idées reçues

Jean-Gabriel Ganascia

Intelligence artificielle

vers une domination programmée ?

Le Cavalier Bleu
EDITIONS

Intelligence artificielle

vers une domination programmée ?

Intelligence artificielle

vers une domination programmée ?

Jean-Gabriel Ganascia

2e édition revue et augmentée

Issues de la tradition ou de l'air du temps, mêlant souvent vrai et faux, les idées reçues sont dans toutes les têtes. Les auteur·e·s les prennent pour point de départ et apportent ici un éclairage distancié et approfondi sur ce que l'on sait ou croit savoir.

Le Cavalier Bleu
EDITIONS

Jean-Gabriel Ganascia

Suite à des études de physique et de philosophie, Jean-Gabriel Ganascia s'est spécialisé d'abord en intelligence artificielle et en apprentissage machine puis en modélisation cognitive. Aujourd'hui il est professeur à l'université Pierre et Marie Curie, *EurAI Fellow* et membre de l'Institut universitaire de France. Il a dirigé pendant 12 ans le diplôme d'études approfondies IARFA (Intelligence artificielle, reconnaissance des formes et applications). Il a aussi été chargé de mission à la direction du CNRS avant de diriger le programme de recherches coordonnées « Sciences cognitives » et d'animer le groupement d'intérêt scientifique « Sciences de la cognition ». Actuellement, il pilote l'équipe ACASA (Agents cognitifs et apprentissage symbolique automatique) au sein du laboratoire d'informatique de Paris-VI. Il est aussi le directeur adjoint du Labex OBVIL où il poursuit, en collaboration avec les équipes de littérature de l'université Paris-Sorbonne, des recherches sur le versant littéraire des humanités numériques. Enfin, il est président du comité d'éthique du CNRS (COMETS) et membre de la CERNA (Commission de réflexion sur l'éthique de la recherche dans les sciences du numérique d'Allistène).

Du même auteur

– *L'Âme machine*, Le Seuil, 1990.
– *L'Intelligence artificielle*, Flammarion, 1993.
– *2001, l'odyssée de l'esprit*, Flammarion, 1999.
– *Gédéon, ou les aventures extravagantes d'un expérimentateur en chambre*, Éditions du Pommier, 2002.
– *Sciences cognitives*, Éditions du Pommier, 2006.
– *Voir et pouvoir : qui nous surveille ?*, Éditions du Pommier, 2009.
– *Le Mythe de la Singularité : faut-il craindre l'intelligence artificielle ?*, Le Seuil, 2017.

sommaire

Des machines et des hommes

L'avenir de l'intelligence artificielle

Annexes

définition

Intelligence artificielle n. f.

La discipline est officiellement née en 1956, au « Dartmouth College », Hanover, New Hampshire aux États-Unis, lors d'une école d'été organisée par quatre chercheurs : John McCarthy, Marvin Minsky, Nicolas Rochester et Claude Shannon. L'année précédente, en 1955, le plus jeune d'entre eux, John McCarthy, avait adressé une demande de subvention en leurs quatre noms à la fondation Rockefeller ; lui, et son comparse Marvin Minsky, étaient âgés de moins de trente ans lorsqu'ils inventèrent le terme d'intelligence artificielle pour frapper les esprits. Et ce terme fit fortune puisque leur projet fut financé et qu'on l'emploie encore pour désigner la discipline informatique qui vise à fabriquer des machines simulant une à une les différentes fonctions de l'intelligence.

