané et doit rester au plus proche de sa nature première. Mais, quelle est cette nature première ? Il n'existe pas de nature humaine en tant que telle. L'effort de civilisation ne peut donc consister en l'éviction de comportements considérés comme « non-naturels » à l'humain. Il s'agit bien plutôt d'accéder à la compréhension de ce que ces comportements, animaux et *machinaux*, ont de proprement humain, dans le but de les apprivoiser et de bénéficier de leurs aspects positifs. C'est ce que nous tenterons de montrer en expliquant la nécessité de reconnaître la *machinité* comme projection humaine.

Nous mènerons cette analyse en trois temps. Il s'agira d'abord d'identifier les projections humaines en termes d'*animalité*, d'*animal-machine* puis de *machinité*. Ensuite, nous verrons en quoi la *machinité* représente la nouvelle menace pour l'être humain. Enfin, nous verrons en quoi la *sur-sophistication machinale* doit sérieusement être prise en considération dans l'effort de civilisation.

## I. Humanité et projection : l'*animalité*, l'*animal-machine*, la *machinité*

Nous aborderons dans cette première partie la triade *animalité*, *humanité*, *machinité* en termes ontologiques dans l'élaboration des deux contre-modèles que sont l'*animalité* et la *machinité*. Nous verrons que l'être humain est par nature indéterminé, qu'il se représente en tant que non animal et que l'époque contemporaine lui enjoint plus que jamais de ne pas devenir une machine et de composer avec son *ambivalence créatrice*.

Nous avons supposé que l'espèce humaine n'est pas déterminée en tant que telle, qu'elle consiste en une projection de caractéristiques considérées comme humaines et qu'elle se singularise par le fait de repousser certains comportements considérés comme inhumains. Par « projection », il faut comprendre la capacité humaine à s'auto-déterminer en définissant des caractéristiques censées être le propre de l'être humain, auxquelles des valeurs sont associées, et à en rejeter d'autres pour le motif inverse, à savoir qu'elles éloigneraient l'être humain de son humanité.

Il est intéressant de se référer à la valeur sémantique attribuée aux mots « *animal* », « *animaux* », « *animalité* ». Cette valeur dépasse le champ sémantique et s'apparente à un jugement de valeur. Proposons d'abord, à la suite de la linguiste Astrid Guillaume, une analyse sémiologique des mots *humanité*, *humain*, *animalité*, *animal* et *bestialité*.

Tout d'abord pour les termes « humanité » et « humain » :

« - Humanité, Sens 1 : Genre humain.

Sens 2 : Compréhension, bienveillance. Synonyme : bonté.

Sens 3 : Caractère de ce qui est propre au genre humain. Synonymes : altruisme, bienveillance, bonté, compassion, douceur, monde.

- Humain, adjetif, Sens 1 : Qui a trait à l'homme.

Sens 2 : Sensible, compréhensif. Synonyme : accessible.

- Humain, nom masculin, Sens 1 : Personne humaine. Synonyme : individu. »<sup>346</sup>

Puis, les termes « animalité », « animal », « bestialité » :

« - Animalité, Sens : Ensemble des caractères propres à l'animal [Zoologie]. Synonyme : bestialité.

- Animal, nom, Sens 1 : Etre vivant organisé, doué d'une sensibilité et capable de mouvement [Zoologie].

Sens 2 : Être vivant animé autre que l'homme. Synonyme : bête.

Sens 3 : Personne grossière, brutale, dépourvue d'intelligence. Synonyme : abruti.

- Animal, adjetif, Sens 1 : Propre aux animaux [Zoologie]. Synonyme : inné.

Sens 2 : Bestial, digne d'un animal [Figuré]. Synonyme : brutal.

- Bestialité : Caractère de ce qui est bestial. Synonyme : brutalité »<sup>347</sup>.

Et Astrid Guillaume de poursuivre :

« Le fossé sémantique et sémiotique entre la sphère de l'animal et la sphère de l'humain est immense et extrêmement manichéen : d'un côté la bonté et la compassion ressortent, de l'autre la grossièreté, l'absence d'intelligence et la brutalité, bref la bêtise de la bête, la confusion du sens propre et du sens figuré œuvrant également dans ce sens (être bête versus la bête). »<sup>348</sup>

La définition des notions d'humanité et d'animalité est rarement neutre. Que cela se produise consciemment ou non, il y a dans ces définitions un

---

<sup>346</sup> A. Guillaume. 2013a. « Humanity and Animality. A Transdisciplinary Approach. » In *Review Human and Social Studies*, éd. Walter de Gruyter, Berlin-Boston. (Version française.), n° 3.

<sup>347</sup> *Ibid.*

<sup>348</sup> *Ibid.*

dessein plus vaste que l'acte de décrire. En ce sens, la notion d'animalité n'est pas descriptive mais elle correspond à l'attribution de valeurs négatives aux animaux, permettant de valoriser la notion d'humanité. Il en va de même concernant la notion de *machinité* : attribuer des caractéristiques négatives aux machines permet aussi à l'être humain de construire une vision positive de l'humanité. Quand il s'agira de projections négatives par rapport à la notion d'animalité, on parlera d'*animalisation projective* ; quand il sera question de projections négatives sur les machines, on parlera de *mécanisation projective*. Dans le processus humain consistant à projeter des contenus sur des objets vivants et non-vivants, l'*animalisation projective* et la *mécanisation projective* désignent la tendance humaine à projeter des caractéristiques, des sentiments et des valeurs négatifs sur les animaux et les machines, dans le but de construire, de renforcer le sentiment positif et la valeur positive attribués à la notion d'humanité.

Parler de nature humaine s'avère donc inadéquat. Être humain est un projet sans cesse à consolider, toujours menacé. Parler de nature humaine introduit une forme d'ambiguïté : l'être humain n'est ni tout à fait un être de nature ni tout à fait un être de culture. Cette ambivalence indécidable provient du fait que l'homme est par nature indéterminé.

Si on doit envisager une nature humaine première, il faut la considérer en tant qu'indétermination naturelle humaine. Cette ambivalence, nous la qualifions de *créatrice*, nous parlerons d'*ambivalence créatrice*, dans le sens où son indétermination naturelle permet à l'être humain de naviguer entre différentes projections de son être.

L'aventure humaine est singulière par la capacité de l'homme à s'auto-constituer, non pas uniquement de façon individuelle mais aussi sous l'influence de dynamiques culturelles qui insufflent un sens plus large à la pensée, aux actions et aux représentations du monde. Mais cette capacité de s'auto-déterminer ouvre aussi la possibilité de s'auto-annihiler. Cette tension consubstantielle à l'être humain s'exprime, si l'on se réfère aux repères historiques de la pensée occidentale dans la mythologie, sous la forme d'Éros et Thanatos qui sont les projections de puissances, finalement bien humaines, soit d'un potentiel de vie, de création et de mort, de destruction. En effet, autant il semble évident que l'histoire humaine est jalonnée de projections d'entités immatérielles, que ce soit dans les pensées animistes, le polythéisme ou les religions révélées, autant on sous-estime la nécessité et l'incidence de projections plus ordinaires, davantage en prise avec le réel. Nous nous intéressons ici particulièrement au triangle *animalité*, *humanité*, *machinité*, car il exprime le potentiel de création et de destruction. Il est alors intéressant d'analyser comment l'être humain gère l'*ambivalence créatrice* de son pouvoir concret de créer et de nuire. Il y aurait beaucoup à dire sur le potentiel positif de l'*animalité* et de la *machinité*. Nous allons ici considérer comment l'être humain projette sa capacité d'auto-annihilation pour sauvegarder et

développer son humanité. La linguiste Astrid Guillaume explique en effet, au sujet de la représentation sémantique de l'humanité *versus* l'animalité :

« Depuis l'Antiquité, l'humanité est définie soit par ce qui fait justement son absence d'humanité (tortures, guerres, génocides), soit en fonction de ce qui la distinguerait *a priori* de l'animal (civilisations, croyances, édifices architecturaux, voire cultures, langues, souffrances, sensibilité, intelligence, ce qui est discutable). »<sup>349</sup>

L'animalité constitue la première des menaces pour les valeurs humaines. Cette menace est facilement identifiable ne serait-ce que dans le langage courant, mais plus largement dans l'effort de civilisation global dont l'un des premiers objectifs consiste à s'éloigner du « sauvage », en d'autres termes de l'animalité.

« [...] quand l'homme se comporte en tortionnaire, volontairement, voire avec un plaisir sadique, quand il est belliqueux ou agressif, quand il commet des meurtres ou viols en série, quand il décime des ethnies entières. Dans ces cas-là, il agit toujours sous l'influence de l'animal qui sommeille en lui, ce n'est alors jamais l'humain qui se trouve aux commandes de l'appareil, c'est l'animal ! Il n'est pourtant ici question ni de comportement animal ni de comportement humain mais véritablement de comportement monstrueux, car aucun animal ne planifierait l'extermination totale des membres de sa communauté. »<sup>350</sup>

Les comportements les plus violents ou indécents sont directement associés à la notion d'animalité. La conduite en société, tant dans la présentation que dans les rapports inter-individuels, la conduite dans la vie intime, sentimentale ou sexuelle, est jugée à l'aune de la capacité à s'éloigner le plus possible d'une posture animale. Que ce soit physiquement ou moralement, l'homme a le devoir de se départir de toute ressemblance à l'animal. Il ne s'agit pas ici de faire la critique morale de la condamnation de l'animalité, mais plutôt d'identifier la menace « animale » envisagée comme critère de condamnation de comportements humains. Cette menace prend sens en tant qu'elle est une représentation, une projection, presque au sens littéral du terme : « jeter au loin ».

Mais qu'est-ce qui est rejeté loin de soi ?

C'est une animalité vécue ou imaginée et non des caractéristiques propres à l'animal. Ce qui est intéressant est de concevoir que cet effort de dissociation

---

<sup>349</sup> A. Guillaume. 2013b. « Humanité vs Animalité. » In *Revue Trimestrielle de La Fondation Droit Animal, Ethique et Sciences*, n° 79 (September). <http://www.fondation-droit-animal.org/documents/revue79.pdf>.

<sup>350</sup> *Ibid.*

s'exprime par la projection de caractéristiques dites non-humaines ou inhumaines. Pourtant, il y a bien un fossé entre la réalité animale et l'*animalisation projective*. La distinction entre humanité et animalité et l'*animalisation projective* ont pour fonction de développer, de sublimer l'humanité de l'être humain en projetant hors de lui toute dénaturation « animalisante ». Finalement, être humain est illusoirement vécu comme un absolu qui doit éviter d'être corrompu par l'animalité qui est considérée comme son contraire.

L'idée d'animal-machine, elle, correspond, en termes de projection, à un glissement dans une pensée idéaliste et dogmatique, au sens religieux du terme, et à un affermissement mécaniste. Ce glissement se manifeste notamment par l'enfermement de l'animal dans le statut de machine, qu'on trouvera exprimé sous la plume de Descartes par le concept d'« animal-machine »<sup>351</sup>. Cette appellation permet d'assurer la sauvegarde du statut singulier de l'être humain dans le cadre d'un paradigme mécaniste : l'être humain n'est pas un animal, il n'est donc pas non plus une machine puisque ce statut est celui de l'animal. En termes de représentation, on peut y voir une forme de parenthèse dans l'histoire intellectuelle du projet d'humanisation. En effet, l'être humain se retrouve pris entre deux menaces : celle de la réduction naturaliste à l'animalité et celle de la réduction mécaniste de la subjectivité humaine. « L'animal-machine » devient donc un moyen terme qui, par la capacité d'*ambivalence créatrice*, permet à l'être humain de s'auto-déterminer en composant avec les différentes projections de son potentiel, en repoussant ce qu'il ne veut pas être et en développant ce qu'il a envie de devenir. Petit à petit apparaît donc une autre menace à cet effort pour lutter et dépasser le « sauvage », indigne de l'être humain. Cette nouvelle menace est celle de la *machinité*.

## II. La *machinité* : la nouvelle menace pour l'humain

Dans cette partie, il sera question de la constitution de la *machinité*, nouveau contre-modèle humain, préfiguré conceptuellement par la pensée mécaniste. Ce contre-modèle s'applique au niveau de l'individu mais aussi au niveau des organisations humaines, il s'exprime par l'idée de *machinisation projective*. L'enjeu de cette approche consiste également à mettre en avant l'idée selon laquelle prendre conscience de l'*animalisation projective* et de la *machinisation projective* permettrait à l'être humain de ne pas demeurer captif de ces représentations.

---

<sup>351</sup> R. Descartes, 1946.

Comme nous venons de le suggérer en fin de première partie, la pensée mécaniste imprime une nouvelle forme au réel compris comme représentation du vivant et de l'être humain. Comme nous l'avons souligné, s'il est déjà difficile de départir l'être humain d'une forme d'animalité il devient compliqué de ne pas le réduire à une simple machine. À ce sujet, dans *Le Mythe de la machine*, Mumford parle du « Crime de Galilée »<sup>352</sup>. Le penseur est accusé (sans que soit porté de discrédit à l'apport de sa pensée) d'un crime symbolique envers la sauvegarde de la subjectivité humaine. Mumford explique que, si la religion a été dogmatique dans sa critique de la nature, le mécanisme peut l'être aussi dans la critique de la subjectivité :

« Pendant près de trois siècles, les savants suivirent les directions de Galilée. Dans la croyance naïve, exposée, [...] qu'ils étaient libres de préjugés métaphysiques, les interprètes orthodoxes de la science supprimèrent toute preuve de comportement humain et organique que l'on ne pouvait faire entrer exactement dans leur image mécanique du monde. Ils commirent ainsi, à rebours, l'erreur des anciens Pères de l'Église, lesquels avaient supprimé tout intérêt pour le monde naturel afin de se concentrer sur le destin de l'âme humaine dans l'éternité. [...] En toute innocence, Galilée avait abdiqué le droit de naissance historique de l'homme : son expérience mémorable et remémorée, en bref, sa culture accumulée. En rejetant la subjectivité, Galilée avait excommunié le sujet central de l'histoire, l'homme multidimensionnel. »<sup>353</sup>

De la projection rassurante « d'animal-machine » au statut « d'homme-machine »<sup>354</sup>, il y a un dépassement de la menace animale : si l'humain peut-être un animal et que l'animal est une machine, l'homme peut aussi être une machine. La triade *animalité*, *humanité*, *machinité* prend forme ; le spectre de la *machinité*, sans qu'il soit distinctement appréhendé, commence à être perçu comme menace existentielle et éthique.

Bien que l'animal ait été significativement relégué au statut de machine, cela ne rend pas définitivement inopérante l'animalisation projective. De fait, comme nous venons de le voir, l'être humain se retrouve entre deux menaces : celle de la dégradation animale et celle de la réduction mécaniste. En effet, comme l'animal est une machine et que l'animalité est une caractéristique humaine, à combattre, on peut même envisager un couple de projection qui serait « *l'animalité-machinité* », pour reprendre l'expression de Descartes d'animal-machine. Comme nous l'avons annoncé en introduction, si l'être humain correspond à une vision plus tempérée du versant animal, il existe

<sup>352</sup> L. Mumford. 1973. *Le mythe de la machine*. Translated by Léo Dilé. 2 vols. Le Phénomène scientifique. Paris : Fayard.

<sup>353</sup> *Ibid.*, p. 74.

<sup>354</sup> La Mettrie, 1999.

aussi un pendant « froid » à l’animalité : la *machinité*. Il s’agit d’une projection humaine de comportements que nous qualifions de *machinaux*, qui correspond également à la qualification de caractéristiques attribuées à la machine. Entendons par là que dans la *machinisation projective* tous les comportements froids, mécaniques, automatiques, dénués d’empathie, qui marquent en somme l’absence de subjectivité, de conscience, de spontanéité et tout simplement de vie dans ce qu’elle offre en relief, en chaleur et en diversité, peuvent être associés à une attitude *machinale*, au sens négatif du terme.

Ainsi, au niveau individuel, une réaction manifestant un degré excessif d’emportement ou de violence, verbale ou physique, pourra être associé à l’animalité. À l’opposé, un comportement froid, calculateur, distancié émotionnellement, relèvera de la catégorie de *machinité*, qui semble plus à même de rendre compte de cette forme de violence visuellement moins extraordinaire mais pourtant potentiellement assez nuisible, voire tout aussi nuisible. Pour reprendre des exemples de groupes, on peut noter la différence entre la vision d’un bûcher en place publique et des chambres à gaz. L’effroi à l’évocation de ces deux supplices de mise à mort est immédiat, pourtant leur nature diffère. La cruauté est le dénominateur commun de ces deux procédés de mise à mort, mais dans un cas c’est la bestialité spectaculaire qui impacte l’esprit, quand dans la seconde c’est la froideur du « système » qui glace le sang.

La *machinité* peut être appréhendée au niveau du système conçu comme principe gouvernant les actions humaines. Le « système » est alors un principe législateur rationnel sans souffle, sans vie, sans dieu, et quasiment sans humain pour le maîtriser. La violence imputée au « système » comme force de déshumanisation, de standardisation, froide et implacable, renvoie à la « mégamachine ». Ce concept de Mumford tend à montrer que l’humanité se retrouve dépourvue de son pouvoir d’action au sein du système. Il écrira au sujet de la mégamachine :

« L’idéologie qui sous-tend et unit les mégamachines anciennes et modernes est une idéologie qui ignore les nécessités et les buts de la vie afin de fortifier le complexe de puissance et d’en étendre la domination. »<sup>355</sup>

Le discours de Mumford sur les conséquences d’un système fondé sur les mégamachines entre dans la logique générale des critiques de l’aliénation humaine par des superstructures depuis Marx. Mais ce qui nous intéresse ici est la déssubjectivisation qui entraîne la machine :

« Nous avons donc à faire face à une culture hyperorganisée, hypermécanisée, hyperdirigée, hyperprévisible. À jouer aux jeux

---

<sup>355</sup> *Ibid.*, p. 73.