Pour John McCarthy et Marvin Minsky comme pour les autres promoteurs de l'école d'été du Dartmouth College, l'intelligence artificielle reposait sur la conjecture selon laquelle toutes les facultés cognitives, en particulier le raisonnement, le calcul, la perception, la mémorisation, voire même la découverte scientifique ou la créativité artistique, pourraient être décrites avec une précision telle qu'il devrait être possible de les reproduire à l'aide d'un ordinateur. Soixante ans plus tard, beaucoup de progrès ont été réalisés dans cette perspective. Désormais, les ordinateurs se sont disséminés partout, dans toutes les activités quotidiennes. Une machine a vaincu, à plusieurs reprises, le champion du monde en titre au jeu d'échecs et même, plus récemment, l'un des meilleurs joueurs au monde au jeu de go ; d'autres démontrent ou aident à démontrer des théorèmes mathématiques ; on construit automatiquement des connaissances à partir de masses immenses de données *(Big Data)*. Grâce à cela,

des automates reconnaissent la parole articulée et comprennent des textes écrits en langage naturel ; des voitures se conduisent seules ; des robots font la guerre à la place des hommes ; certains scientifiques cherchent même à vaincre la mort en déterminant les mécanismes du vieillissement… Non seulement, la plupart des dimensions de l'intelligence – sauf peut-être l'humour – font l'objet d'analyses et de reconstructions rationnelles avec des ordinateurs, mais de plus les machines outrepassent nos facultés cognitives dans la plupart des domaines, ce qui fait craindre à certains un risque pour le futur de l'humanité. En dépit des progrès époustouflants enregistrés ces dernières années, l'étude de l'intelligence artificielle repose toujours sur la même conjecture que rien, jusqu'à présent, n'a permis ni de démentir, ni de démontrer irréfutablement.

introduction

Telle une étincelle, l'esprit jaillit parfois de l'entrechoquement de mots contraires. Les ardents et subtils rhétoriqueurs de l'âge baroque le savaient et en usaient à merveille ; les informaticiens contemporains, nourris de syllogismes et de hamburgers, ne l'ont pas oublié… Les succès qu'alimentèrent les controverses nées autour de l'intelligence artificielle leur ont donné raison : l'accolement des deux mots « intelligence » et « artificielle » fait toujours scandale, à défaut de faire recette.

Et pourtant, à bien y réfléchir, les termes sont tout à fait appropriés : au sens étymologique, l'intelligence artificielle est bien un « artifice », c'est-à-dire un art consommé qui fait illusion en produisant des leurres fabriqués tout exprès pour nous tromper, en faisant accroire que les machines seraient effectivement intelligentes.

Qu'une machine interprète les ordres que nous lui donnons en langue ordinaire, par écrit ou oralement, pour s'exécuter conformément à nos souhaits ; qu'elle pose quelques questions pertinentes avant de suggérer un diagnostic médical ; qu'elle localise d'elle-même les pièces défaillantes d'une voiture atteinte de hoquet ; qu'elle démontre quelques théorèmes mathématiques inédits ; qu'elle reconnaisse des cellules malignes dans une coupe biologique grossie plusieurs centaines de fois au microscope ; qu'elle conduise une voiture au milieu de la circulation ; qu'elle repère, dans le flux des questions posées par les internautes à un moteur de recherche, la manifestation

des symptômes propres à la naissance d'une épidémie ; qu'elle joue au jeu de go et gagne une partie contre l'un des meilleurs joueurs au monde… l'intelligence artificielle est là, et tous s'exclament : « Ô prodige ! ». Serait-ce que les machines, au terme de longs calculs, seraient vraiment devenues intelligentes, et qu'elles posséderaient un esprit, voire que les ingénieurs, à force de les instruire, les auraient dotées d'une conscience ? Point n'est besoin d'aller jusque-là, et d'ailleurs, personne n'est vraiment dupe. Il suffit qu'un ensemble de techniques mises au point par des informaticiens simule des capacités cognitives ordinaires de compréhension du langage naturel, de reconnaissance de la parole, de raisonnement, de résolution de problèmes, de vision, de planification, de jeux, d'apprentissage…

Ces techniques font toutes intervenir des opérations informatiques ordinaires sur des chaînes de caractères, c'est-à-dire sur des mots ou, plus exactement, sur des textes qui symbolisent des sons, des images, des sensations, des états d'esprit… Comme les autres techniques de l'informatique, elles font appel à la logique, aux mathématiques discrètes*, à l'algorithmique, à l'optimisation, à la programmation…