économiques et sociaux vides qui servent ce processus automatique, les êtres humains deviennent des “objets” ou des “pions” destinés à être traités de la même façon que n’importe quel échantillon fortuit de matière brute. »<sup>356</sup>

Le concept de mégamachine est intéressant en ce qu’il décrit un lien direct entre le développement de la pensée mécaniste, la machine et le développement de sociétés capitalistes. Mais il semble important de ne pas perdre de vue que la tendance *machinale* est un fait humain et non une nouvelle forme de Mal extérieur. Ce ne serait donc pas la machine, peu importe ses formes (objet, système associé à un système de valeur), qu’il faudrait combattre, interroger en tant que telle, mais bien plutôt la tendance *machinale* de l’homme.

La *machinité* constitue la nouvelle menace envers l’humain, au niveau individuel. On parle alors « d’humain simplifié »<sup>357</sup> par opposition à « l’humain augmenté » qui serait une savante fusion entre l’homme et la machine. La *machinité* constitue également une menace au niveau des relations inter-individuelles et des rapports de groupe. Elle a probablement toujours existé. Elle a commencé à se faire plus signifiante avec l’apparition du paradigme mécaniste (qui peut aller jusqu’à dévaluer la subjectivité), qui remplace le paradigme déiste (*qui peut aller jusqu’à dévaluer* contre le « naturel »), conférant au réel et à la réalité vécue une dimension machinale forte. Pourtant, bien qu’il soit nécessaire pour l’être humain que la *machinité* (et l’animalité) soit perçue selon des valeurs négatives afin que l’humain s’auto-constitue comme humain, en rester là correspondrait à une absence de recul qui limiterait la liberté, et peut-être le projet global d’épanouissement, de l’humanité. En effet, le Mal absolu fut longtemps incarné par l’animalité, et nous affirmons que la nouvelle menace de la mégamachine écrasante serait symbolisée par la *machinité*. Projections internes dont la dialectique interne est partout, l’animalité et la *machinité* sont des caractéristiques humaines qu’il ne s’agit pas de combattre de l’extérieur.

---

<sup>356</sup> *Op. cit.* L. Mumford. *Le mythe de la machine*. p. 489.

<sup>357</sup> Cf. la notion d’homme simplifiée développée en opposition avec l’idée transhumaniste d’homme augmenté, J.-M. Besnier. 2012. *L’homme simplifié : le syndrome de la touche étoile*. 1 vol. Paris : Fayard.

### III. De la sauvagerie à la *sur-sophistication machinale* : *machinité* et enjeux de civilisation

Dans cette dernière partie, il sera question des enjeux éthiques de la notion de *machinité* concernant l'effort de civilisation. Du point de vue éthique, la notion de *machinité* devrait être assumée en tant que versant humain et non condamnée pour des raisons morales extérieures, imputées à un déterminisme technologique. La *machinité* est aussi une caractéristique humaine et non une donnée extérieure propre à la « machine ». Mumford explique que la machine n'est pas qu'un produit autonome ; il est avant tout un produit de l'activité humaine. En tant que tel, comprendre la machine, c'est aussi comprendre l'être humain :

« La machine, cependant, est un produit de l'ingéniosité et des efforts humains. Aussi, comprendre la machine, ce n'est pas simplement faire un premier pas pour réorienter notre civilisation, c'est aussi trouver un moyen de comprendre la société et nous comprendre nous-mêmes. Le monde de la technique n'est pas isolé et autonome. »<sup>358</sup>

Pourtant, comprendre la *machinité* n'est pas seulement un effort de compréhension de la *machinisation projective*, c'est aussi la possibilité d'attribuer une valeur à la *machinité* dans l'effort de civilisation. Une société trop « animale », dans le sens péjoratif du terme, conduit l'être humain à sa propre perte, à la sauvagerie humaine. De la même façon, une société trop « machinale » mène à une *sur-sophistication machinale*. Toute action, logique ou comportement extrêmement *sophistiqué* peut mener à la perte du sens commun, du sens des réalités. C'est-à-dire que la *machinité* humaine comprise comme complexification excessive, ou *sur-sophistication*, donne une vision de l'être humain désincarné, soumis à une logique froide. Il devient une belle machine vide tournant sur elle-même.

Les exemples du *body-building* et de l'emploi abusif de la chirurgie permettent une compréhension immédiate de la *sur-sophistication* des corps. Que ce soit dans la pratique excessive de la musculation, avec l'exemple du *body-building*, ou dans l'emploi non raisonné de la chirurgie esthétique, le corps se retrouve figé, le mouvement naturel limité, l'expression générale raide et mécanique. C'est une forme de violence qui a trait à la maîtrise, à l'adoption de comportements standardisés et *machinaux*. Le risque est alors d'être guidé par une logique *machinale* qui prendrait le pas sur l'*ambivalence créatrice* naturelle de l'être humain (qui s'exprime dans la triade *animalité, humanité, machinité*). En effet, dans le cas d'une *sophistication* importante,

---

<sup>358</sup> L. Mumford. 1976. *Technique et civilisation*. Collections Esprit. Paris : Éditions du Seuil. p. 19.

l'être humain amoindrit sa capacité à naviguer entre les différentes projections de son être.

Ainsi, la *sur-sophistication*, comprise comme conséquence négative de comportements *machinaux* (de la même façon que la sauvagerie est la conséquence de comportements considérés comme animaux) constitue une menace à l'encontre de l'ambivalence créatrice humaine, en insistant sur la polarité « machine » du triangle *animalité*, *humanité*, *machinité*. Une civilisation qui serait guidée par le seul désir de développer des technologies performantes, dont les machines préfigurent la tendance *machinale* humaine, serait confrontée à son propre vide, à un mouvement automatique, froid. On peut y voir une forme de dissolution du rapport à soi-même, des rapports interindividuels et de l'effort de civilisation. Le fil conducteur de ce questionnement consiste à insister sur la valeur de l'action humaine face aux projections qui lui sont propres, en l'occurrence l'*animalisation* et la *machinisation projectives*, et non face à un ennemi qui serait considéré comme extérieur. Contrairement à ce qu'affirment les détracteurs de la technologie tels que Jacques Ellul<sup>359</sup>, la menace n'est pas le système en lui-même. Par extension, la menace n'est pas la *sur-sophistication* liée à la *machinité*. L'erreur d'identification de la nature de cette menace maintient l'être humain dans sa tendance *machinale* parfois excessive.

Ce sont les objets et le système, qui permet la production de ces objets, qui sont incriminés. Pourtant, la menace réside dans l'être humain lui-même ; cette méconnaissance implique une mauvaise évaluation des possibilités de contourner la menace de mécanisation. Par exemple, il ne s'agit peut-être pas de condamner les possibilités ouvertes par les nouvelles technologies en général, mais plutôt de reconnaître que la propension à adopter des comportements *machinaux* est naturelle. Naturel<sup>360</sup> ne signifie pas nécessairement bon. Cela est facilement identifiable dès qu'on parle d'*animalité* dans le cadre de l'*animalisation projective*. Mais l'*animalité*, tout comme la *machinité*, n'est pas non plus forcément nuisible à l'être humain. Dans les rapports humains, dans l'intimité, il est parfois bon d'être spontané, de s'en remettre à son instinct, d'être connecté à ses émotions primitives. Dans certaines activités, on pensera par exemple à l'action du chirurgien en salle d'opération, il est parfois bon d'avoir un regard purement analytique, d'adopter un comportement automatique, de laisser de côté ses émotions etc. À l'inverse, la rigidité trop animale, trop humaine, trop *machinale* est rarement souhaitable. Finalement, reconnaître l'*ambivalence créatrice* humaine permet

---

<sup>359</sup> Cf. le concept de technocratie, J. Ellul. 1977. *Le Système technicien*. Liberté de l'esprit. Paris : Calmann-Lévy.

<sup>360</sup> Le terme naturel renvoie soit à un système de valeur lié à des projections humaines sur ce qui est considéré comme naturel, soit au naturel lui-même. Dans un cas comme dans l'autre le naturel n'est pas intrinsèquement bon.

de naviguer consciemment entre les différentes possibilités de l'être humain en identifiant les moments où la polarité est trop déséquilibrée dans le sens de l'animalité ou de la *machinité*.

En termes d'effort de civilisation, cela signifie qu'il est utile d'identifier la logique gouvernant la vie de la civilisation. Que ce soit la barbarie, les religions, l'économie ou la *sur-sophistication machinale*, l'identification des forces actives permet à l'être humain de s'extraire de la pénombre dans la grotte, pour faire référence au fameux mythe de Platon. En revanche, il est peut-être vain de vouloir combattre la pénombre elle-même en la considérant comme une nuisance extérieure à l'être humain.

Reconnaitre la triade *animalité, humanité, machinité* dans sa dimension projective est l'occasion de sortir du rapport attraction/répulsion à l'encontre des technologies en général et dans le cadre de la problématique contemporaine par rapport aux nouvelles technologies NBIC<sup>361</sup>. Ce rapport anxiogène (répulsion) ou hypnotisant (attraction) a pour conséquence de scléroser la pensée dans une forme d'abattement face aux potentiels de plus grande fusion homme-machine, ou alors de précipiter des comportements sur-sophistiqués. Or assumer l'*ambivalence créatrice* et la triade *animalité, humanité, machinité* permet de placer l'être humain dans de meilleures dispositions pour appréhender, percevoir, mesurer, réfléchir aux dimensions qu'il souhaite privilégier, tout en naviguant au travers de projections assumées, animales et *machinales*.

Il faudra retenir du triangle *animalité, humanité, machinité* que les notions d'*animalité* et de *machinité* ne sont pas que descriptives ; elles correspondent aussi à l'attribution de valeurs négatives aux animaux et aux machines, permettant ainsi de donner un sens positif à la notion d'*humanité*. On parlera alors d'*animalisation projective* et de *machinisation projective*. Ainsi, l'*animal* et la *machine* font l'objet de projections qui permettent à l'être humain, indéterminé par nature, de s'*auto-constituer* comme humain. L'être humain se représente comme étant non-animal, mais il doit dorénavant souhaiter aussi ne pas devenir une machine. De fait, du point de vue ontologique, la nécessité de s'*auto-déterminer* comme humain est en rapport avec deux contre-modèles. La *machinité* est le nouveau contre-modèle, du moins en termes de perceptions conscientes, car l'être humain a vraisemblablement toujours eu une tendance *machinale*.

Il s'agit donc de prendre conscience de cette tendance *machinale*, en ce qu'elle est constitutive du fonctionnement humain, tout comme l'*animalité* est, finalement, une représentation humaine. En prendre conscience, c'est aussi prendre conscience de la nécessité de la projection de comportements

---

<sup>361</sup> À ce sujet, se reporter au rapport américain de 2003 (Rocco and Brainbridge, 2003), et son équivalent européen le rapport Nordmann de 2004 (Nordmann, 2005).

« animaux » et « *machinaux* » en dehors de la dimension humaine. En effet, il est nécessaire, pour l'être humain, que ce mécanisme soit perçu selon des valeurs négatives. Ce passage est peut-être nécessaire dans l'histoire humaine, mais ne pas en prendre conscience laisse l'être humain dans une double erreur : celle de sa propre connaissance et de la découverte d'une altérité. Par ailleurs, on ne comprend jamais bien ce qu'on rejette. Or ne pas saisir la nature de ce qui est rejeté ni même les mobiles pour lesquels on les rejette participe d'une absence de recul et constitue une limitation de liberté.

Le rétrécissement de liberté s'exprime par la *sur-sophistication machinale* dont nous avons parlé, que ce soit au niveau individuel ou de la civilisation. Il serait préférable d'adopter une posture éthique face à la *machinité humaine*. Reconnaître la triade *animalité*, *humanité*, *machinité* dans sa dimension projective offre la possibilité d'assumer l'*ambivalence créatrice* humaine qui est en soi une force active et positive d'auto-détermination des valeurs humaines. On est conscient depuis quelques décennies qu'il est important d'assumer le versant animal de l'être humain, sans pour autant sombrer dans l'animalité sauvage, qui reste une représentation humaine fantasmée. Mais la perception de la *machinité* comme projection et comme versant de l'humain est un des nouveaux défis de la civilisation. L'être humain pourrait y faire face de manière créative en sortant de l'alternative attraction/répulsion.

Finalement, l'être humain dans son effort d'humanisation et de civilisation essaie de faire preuve de tempérance. Il tend à être ce milieu tempéré entre le pôle animal conçu comme « chaud » et le pôle *machinal* conçu comme « froid ».

## Références

- ARENDT, Hannah. 1993. *Condition de l'homme moderne*. Translated by Georges Fradier. 1 vols. Agora 24. Paris : Presses Pocket.
- . 2012. *L'humaine condition*. Edited by Philippe Raynaud. 1 vols. Quarto. Paris : Gallimard.
- BESNIER, Jean-Michel. 2012a. *Demain les posthumains : le futur a-t-il encore besoin de nous ?*. 1 vols. Pluriel. Paris : Pluriel.
- . 2012b. *L'homme simplifié : le syndrome de la touche étoile*. 1 vols. Paris : Fayard.
- CASSIRER, Ernst. 1975. Essai sur l'homme. Translated by Norbert Massa. 1 vols. Le sens commun 41. Paris : Ed. de Minuit.

- DESCARTES, René. 1946. *Discours de la méthode, suivi de Méditations métaphysiques*. [Avant-propos de René Ristelhueber]. Montréal, les Éditions Variétés (imprimé à Québec).
- DURAND, Gilbert. 1960. *Les Structures anthropologiques de l'imaginaire introduction à l'archétypologie générale*. Grenoble : impr. Allier.
- ELLUL, Jacques. 1977. *Le Système technicien*. Liberté de l'esprit. Paris : Calmann-Lévy.
- GUILLAUME, Astrid. 2013a. « Humanity and Animality. A Transdisciplinary Approach. » In *Review Human and Social Studies*, éd. Walter de Gruyter, Berlin-Boston. (Version française.), n° 3.
- . 2013b. « Humanité vs Animalité. » *Revue Trimestrielle de La Fondation Droit Animal, Ethique et Sciences*, n° 79 (September). <http://www.fondation-droit-animal.org/documents/revue79.pdf>.
- LA METTRIE, Julien Offray de. 1999. *L'homme-machine*. Collection Folio. Paris : Gallimard.
- MUMFORD, Lewis. 1973. *Le mythe de la machine*. Translated by Léo Dilé. 2 vols. Le Phénomène scientifique. Paris : Fayard.
- . 1976. *Technique et civilisation*. Collections Esprit. Paris : Éditions du Seuil.
- NORDMANN, Alfred. 2005. *Technologies Convergentes, Façonner L'avenir Des Sociétés Européennes*. Rapport du groupe d'experts de haut niveau, Commission européenne. [http://ec.europa.eu/research/social-sciences/pdf/ntw-report-alfred-nordmann\\_fr.pdf](http://ec.europa.eu/research/social-sciences/pdf/ntw-report-alfred-nordmann_fr.pdf).
- ROCCO, Mihail, and William BRAINBRIDGE. 2003. *Converging Technologies for Improving Human Performances : Nanotechnology, Biotechnology, Information Technology and Cognitive Science*. Springer. [http://www.wtec.org/ConvergingTechnologies/Report/NBIC\\_report.pdf](http://www.wtec.org/ConvergingTechnologies/Report/NBIC_report.pdf).

# L'intelligence artificielle

*Dimension prospective*



# Regard prospectif sur l'intelligence artificielle et les interactions émotionnelles humains-machines : bénéfices et risques

**Laurence Devillers**

*La science est un outil puissant.*

*L'usage qu'on en fait dépend de l'homme, pas de l'outil.*

Albert Einstein

L'intelligence des machines est une sorte de quête du Graal. Elle oblige à comprendre notre intelligence d'êtres vivants pour la modéliser. Pour la majorité des chercheurs, le rêve d'une intelligence artificielle rivalisant avec celle des humains, voire la surpassant, est plutôt un fantasme qui n'a rien à voir avec la réalité. On distingue l'intelligence artificielle « faible », qui vise à imiter certaines fonctions de l'intelligence pour répondre à des missions spécifiques — il s'agit des programmes les plus utilisés, la machine ne faisant que donner une impression d'intelligence — et l'intelligence artificielle « forte », qui, quant à elle, vise à faire naître une conscience de soi, à se rapprocher du raisonnement humain en développant la notion d'apprentissage.

Durant la phase d'innovation dans laquelle nous vivons, de plus en plus d'objets embarquant de l'Intelligence Artificielle (IA) sont utilisés pour des tâches jusqu'à présent réservées à l'être humain sans qu'on se demande quel pourrait être l'effet à long terme de ces technologies. Il y aura, en 2020, en moyenne six objets connectés par foyer : le téléphone, l'ordinateur, l'enceinte, la voiture, la montre et les vêtements connectés. L'IA ouvre d'importantes

perspectives, par exemple dans le domaine de la santé avec l'aide au traitement et de meilleurs diagnostics et suivis des patients. Si l'on considère l'intelligence du vivant comme la propension à ressentir, agir et interagir ou être, l'expression « intelligence artificielle » au sujet des machines non conscientes est un oxymore. Cette figure de style allie les deux termes « intelligence » (propre à l'humain) et « artificiel » (propre à la machine) dont le rapprochement est inattendu et crée une formule en apparence contradictoire. Une des différences fondamentales entre les deux est qu'un programme informatique s'appuie sur un substrat abstrait (l'information) s'exécutant sur une machine statique par essence, tandis que le comportement biologique des êtres vivants repose sur un substrat concret organique et dynamique par essence. L'intelligence est souvent décrite telle une faculté d'apprentissage pour s'adapter à l'environnement ou, au contraire, comme une faculté de modifier l'environnement pour l'adapter à ses propres besoins. L'intelligence artificielle n'a rien à voir avec l'intelligence des êtres humains. Le principe d'apprentissage des robots est lié à l'imitation et à l'association. Mais savoir comment une intelligence peut devenir réflexive, revenir sur elle-même et devenir consciente est un grand mystère. C'est une capacité que permet le cerveau humain grâce au langage<sup>362</sup>.