Parfois, le rapprochement des mots « intelligence » et « artificielle » fait resurgir d'un lointain passé de vieux mythes, tels ceux du Golem de Prague, des automates d'Artus de Bretagne ou de l'Ève future de Villiers de l'Isle-Adam, avec leur cortège de légendes et de maléfices. Or, en dépit des craintes que beaucoup nourrissent et des déclarations enflammées d'hommes investis qui abusent de leur autorité, comme Stephen Hawking, Bill Gates ou Elon Musk, il n'en est rien. Les machines fabriquées par l'intelligence artificielle ne possèdent pas, par elles-mêmes, la capacité de prendre le pouvoir sur l'espèce humaine et de la réduire à l'esclavage ;

* *Les mots signalés par un astérisque renvoient au glossaire en fin d'ouvrage.*

d'ailleurs, pour se prémunir de leurs dangers, il suffit de les débrancher.

De plus, l'intelligence artificielle ne vise aucunement à destituer l'homme de son privilège de penser, pour lui substituer une machine pensante. Elle ne bâtit que des théâtres imaginaires, où se meuvent des personnages chimériques dotés d'aptitudes partielles. Elle n'est qu'une intelligence fabriquée au moyen de techniques informatiques ; autrement dit, elle n'est qu'une « intelligence artificielle »...

Le Golem

La tradition cabalistique juive rapporte l'existence d'une statue d'argile fabriquée par le rabbin Loew, plus connu sous le nom de « Maharal de Prague », vers la fin du XVI^e^ siècle. À l'instar des ordinateurs contemporains, cette machine s'animait lorsqu'on passait un message derrière ses dents. Usuellement, elle vaquait aux occupations domestiques quotidiennes, comme un serviteur zélé et assidu.

Beaucoup de légendes ont couru autour de cette statue extraordinaire. Selon l'une d'entre elles, le rabbin Loew aurait oublié, un samedi, jour de prière, d'enlever le message derrière les dents du Golem et celui-ci aurait commencé à s'agiter, à crier et à effrayer tous les voisins pendant que son maître remplissait ses devoirs saints à la synagogue. De retour chez lui, le rabbin Loew aurait détruit son œuvre, de peur qu'elle ne recommence à prendre de fâcheuses initiatives.

Selon une autre légende, sur le front du Golem était écrit le mot *emeth*, qui signifie « vérité » en hébreu ; or, on dit qu'un jour celui-ci aurait pris un couteau pour effacer la première lettre du mot *emeth*, ce qui aurait donné le mot *meth*, soit « mort » en hébreu...

Il résulte de toutes ces mythologies une ambivalence du Golem, qui annonce celle de la technique contemporaine. D'un côté, le rabbin Loew, capable, par son savoir, de fabriquer un objet si perfectionné, fut grandement loué, et même vénéré, au point que le fauteuil sur lequel il s'asseyait est toujours visible dans la synagogue Vieille-Nouvelle de Prague. D'un autre côté, un tel Golem risque parfois d'échapper à ses

maîtres et créateurs, lesquels doivent toujours se garder d'une telle éventualité.

À cet égard, il n'est pas anodin que le père de la cybernétique, Norbert Wiener ait intitulé *God and Golem, Inc.: A Comment on Certain Points Where Cybernetics Impinges on Religion* (« Dieu et le Golem : Un commentaire sur certains points où la cybernétique empiète sur la religion) son dernier ouvrage publié en 1964, l'année de sa mort.

Pour plus d'informations sur le Golem, on pourra lire *Le Golem*, par Moshe Idel, Henri Atlan, et Cyril Aslanof, Le Cerf éditions, 1992.