Un agent conversationnel (*chatbot*) est un système informatique qui dialogue de façon autonome avec des utilisateurs en langage naturel. Ce système embarque des modules d'intelligence artificielle dont certains utilisent des algorithmes d'apprentissage machine. Plusieurs milliers de *chatbots* ont été créés ces dernières années, cela pour rendre des services extrêmement variés : améliorer l'expérience client, gérer un outil, donner des conseils médicaux, optimiser la stratégie de comportement etc. Il existe également des *chatbots* sans finalité ou domaine d'application précis susceptibles de dialoguer avec l'homme sur tous les sujets. Ces agents peuvent être embarqués sur des robots.

Un robot est une plateforme qui embarque un grand nombre de logiciels informatiques avec différentes approches algorithmiques en perception, décision et action dans notre environnement. Cette machine comporte trois éléments : des capteurs pour comprendre l'environnement dans lequel elle se trouve, des processeurs qui analysent les paramètres recueillis pour prendre une décision et des actionneurs qui lui permettent d'agir dans le monde réel,

---

<sup>362</sup> E. Morin, *Penser global*, Robert Laffont, 2015. Pour qu'un robot puisse apprendre de son environnement, prendre des décisions et des initiatives, faire sa propre expérience, il est nécessaire qu'il ait un premier niveau de conscience de soi. En ce sens, la phénoménologie a été développée par le philosophe Husserl (1859-1938). Au cœur de la phénoménologie se trouve le concept d'intentionnalité, qui signifie principalement que « la conscience est conscience de quelque chose ».

matériel<sup>363</sup>. Certains robots sont aussi doués de la possibilité de dialoguer... Même si chacun des modules de perception d'objets ou de reconnaissance de visages est entraîné grâce à des algorithmes d'apprentissage machine, l'automatisation de l'ensemble des modules est très complexe à régler et l'adaptation à son environnement émergente. Pour donner au robot la capacité d'apprendre de son environnement de façon autonome, des algorithmes de renforcement, qui nécessitent que l'humain conçoive des métriques de récompenses, sont utilisés. Le robot apprend par essais et erreurs en fonction des récompenses programmées. De façon maladroite et laborieuse, il associe des actions dans le monde et des représentations internes pour réaliser certaines tâches pour lesquelles il est conçu.

Plus de 150 experts, en majorité anglo-saxons, ont donné leurs avis sur la machine pensante, l'intelligence artificielle forte et la conscience à l'occasion du livre *What to Think about Machines that Think* de J. Brockman, paru en 2015.

\* Stanislas Dehaene, cogniticien et neuroscientifique, qui travaille au Collège de France et est l'auteur de *Consciousness and the Brain*, rappelle que, quand Turing inventa l'ordinateur, il confessa qu'il cherchait à imiter l'humain. Selon lui, étudier le cerveau humain est encore notre meilleur moyen de créer des machines pensantes. Il y a cependant deux aspects étudiés en sciences cognitives qui sont peu développés en simulation sur machine. Le premier est un espace de travail global : les programmes actuels sont modulaires, chaque module est indépendant des autres et ne partage pas un espace commun. Le deuxième aspect concerne la théorie de l'esprit : une représentation mentale des personnes avec qui on interagit est codée également dans notre cerveau. Actuellement, les systèmes les plus avancés essayent de donner cette capacité au robot.

\* Nick Bostrom, qui travaille à l'université britannique d'Oxford, dans l'Institut du Futur de l'Humanité, et qui est l'auteur de *SuperIntelligence : Paths, Dangers, Strategies*, publié en 2012, explique dans ce livre que « *La machine est encore peu douée pour penser, sauf dans des domaines spécialisés. Cependant, les machines sont déjà beaucoup plus fortes physiquement et plus rapides que n'importe quelle créature biologique. On ne sait absolument pas quand elle aura une intelligence humaine.* »

\* Stuart Russel, qui mène ses recherches à l'université californienne de Berkeley et est l'auteur de *Artificial Intelligence : A modern Approach*, pense qu' « *il faut construire une intelligence*

---

<sup>363</sup> [http://www.symop.com/wp-content/uploads/2016/10/LivreBlanc\\_Robotique\\_symop\\_BD.pdf](http://www.symop.com/wp-content/uploads/2016/10/LivreBlanc_Robotique_symop_BD.pdf).

*machine alignée sur les valeurs humaines* ». Mais n'est-ce pas dangereux si elle prend des décisions à notre place ? Celui qui décidera des règles aura alors le pouvoir.

\* Max Tegmark, chercheur au MIT (Massachusetts Institute of Technology), au « Future of Life Institute », explique que la question la plus importante en matière d'intelligence artificielle est « *Que voulons-nous en faire ? Et qu'arrivera-t-il aux humains si les machines les remplacent graduellement sur le marché du travail ?* ».

\* Demis Hassabis, président directeur général de Google DeepMind, système qui a battu le champion Lee Sedol au jeu de Go, est préoccupé par l'éthique et la sécurité et trouve que beaucoup de personnes sont trop optimistes au sujet de l'intelligence artificielle en pensant que tout serait possible en peu de temps. Il cite Peter Norving qui trouve que les propos tenus sur l'intelligence artificielle se sont radicalement transformés. Maintenant, au lieu de « *Isn't terrible that Artificial Intelligence is a failure?* », nous entendons dans les médias « *Isn't terrible that Artificial Intelligence is a success?* »

Il ne se passe pas un jour sans annonce de nouvelles découvertes sur le cerveau et sur la manière d'optimiser ses capacités. Notre cerveau se nourrit de toutes nos expériences. Les découvertes sont impressionnantes en médecine et en neurosciences, les avancées plus incroyables les unes que les autres, mais nous n'en savons pas beaucoup plus sur l'organisation de l'ensemble du cerveau. Le projet « *Human Brain Machine* »<sup>364</sup> de l'École Polytechnique Fédérale de Lausanne (EPFL) a pour but de recopier le cerveau d'un mammifère sur une machine... mais nous avons encore beaucoup de progrès à faire pour comprendre le fonctionnement d'un cerveau, ne serait-ce que celui d'une souris ! La reproduction du cerveau de l'Homme sur machine se heurte donc à des difficultés considérables qui sont de nature intrinsèquement biologique, au-delà des difficultés tout aussi réelles pour développer l'intelligence artificielle au niveau nécessaire. Si on considère le problème des matériaux et des fonctions, la machine n'a pas de corps au sens de viscères, de cellules et d'hormones. Elle n'a pas non plus d'intentions, d'émotions, de plaisirs ni de désirs propres. Elle n'a pas de *conatus* au sens de Spinoza (1861), cette intention puissante de se maintenir en vie : « L'effort par lequel chaque chose tend à persévérer dans son être n'est rien d'autre que l'essence actuelle de chaque chose »<sup>365</sup>. Le concept de l'homéostasie, proche du *conatus*, a été évoqué pour la première fois en 1866 par le médecin et physiologiste français Claude Bernard. Ce mécanisme d'homéostasie est propre au vivant et est construit de manière à maintenir la cohérence de ses structures et de ses fonctions et à les réparer en cas d'atteintes extérieures. Le

---

<sup>364</sup> Human Brain Machine, EPFL.

<sup>365</sup> B. Spinoza, *Éthique III*, Proposition VII, 1861.

vivant pourrait être défini par ces deux aspects : l'homéostasie et, bien sûr, la reproduction.

Les avancées de la médecine dite « régénérative » (thérapie génique, cellules souches, greffes, interface cerveau-machine etc.) apportent des solutions ou suscitent des espoirs pour réparer le cerveau mais peu de solutions thérapeutiques existent encore. Comprendre le fonctionnement du cerveau, pour le préserver, augmenter ses performances, le réparer et le guérir, constitue un projet enthousiasmant pour les générations à venir, même si personne ne peut affirmer que nous atteindrons un jour une connaissance totale de cet organe qui fait de chacun de nous un individu singulier.

La question est aussi de savoir jusqu'à quel point il faut s'inspirer de l'être humain pour concevoir un robot utile pour l'être humain.

## I. L'imitation du vivant

La robotique de science-fiction a toujours été hantée par la confusion entre humains et robots. Il y a deux avantages aux robots humanoïdes : leur apparence anthropomorphique facilite l'interaction des humains avec les machines et ils peuvent vivre dans les maisons et utiliser les objets des humains.

Vouloir recopier l'intelligence de l'homme dans une machine est également très narcissique car que connaît-on de notre intelligence ? Nous ne connaissons pas le substrat de la pensée, nous ne sommes conscients que d'une petite partie de nos perceptions et de notre activité cérébrale. Il n'existe pas de terme plus polysémique et sujet à interprétation que celui de « conscience » : il évoque pour certains la conscience de soi, pour d'autres la conscience du prochain, ou encore la conscience phénoménale, la conscience morale etc.