Étapes du développement de l'intelligence artificielle

Née en 1956, l'intelligence artificielle a connu de nombreuses évolutions au cours de sa courte existence. On peut les résumer en six étapes.

i. Le temps des prophètes

Tout d'abord, dans l'euphorie des origines, les chercheurs se sont laissés aller à des déclarations un peu inconsidérées qu'on leur a beaucoup reprochées par la suite. C'est ainsi qu'en 1958, Herbert Simon, qui deviendra par la suite prix Nobel d'économie, a déclaré que d'ici dix ans les machines seraient championnes du monde aux échecs, si elles n'étaient pas exclues des compétitions internationales. Et que, toujours d'ici dix ans, elles démontreraient des théorèmes originaux en mathématiques, qu'elles composeraient de la musique douée d'une indéniable valeur esthétique, que les théories psychologiques prendraient toutes la forme de programmes informatiques, etc.

ii. Les années sombres

Au milieu des années 1960, les progrès tardèrent ; un enfant de dix ans a battu un ordinateur au jeu d'échecs en 1965 ; un rapport commandé par le sénat américain fit état, en 1966, des limitations intrinsèques de la traduction automatique. L'intelligence artificielle eut alors mauvaise presse, et c'est ainsi que commencèrent quelques années sombres.

iii. L'intelligence artificielle sémantique

Les travaux ne s'interrompirent pas pour autant, mais on axa les recherches dans de nouvelles directions. On s'intéressa à la mémoire, aux mécanismes de compréhension, que l'on chercha à simuler sur un ordinateur, et au rôle de la connaissance dans le raisonnement. C'est ce qui donna naissance aux techniques de représentation des connaissances, qui se développèrent considérablement dans le milieu des années 1970, avec entre autre les réseaux sémantiques et les « cadres de données ». Cela conduisit aussi à développer des systèmes dits experts*, parce qu'ils recourraient au savoir d'hommes de métiers pour reproduire leurs raisonnements. Ces derniers suscitèrent d'énormes espoirs au début des années 1980.

iv. Néo-connexionnisme* et apprentissage machine

Parallèlement à l'essor de l'intelligence artificielle au début des années 1980, les techniques issues de la cybernétique* et du connexionnisme se perfectionnèrent, s'affranchirent de leurs limitations initiales et firent l'objet de multiples formalisations mathématiques. Cela donna naissance à de nombreux développements théoriques puis à des applications industrielles, où les approches se combinèrent pour donner des systèmes hybrides, faisant côtoyer des techniques issues de l'intelligence artificielle, de la recherche opérationnelle, de la cybernétique, de la théorie des systèmes, de la vie artificielle, de l'apprentissage statistique ou de la programmation dynamique.

v. De l'intelligence artificielle à l'informatique animiste...

À partir de la fin des années 1990, on coupla l'intelligence artificielle à la robotique et aux interfaces homme-machine, de façon à produire des agents intelligents qui suggèrent la présence d'un autre. Plus généralement, les réactions des machines usuelles sont désormais calculées de façon à provoquer en nous, à leur contact, l'illusion d'une intelligence les animant, c'est-à-dire d'une âme au sens aristotélicien de « souffle qui anime ». Cette tendance actuelle de l'intelligence artificielle peut éventuellement se caractériser comme une forme d'animisme informatique en cela qu'elle s'emploie à susciter la projection d'un souffle de vie sur les objets quotidiens de notre environnement.

vi. Renaissance de l'intelligence artificielle

Depuis environ 2010, la puissance des machines permet d'exploiter des grandes masses de données (ce que l'on appelle couramment les *Big Data*) avec des techniques d'apprentissage machine qui se fondent sur le recours à de l'apprentissage par renforcement ou à des réseaux de neurones formels, c'est-à-dire à des techniques relativement anciennes que l'on déploie aujourd'hui sur des architectures de dimensions beaucoup plus conséquentes qu'auparavant. Les applications très fructueuses de ces techniques à tous les domaines de l'intelligence artificielle (reconnaissance de la parole, vision, compréhension du langage naturelle, pilotage automatique de voiture, etc.) conduisent à parler d'une renaissance de l'intelligence artificielle qui bouleverse désormais tous les secteurs d'activités (commerce, industrie, banque, assurances, robotique, etc.) en modifiant les métiers, les rôles et les pouvoirs.

L'INTELLIGENCE ARTIFICIELLE, QUI ET QUOI ?