Avec une conception philosophique matérialiste de la vie, on peut considérer que l'ordinateur et le cerveau humain sont des systèmes comparables, capables de manipuler des informations. Les modélisations numériques les plus performantes, comme le *deep learning* (apprentissage profond), s'appuient sur une modélisation simplifiée du neurone (neurone formel) intégrée dans une machine à états discrets simulée sur ordinateur. Le nombre de couches cachées de l'architecture du modèle correspond à la profondeur. Pour l'instant, nous sommes très loin de la complexité du vivant ! Les expériences menées à Neurospin par l'équipe de Stanislas Dehaene, notamment à l'aide d'images subliminales, ont démontré que notre cerveau fonctionne principalement selon un mode non conscient. Les actions routinières, la reconnaissance des visages, des mots, par exemple, sont

réalisées sans recours à la conscience. Afin d'accéder à une conscience, le cerveau humain met en place deux types de traitement de l'information : un premier niveau, baptisé « disponibilité globale », qui correspond au vaste répertoire d'informations, de programmes modulaires que l'on peut convoquer à tout moment pour s'en servir ; et un deuxième type de traitement de l'information qui lui est associé, propre à la conscience humaine : l'auto-monitoring ou auto-évaluation, « soit la capacité à traiter des informations sur soi-même, que l'on peut appeler également métacognition ». Ainsi, le cerveau est capable d'introspection, de contrôler son propre processus et d'obtenir des informations sur lui-même, ce qui le mène à l'autonomie.

Les systèmes actuels d'intelligence artificielle ont la capacité de calculer des corrélations de faits, avec les approches d'apprentissage profond par exemple, de prendre des décisions et d'apprendre, sans en avoir conscience. Certains prototypes de robot ont pourtant déjà des embryons de niveaux de « conscience » comparables à ceux que décrit Stanislas Dehaene. Ils sont simulés par des mécanismes de partage de connaissances et d'introspection. Pour autant, ces machines ne sont pas conscientes comme peut l'être un humain, elles n'ont ni conscience morale ni conscience phénoménale associée à une expérience qualitative telle que la sensation de chaud ou de froid, le sentiment d'anxiété etc. car elles n'ont pas de viscères ni de ressentir, à moins, là encore, de les simuler. Une conscience artificielle, dotée de sentiments, de pensées et de libre arbitre sans programmation humaine, a donc peu de chance d'émerger spontanément avec les architectures actuelles d'ordinateurs.

## II. L'interaction affective avec les robots

La parole des *chatbots* et des robots, quoique générée à travers la démarche d'imitation des conversations humaines, en diffère de manière significative autant par sa structure que par son sens. Cependant, la parole est un élément constitutif de l'identité de l'homme et le fondement de sa vie en société. Les machines parlantes sont ainsi comparées à l'homme de façon immédiate et spontanée. Souvent marquée par le manque de distinction et le brouillage des limites, cette comparaison est nourrie par les mythes et fictions<sup>366</sup>. Elle est aussi susceptible d'attribuer aux *chatbots* un statut social à travers la projection des caractéristiques humaines sur les machines.

Un robot social est une machine artificiellement intelligente grâce à des modèles informatiques conçus par des êtres humains : pour peu qu'on l'ait programmée pour dialoguer, détecter et reconnaître des indices émotionnels

---

<sup>366</sup> L. Devillers, 2017.

et conversationnels et pour s'adapter à l'être humain, voire même simuler de l'humour, cette machine peut sembler chaleureuse. Mais cette simulation affective qui est un domaine de recherche qui a émergé il y a une vingtaine d'années avec les travaux de Rosalind Picard au MIT<sup>367</sup> ne doit pas occulter le fait que la machine a un « organisme » artificiel.

L'homme a tendance à surévaluer les capacités du robot et à lui prêter des intentions qu'il n'a pas, surtout lorsqu'on lui a donné des capacités langagières, émotionnelles et empathiques. Le robot capable de parler avec une voix proche de celle d'un humain est difficilement considéré comme un objet. Il est nécessaire de faire comprendre aux êtres humains qu'ils vont projeter de façon très naturelle des intentions et des émotions sur les machines robotiques. Nous allons souffrir pour un robot qui imitera la souffrance. Mais comment résister à l'empathie ? — d'ailleurs, il ne le faut pas car que deviendrait notre société si nous devenions non émotionnels ?

Les robots, même simples, sont parfois perçus comme capables d'établir des relations avec les humains et semblent même doués d'une présence qui les amène à réclamer notre attention. Les robots sociaux vont partager notre espace, habiter nos maisons, nous aider dans notre travail et notre vie quotidienne et, également, partager avec nous une certaine histoire. Pourquoi ne pas les doter d'humour machine ? L'humour joue un rôle capital dans les relations sociales : il amortit le stress, met en confiance, crée une connivence entre les interlocuteurs. Si vous êtes seul et malheureux, le robot pourrait plaisanter pour vous réconforter ; si vous êtes en colère, il pourrait vous aider à relativiser, dire que la situation n'est pas si terrible. Il pourrait également faire preuve d'autodérision s'il commet des erreurs et s'en rend compte !

Comment ces robots « empathiques » sont-ils accueillis ? Pour le savoir, il est important de mener des études perceptives sur l'interaction humain-machine. Le Limsi-CNRS a conduit de nombreux tests en laboratoire et en Ehpad avec des personnes âgées, ou dans des centres de rééducation avec l'association Approche, dans le cadre du projet BPI ROMEO2, piloté par Softbank robotics<sup>368</sup>. Crée en 1991, l'association Approche a pour principale mission de promouvoir les nouvelles technologies (robotique, électronique, domotique, technologies de l'information et de la communication...) au service des personnes en situation de handicap quels qu'en soient l'âge et le milieu de vie. Nous avons également mené des études autour de scénarios de la vie de tous les jours et de jeux avec l'équipe du Professeur Anne-Sophie Rigaud au Living Lab de l'hôpital Broca<sup>369</sup>. Toutes ces expériences ont montré que les robots étaient plutôt bien acceptés par les patients lorsqu'ils ont le

---

<sup>367</sup> R. Picard, 1997.

<sup>368</sup> Kumar et al., 2014.

<sup>369</sup> Garcia et al., 2017.

temps de les expérimenter. De nombreux travaux de recherche sont encore nécessaires pour étudier la relation à long terme avec ces machines.

Actuellement, les robots ne sont pas réellement autonomes : ils n'ont donc pas de conscience, ni d'émotions, ni de désirs comme les humains. Cependant, nous projetons sur eux de façon très naturelle des capacités humaines<sup>370</sup>. Pour cette raison, il est fondamental de réfléchir à la régulation de leur fonctionnement, mais aussi de démystifier leurs capacités, par l'éducation notamment, et de rendre plus transparents et compréhensibles les programmes informatiques qu'ils embarquent.

### III. Réflexions éthiques

Il est nécessaire d'engager une réflexion sur les problèmes éthiques liés à ces aspects de la coévolution homme-machine, aux niveaux international et national, et de rappeler les limitations fondamentales de la construction automatique du dialogue en analysant notamment les problèmes éthiques posés par le choix des métriques dans la conception des *chatbots* et des robots. La CERNA a rappelé les enjeux éthiques de l'affectivité automatique, de la personnification des *chatbots*/robots émotionnels et de l'imitation du vivant, ainsi que les biais susceptibles d'émerger dans le fonctionnement des *chatbots*/robots à la suite de données d'apprentissage partielles ou sélectives dans plusieurs avis, notamment celui sur la robotique<sup>371</sup> et celui sur l'apprentissage machine<sup>372</sup>.

Il est nécessaire également de s'interroger sur les spécificités socio-culturelles et les divergences dans l'utilisation de l'IA et sur les représentations des humains dans les *chatbots* et les robots à l'intérieur de différentes parties du monde. En tenant compte de ces spécificités, il est nécessaire de s'interroger sur les valeurs impliquées dans l'évaluation éthique des *chatbots*, notamment sur le besoin de transparence, sur l'information que le concepteur doit fournir à l'utilisateur et sur la finalité et les effets non-intentionnels de ces machines qui sont des sujets de recommandations opérationnelles pour les concepteurs et les utilisateurs, ainsi que des sujets de recherche.

Les robots sociaux et affectifs soulèvent de nombreuses questions éthiques, juridiques et sociales. Qui est responsable en cas d'accident : le fabricant, l'acheteur, le thérapeute, l'utilisateur ? Comment réguler leur

---

<sup>370</sup> Nass & Reeves, 1996.

<sup>371</sup> Grinbaum et al., 2016.

<sup>372</sup> Devillers et al., 2017.

fonctionnement ? Faut-il intégrer des règles morales dans leur programmation ? Faut-il contrôler leur utilisation par des permis ? Pour quelles tâches souhaitons-nous créer ces entités artificielles ? Comment préserver notre intimité, nos données personnelles ? Plus de 80 % des chercheurs travaillant sur les algorithmes d'apprentissage automatique et ces objets artificiels sont des hommes, plus de 80 % des objets (*chatbots*, robots) créés ont des noms féminins (comme Alexa, Sofia, Samantha etc.), ont des voix et corps de jeunes femmes et ont des rôles de subalternes (assistantes, infirmières) que l'on peut éteindre facilement. La prédominance des assistantes vocales peut s'expliquer par la prédominance masculine dans les équipes de conception de tels assistants<sup>373</sup>. Cette représentation des femmes dans les objets renforce les stéréotypes de genre dans le monde réel. Il s'agit avant tout de susciter des questionnements sur ces assistants artificiels et leurs apports et risques pour la société, notamment dans les secteurs de la santé, du bien-être et de l'éducation.

Tout système doit être évalué avant d'être mis dans les mains de ses utilisateurs<sup>374</sup>. Comment évaluer une intelligence artificielle qui apprend des humains et s'adapte à eux, ou qui apprend seule ? Peut-on prouver qu'elle se cantonnera aux fonctions pour lesquelles elle a été conçue, qu'elle ne dépassera pas les limites fixées ? Les données que la machine exploite pour son apprentissage la dirigent vers certaines actions. Qui supervisera la sélection de ces données ?

Ces questions prégnantes ne sont évoquées que depuis peu. Les progrès spectaculaires du numérique permettront un jour d'améliorer le bien-être des personnes, à condition de réfléchir non à ce que nous pouvons en faire, mais à ce que nous voulons en faire. C'est pourquoi la plus importante association professionnelle internationale du numérique, l'Institut des ingénieurs électriciens et électroniciens (IEEE), organisation savante mondiale, a lancé une initiative pour réfléchir à une éthique relative aux systèmes autonomes<sup>375</sup>. Dans la foulée, une douzaine de groupes de travail sur les normes et les standards IEEE ont émergé<sup>376</sup>, notamment sur le *nudging* des robots (manipulation incitative). Aux États-Unis, il y a eu également les 23 « *Asilomar AI Principles* » du *Future of Life Institute* signés par 5 000 experts. Puis la déclaration de Montréal pour une IA responsable a été rédigée à l'Université de Montréal, en 2017. La Cerna<sup>377</sup> s'est également emparée de ces sujets en France. Le Japon et l'Allemagne ont également publié des chartes de régulation et d'éthique. Quant aux GAFA, ils ont construit le *Partnership on AI* (PAI) en 2017.

---

<sup>373</sup> West et al., 2019.

<sup>374</sup> Dubuisson & Devillers, 2015.

<sup>375</sup> IEEE, 2016.

<sup>376</sup> Fagela, 2019.

<sup>377</sup> CERNA, 13-19.

Il faut éviter un déficit de confiance mais également une confiance trop aveugle dans les programmes d'intelligence artificielle. Un certain nombre de valeurs éthiques sont importantes : la déontologie et la responsabilité des concepteurs, l'émancipation des utilisateurs, l'évaluation, la transparence, l'explicabilité, la loyauté et l'équité des systèmes, enfin l'étude de la coévolution humain-machine — la machine s'adaptera à l'humain et l'humain à la machine.

Du côté de la Commission européenne, une phase pilote pour tester sept principes éthiques pour une intelligence artificielle de confiance a été annoncée. De juin à novembre 2019, une phase de consultation a permis de recueillir des remarques additionnelles autour de ces sept principes :

*Facteur humain et contrôle humain* : les systèmes d'IA devraient être les vecteurs de sociétés équitables en se mettant au service de l'humain et des droits fondamentaux, sans restreindre ou dévoyer l'autonomie humaine.

*Robustesse et sécurité* : une IA digne de confiance nécessite des algorithmes suffisamment sûrs, fiables et robustes pour gérer les erreurs ou les incohérences dans toutes les phases du cycle de vie des systèmes d'IA.

*Respect de la vie privée et gouvernance des données* : il faut que les citoyens aient la maîtrise totale de leurs données personnelles et que les données les concernant ne soient pas utilisées contre eux à des fins préjudiciables ou discriminatoires.

*Transparence* : la traçabilité des systèmes d'IA doit être assurée.

*Diversité, non-discrimination et équité* : les systèmes d'IA devraient prendre en compte tout l'éventail des capacités, aptitudes et besoins humains, et leur accessibilité devrait être garantie.

*Bien-être sociétal et environnemental* : les systèmes d'IA devraient être utilisés pour soutenir des évolutions sociales positives et renforcer la durabilité et la responsabilité écologique.

*Responsabilisation* : il convient de mettre en place des mécanismes pour garantir la responsabilité à l'égard des systèmes d'IA et de leurs résultats, et de les soumettre à une obligation de rendre des comptes.

L'intelligence artificielle peut apporter de meilleurs diagnostics en santé, des outils de stimulation, des instruments de détection des comportements anormaux, une meilleure assistance (notamment pour le handicap ou la perte d'autonomie). Les machines pourront sûrement apprendre seules, mais ne pourront pas savoir si ce qu'elles ont appris est intéressant, car elles n'ont pas de conscience. Le contrôle par des humains sera toujours primordial. Il est nécessaire de développer des cadres éthiques pour les robots et agents sociaux, ainsi que des normes, notamment en santé, et de comprendre le niveau de

complémentarité humain-machine. Nous avons besoin de démystifier, de former à l'intelligence artificielle et de remettre au centre de la conception de ces systèmes robotiques les valeurs de l'humain. Il est urgent de travailler en interdisciplinarité pour construire ensemble des IA éthiques *by design*.

## Références

Association Approche <http://www.approche-asso.com>

Comité CERNA d'Allistène, (2013-19) : Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene, [cerna-ethics-allistene.org](http://cerna-ethics-allistene.org)

L. Devillers, (2017) *Des robots et des hommes : mythes, fantasmes et réalité*, Plon

L. Devillers et al., (2017), *Éthique de la recherche en apprentissage machine*, CERNA

G. Dubuisson Duplessis et L. Devillers, (2015) « Towards the consideration of dialogue activities in engagement measures for human-robot social interaction », *International Conference on Intelligent Robots and Systems, Designing and Evaluating Social Robots for Public Settings Workshop*, 2015, pp. 19-24

IEEE et l'éthique des systèmes autonomes, (2016) [standards.ieee.org/develop/indconn/ec/autonomous\\_systems.html](http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html)

D. Fagella, (2019), « Initiative IEEE standards et normes » : [techemergence.com/ethics-artificial-intelligence-business-leaders/](http://techemergence.com/ethics-artificial-intelligence-business-leaders/)

Grinbaum, R. Chatila, L. Devillers, J.-G. Ganascia, C. Tessier et M. Dauchet, (2017), « Ethics in robotics research », *IEEE Robotics & Automation Magazine*, vol. 24 n° 3, pp. 139-145

K. Pandey, R. Gelin, R. Alami, R. Vitry, A. Buendia, R. Meertens, M. Chetouani, L. Devillers, M. Tahon, D. Filliat, Y. Grenier, M. Maazaoui, A. Kheddar, F. Lerasle et L. Fitte Duval, (2014) « Ethical considerations and feedback from social human-robot interaction with elderly people », AIC.

R. Reeves et C. Nass, (1996), *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Cambridge University Press

M. West, R. Kraut, & H. Ei Chew (2019). *I'd blush if I could: closing gender divides in digital skills through education*



## Table des matières

Sommaire .....	7
Les auteurs .....	9

### Introduction

<b>L'intelligence de l'intelligence artificielle (par Boris Barraud) .....</b>	15
--	----

I. Une problématique humaine : l'intelligence de la machine au service de l'intelligence de l'homme.....	24
II. Une problématique technique : au cœur des cerveaux artificiels (apprentissage automatique et apprentissage profond).....	33
III. Une problématique sociale, politique et économique : l'intelligence artificielle et les données, moteur et carburant d'un changement de civilisation .....	45
IV. Une problématique scientifique : le besoin d'une science grand public de l'intelligence artificielle.....	58

### L'intelligence artificielle : *dimension historique*

<b>Une histoire de l'intelligence artificielle (par Frédéric Fürst) .....</b>	81
---	----

I. La naissance de l'intelligence artificielle.....	81
II. La pré-histoire de l'intelligence artificielle .....	85
III. L'âge d'Or de l'intelligence artificielle .....	90

IV. Le premier hiver de l'intelligence artificielle.....	96
V. Le deuxième cycle de l'IA .....	102
VI. L'IA au XXIe siècle.....	104
Références.....	105

## *L'intelligence artificielle : dimension politique*

<b>Nouvelle gouvernementalité et fin du politique – Libéralisme, contrôle social et privatisation du pouvoir (par Charles Thibout) .....</b>	109
I. La technique au secours du libéralisme.....	110
A. La communication contre l'aliénation .....	111
B. Contre Nietzsche et Freud : le triomphe de la raison et de la liberté .....	112
II. La gouvernementalité selon l'intelligence artificielle .....	114
A. Extension du domaine de l'administration technicienne .....	114
B. De la technologie politique des corps au « Panoptique inversé »....	116
III. Vers la privatisation du pouvoir ?.....	119
A. Entre nécessité et dépendance, États et firmes numériques.....	119
B. La capacité d'influence politique des producteurs d'IA .....	121
Conclusion : Les jalons d'une politique de l'IA .....	122