« Alan Turing est l'inventeur de l'intelligence artificielle. »

Dans un article qui a eu une immense influence, Alan Turing a soutenu qu'il était possible de concevoir une expérience prouvant que l'intelligence de l'ordinateur ne pouvait pas être distinguée de celle d'un être humain. Le pari de Turing a éveillé l'ambition de l'intelligence artificielle.

« L'Ordinateur et l'intelligence », site de Michel Volle

Dès 1936, à l'âge de 24 ans, avec les machines dites de Turing, Alan Turing jette les fondements théoriques de l'informatique en établissant un pont entre une formalisation mathématique du calcul et les automates à états finis, autrement dit, les ordinateurs. Il démontre alors qu'une machine très simple est à même de simuler le comportement de n'importe quel ordinateur. Quelques années plus tard, pendant la Seconde Guerre mondiale, il rentre dans les services de renseignement anglais où il emploie ses talents de mathématicien au décryptage des messages ennemis interceptés sur les ondes. Il fait alors appel aux techniques de l'électronique naissante pour fabriquer des calculateurs rapides. Après la guerre, il contribue à la construction d'un des premiers ordinateurs électroniques, puis il poursuit des travaux plus spéculatifs sur les capacités des machines futures à penser, et il préfigure ainsi ce que sera l'intelligence artificielle. Il travaille ensuite sur des simulations informatiques de la croissance des cellules biologiques pour apporter une

contribution originale à la compréhension de la morphogenèse des organismes vivants. Aujourd'hui, ses travaux font toujours l'objet de bien des discussions et alimentent des débats scientifiques enflammés dans les communautés de l'intelligence artificielle et de la biologie. C'est tout particulièrement le cas du test de Turing, qui tente d'apporter une réponse expérimentale à une question souvent rebattue et un peu académique, mais toujours stimulante : « Une machine peut-elle penser ? » Turing imagine une mise en scène, le jeu de l'imitation, dans lequel un interrogateur tente de discerner une femme d'un homme qui travestit ses réponses pour ressembler à une femme. Tout l'attrait tient au dispositif télématique par l'intermédiaire duquel les messages transitent, les personnages ne communiquant que par l'écrit, sans accéder ni à la voix, ni au visage de leurs interlocuteurs. Hormis l'apparence physique, existe-t-il une différence entre l'homme et la femme dans l'ordre de l'intelligence ? Cette question subsiste certainement dans l'esprit d'Alan Turing et le jeu de l'imitation y apportera peut-être une réponse. Mais dans ses articles scientifiques, il la double d'une autre question : existe-t-il, entre l'homme et l'ordinateur, une différence dans l'ordre de l'intelligence, en dépit de leur différence de constitution physique ? Et, pour tenter d'y répondre, il superpose à la première simulation de la femme par l'homme, une seconde simulation en remplaçant, à l'insu de l'interrogateur, l'homme qui imite la femme par un ordinateur qui imite l'homme qui imite la femme. Turing prédit, en 1950, que d'ici 50 ans – c'est-à-dire en l'an 2000 – l'interrogateur n'aura pas plus de 70 % de chance de percer la supercherie en jouant au jeu de l'imitation contre un ordinateur pendant cinq minutes.

De nombreux informaticiens réalisent aujourd'hui des automates qui prétendent fourvoyer les hommes jouant au jeu de l'imitation et, en conséquence, passer ledit test de Turing. On appelle ces automates des « *chatbots* » par contraction de *chat* – bavarder en anglais – et de « robot ». Il existe même un prix, le prix Loebner, qui récompense tous les ans le robot bavard, c'est-à-dire le « chatbot », le plus convaincant. Au-delà de cette conception pragmatique et empirique, certains chercheurs imaginent un test de Turing qualifié de « total » où la machine ne se distinguerait plus du tout d'un homme (ou d'une femme)... Sorti sur les écrans en 2015, le film d'Alex Garland *Ex Machina* illustre parfaitement cette nouvelle perspective où le test de Turing se trouve en quelque sorte inversé : un robot à l'image non pas d'un homme, mais d'une femme, persuade son interlocuteur à l'issue de longues scènes de séduction, qu'en dépit des apparences, seule une vraie femme, cachée dans la machine, peut l'animer...

L'invention de ce test d'intelligence pour les machines fait-elle de Turing un des précurseurs de l'intelligence artificielle ? Certainement, car Turing a imaginé ce que serait l'intelligence des machines et il a répondu à toutes les objections que l'on opposait – et que l'on oppose toujours – à l'idée qu'une machine puisse penser. Cependant, il existe bien d'autres penseurs qui pourraient figurer au rang de précurseurs. Ainsi en va-t-il de Leibniz qui conçut, au XVII^e^ siècle, une machine à raisonner. Il s'ensuit qu'il est difficile d'affirmer que Turing est l'inventeur de l'intelligence artificielle. Qui plus est ce n'est pas Turing qui a inventé le terme « intelligence artificielle » ; il est mort d'ailleurs avant que ce mot n'existe. Ce n'est pas lui non plus qui est à la

source des outils développés ces soixante dernières années pour réaliser ces machines pensantes dont il avait eu l'intuition. Enfin, il existe beaucoup de dimensions de l'intelligence artificielle qui échappent à la réalisation de *chatbots* et que Turing a volontairement éludées dans ses premiers articles sur l'intelligence des machines. C'est en particulier le cas de la simulation de la perception à partir de flux de sensations, par exemple, de la reconnaissance de la parole ou de la vision. C'est la raison pour laquelle ledit test de Turing a souvent été critiqué, parce que réduit aux seules dimensions symboliques de l'intelligence.

En dépit de ces réserves, on doit noter qu'Alan Turing mit l'accent sur les dimensions essentielles de ce qui fera l'objet des investigations ultérieures des chercheurs en intelligence artificielle. Plus précisément, dans les deux articles qu'il écrivit en 1947 et en 1950 sur l'intelligence des machines, il insista sur le rôle central que jouent les connaissances dans la réalisation d'une machine intelligente, c'est-à-dire d'une machine capable de jouer au jeu de l'imitation et de tromper un homme. Il anticipa ainsi ce que seront lesdits « systèmes à base de connaissances » ou « systèmes experts ». Il mit ensuite l'accent sur la simulation informatique des phénomènes d'apprentissage grâce à laquelle une machine serait en mesure de construire par elle-même des connaissances à partir de ses propres expériences. Il mentionna, enfin, différentes métaphores qui, selon lui, devaient aider à réaliser une machine intelligente. C'est ainsi qu'il suggéra de prendre pour modèle soit les capacités cognitives humaines – c'est-à-dire notre psychisme – soit le cerveau qui est la source de bien des comportements intelligents – ce qui débouchait sur l'emploi de réseaux d'automates et de réseaux de neurones

formels – soit encore l'évolution des espèces – ce qui anticipait les notions d'algorithme génétique* et d'informatique évolutionniste* – soit enfin les phénomènes d'intelligence collective*, qu'il s'agisse de l'intelligence en essaim, c'est-à-dire de l'intelligence d'une ruche ou d'une termitière, ou des idées partagées par l'ensemble des membres d'une société. Toutes ces métaphores alimentèrent l'imagination de nombreux chercheurs en intelligence artificielle et en sciences cognitives pendant les soixante années qui suivirent, et elles continuent de susciter les travaux de spécialistes. Ainsi, si Alan Turing n'est pas à proprement parler l'inventeur de l'intelligence artificielle, il en est certainement le précurseur le plus influent.

En conclusion de cette évocation du rôle d'Alan Turing, rappelons que la vie tragique de ce personnage hors du commun, mort à 42 ans, semble faire écho au jeu de l'imitation qu'il a mis en scène : homosexuel anglais traqué par la société victorienne de l'immédiat après-guerre, Alan Turing fut arrêté par la police, accusé, jugé, puis condamné à subir un traitement hormonal qu'il ne supporta pas, ce qui le conduisit à se suicider.