## *L'intelligence artificielle : dimension sociologique*

<b>Une intelligence artificielle au travail – Cinq histoires d'Homme (par Yann Ferguson) .....</b>	129
I. L'homme remplacé : des routines aux algorithmes, de l'IA à l'emploi.....	131
II. L'homme dominé : singularité technologique, passivité humaine ....	133

III. L'homme augmenté : former des « centaures ».....	135
IV. L'humanité divisée : l'intelligence au cœur de la fracture sociale ...	137
V. L'homme réhumanisé : la redécouverte du proprement humain.....	139
Conclusion : Descartes et Taylor automatisés .....	141
Références.....	142

## *L'intelligence artificielle : dimension économique*

<b>Vers une quatrième révolution industrielle ? L'IA comme moteur, les données comme carburant (par Boris Barraud) .....</b>	147
I. L'explosion de la demande d'IA et du marché de l'IA .....	150
II. L'IA, nouveau facteur de production et nouveau vecteur de croissance.....	154
III. Les effets de l'IA sur le marché du travail : destruction ou réorientation des emplois ? .....	159
IV. Concentrations dans le secteur de l'économie des données .....	161
V. Leaders et suiveurs de l'économie des données .....	165
VI. La stratégie IA de la France.....	169
VII. Souplesse, flexibilité, adaptabilité.....	171
Conclusion .....	173

## *L'intelligence artificielle : dimension anthropologique*

<b>Une intelligence artificielle anthropocentréé (par Magalie Ochs) .....</b>	177
I. Une intelligence socio-émotionnelle artificielle.....	179

II. Améliorer l'humain à travers les machines : l'informatique prosociale .....	182
III. Stéréotypes et discriminations du monde réel au monde virtuel .....	184
Conclusion .....	187
Références.....	188

### *L'intelligence artificielle : dimension juridique*

<b>Les intelligences artificielles du droit (par Boris Barraud) .....</b>	195
I. Des intelligences artificielles à l'origine du droit .....	196
II. Des intelligences artificielles au service du droit.....	199
Conclusion .....	202

### *L'intelligence artificielle : dimension chinoise*

<b>L'intelligence artificielle en Chine : un état des lieux (par Aifang Ma) .....</b>	207
I. Les politiques chinoises pour le développement de l'intelligence artificielle.....	208
A. Proclamation de stratégies précises et mesurables .....	209
B. Investissements colossaux et soutien public constant.....	211
C. Achat de technologies étrangères .....	213
D. Création d'un environnement social favorable .....	213
II. Un état des lieux du développement de l'intelligence artificielle en Chine.....	214
A. Dynamique des start-up et des BATX.....	214
B. E-santé .....	217
C. Préservation de l'ordre public .....	218
D. Protection de l'environnement .....	220

E. Défense nationale.....	222
F. Rivalité avec les Etats-Unis .....	224
 III. Les visées stratégiques des investissements colossaux en IA de la Chine .....	226
Conclusion .....	229
Références.....	231
 <b>Un regard chinois sur la relation Homme-Machine : le cas du jeu de go (par JianMing Zhou et Yu Li) .....</b>	235
I. Une pierre génère mille vagues .....	236
II. Les particularités d'AlphaGo parmi les IA joueuses .....	236
III. Les décisions d'AlphaGo et de l'Homme en fonction des configurations du go.....	238
IV. La technique, l'art et le tao du go.....	240
Conclusion .....	241
 <b>L'intelligence artificielle : dimension socio-scientifique</b>	
 <b>Quelques repères sur l'apport des sciences sociales pour l'étude de l'intelligence artificielle (par Jean-Sébastien Vayre) .....</b>	245
I. L'enquête épistémologique .....	247
II. L'enquête pragmatique .....	250
III. L'enquête interactionniste .....	253
IV. L'enquête structurelle .....	256
Conclusion .....	259
Références.....	261

## L'intelligence artificielle : *dimension philosophique*

<b>La triade animalité, humanité, <i>machinité</i> – De la Sauvagerie animale à la <i>Sur-sophistication</i> de la machine (par Anne-Laure Thessard) .....</b>	267
I. Humanité et projection : l'animalité, l'animal-machine, la <i>machinité</i> .....	272
II. La <i>machinité</i> : la nouvelle menace pour l'humain.....	276
III. De la sauvagerie à la <i>sur-sophistication</i> machinale : <i>machinité</i> et enjeux de civilisation .....	279
Références .....	283

## L'intelligence artificielle : *dimension prospective*

<b>Regard prospectif sur l'intelligence artificielle et les interactions émotionnelles humains-machines : bénéfices et risques (par Laurence Devillers) .....</b>	287
I. L'imitation du vivant.....	291
II. L'interaction affective avec les robots.....	292
III. Réflexions éthiques .....	294
Références .....	297





## **STRUCTURES ÉDITORIALES DU GROUPE L'HARMATTAN**

### **L'HARMATTAN ITALIE**

Via degli Artisti, 15  
10124 Torino  
[harmattan.italia@gmail.com](mailto:harmattan.italia@gmail.com)

### **L'HARMATTAN HONGRIE**

Kossuth l. u. 14-16.  
1053 Budapest  
[harmattan@harmattan.hu](mailto:harmattan@harmattan.hu)

---

### **L'HARMATTAN SÉNÉGAL**

10 VDN en face Mermoz  
BP 45034 Dakar-Fann  
[senharmattan@gmail.com](mailto:senharmattan@gmail.com)

### **L'HARMATTAN MALI**

Sirakoro-Meguetana V31  
Bamako  
[syllaka@yahoo.fr](mailto:syllaka@yahoo.fr)

### **L'HARMATTAN CAMEROUN**

TSINGA/FECAFOOT  
BP 11486 Yaoundé  
[inkoukam@gmail.com](mailto:inkoukam@gmail.com)

### **L'HARMATTAN TOGO**

Djidjole – Lomé  
Maison Amela  
face EPP BATOME  
[ddamela@aol.com](mailto:ddamela@aol.com)

### **L'HARMATTAN BURKINA FASO**

Achille Somé – [tengnule@hotmail.fr](mailto:tengnule@hotmail.fr)

### **L'HARMATTAN CÔTE D'IVOIRE**

Résidence Karl – Cité des Arts  
Abidjan-Cocody  
03 BP 1588 Abidjan  
[espace\\_harmattan.ci@hotmail.fr](mailto:espace_harmattan.ci@hotmail.fr)

### **L'HARMATTAN RD Congo**

185, avenue Nyangwe  
Commune de Lingwala – Kinshasa  
[matangilamusadila@yahoo.fr](mailto:matangilamusadila@yahoo.fr)

### **L'HARMATTAN ALGÉRIE**

22, rue Moulay-Mohamed  
31000 Oran  
[info2@harmattan-algerie.com](mailto:info2@harmattan-algerie.com)

### **L'HARMATTAN CONGO**

67, boulevard Denis-Sassou-N'Gesso  
BP 2874 Brazzaville  
[harmattan.congo@yahoo.fr](mailto:harmattan.congo@yahoo.fr)

### **L'HARMATTAN MAROC**

5, rue Ferrane-Kouicha, Talaâ-Elkbira  
Chrablyine, Fès-Médine  
30000 Fès  
[harmattan.maroc@gmail.com](mailto:harmattan.maroc@gmail.com)

---

## **NOS LIBRAIRIES EN FRANCE**

### **LIBRAIRIE INTERNATIONALE**

16, rue des Écoles – 75005 Paris  
[librairie.internationale@harmattan.fr](mailto:librairie.internationale@harmattan.fr)  
01 40 46 79 11  
[www.librairieharmattan.com](http://www.librairieharmattan.com)

### **LIB. SCIENCES HUMAINES & HISTOIRE**

21, rue des Écoles – 75005 Paris  
[librairie.sh@harmattan.fr](mailto:librairie.sh@harmattan.fr)  
01 46 34 13 71  
[www.librairieharmattansh.com](http://www.librairieharmattansh.com)

### **LIBRAIRIE L'ESPACE HARMATTAN**

21 bis, rue des Écoles – 75005 Paris  
[librairie.espace@harmattan.fr](mailto:librairie.espace@harmattan.fr)  
01 43 29 49 42

### **LIB. MÉDiterranée & Moyen-Orient**

7, rue des Carmes – 75005 Paris  
[librairie.mediterranee@harmattan.fr](mailto:librairie.mediterranee@harmattan.fr)  
01 43 29 71 15

### **LIBRAIRIE LE LUCERNAIRE**

53, rue Notre-Dame-des-Champs – 75006 Paris  
[librairie@lucernaire.fr](mailto:librairie@lucernaire.fr)  
01 42 22 67 13