« Désormais des machines passent le test de Turing. »

Il s'appelle Eugene Goostman. Il vient d'Odessa (Ukraine), il a 13 ans, de grandes lunettes rondes et un petit sourire mutin. Ce jeune garçon vient de réussir le légendaire test de Turing, selon l'université de Reading. Vous ne pourrez néanmoins pas le féliciter, car ce n'est pas un vrai garçon, mais un programme informatique.

Martin Untersinger, « Réussite contestée d'un ordinateur au légendaire test de Turing », *Le Monde*, 9 juin 2014

Pour marquer le 60e anniversaire de la mort l'Alan Turing, deux enseignants britanniques, Huma Shah, maître de conférence à l'université de Coventry, et Kevin Warwick, professeur de cybernétique à l'université de Reading, organisèrent une compétition solennelle parrainée par la très prestigieuse Royal Society de Londres. Ils y convoquèrent les médias pour constater le pas que l'on allait assurément franchir, un grand pas dans l'histoire de l'humanité, selon eux : en effet, à l'issue de cet événement, le 7 juin 2014, la presse annonça que le vainqueur, un dénommé Eugene Goostman, avait « passé » le test de Turing, car il avait trompé 10 interrogateurs sur les 30 qui avaient été invités à dialoguer avec lui sur une durée de cinq minutes, ce qui dépassait très légèrement, avec quelques années de retard, la prédiction d'Alan Turing selon laquelle d'ici l'an 2000, un interrogateur humain aurait plus de 30 % de chance de se laisser berner par une machine dans une conversation de cinq minutes. Précisons ici qu'Eugene Goostman avait

l'image d'un jeune garçon ukrainien âgé de 13 ans, mais que ses répliques étaient générées par un « robot bavard » (*chatbot* en anglais ou « agent conversationnel » en français académique) d'origine russe, conçu en 2001, soit 13 ans plus tôt. On expliquait qu'au fil des années, il avait amélioré ses performances par apprentissage machine jusqu'à devenir le meilleur… Le « cap » du test de Turing aurait donc été franchi par un ordinateur en 2014 ! Or, une telle affirmation demeure toujours controversée. Pour essayer de comprendre ce débat, sans nous attacher au détail de l'expérimentation, revenons sur l'origine et la signification de ce test. Rappelons d'abord que Turing l'a décrit à deux reprises sous deux formes différentes, dans les deux articles qu'il a consacrés à l'intelligence des machines, l'un publié en 1948 et intitulé « Intelligent Machinery », l'autre paru en 1950 dans la revue *Mind* sous le titre « Computing Machinery and Intelligence ».

Dans le premier, une personne joue aux échecs en ligne, sur un terminal, contre un adversaire inconnu qu'il croit être un homme, alors qu'on lui a substitué à son insu un programme informatique. Dans le second, le décor change, du jeu d'échec on passe au *jeu de l'imitation* qui se joue à trois : un interrogateur humain, C, et deux personnages, B, une femme, et A, un homme qui imite une femme. On substitue, toujours à l'encontre de l'interrogateur C, à l'homme qui imite la femme, à savoir à A, un ordinateur programmé pour converser en imitant un homme qui imite une femme. Dans tous les cas, la machine est dite intelligence si elle fait illusion en simulant des facultés cognitives humaines, l'aptitude à jouer aux échecs dans le premier cas, l'aptitude à jouer au jeu de l'imitation dans le second. Dans

l'article de 1950 où il décrit le jeu de l'imitation, Turing se livre en plus à une prophétie en prédisant que d'ici 50 ans, c'est-à-dire en l'an 2000, les progrès techniques réalisés permettront de fabriquer une machine qui trompera au moins 30 % des interrogateurs sur une durée de 5 minutes.

Littéralement parlant, on doit donc se rendre à l'évidence : en dépit des critiques portant sur tel point de détail du protocole expérimental, la performance d'Eugene Goostman, en accomplissant la prophétie d'Alan Turing, laisserait entendre que désormais toutes les autres prophéties concernant les ordinateurs et l'intelligence artificielle devraient être prises au sérieux, car elles seraient, elles aussi, susceptibles de se réaliser un jour.

On est pourtant en droit de se demander si l'épreuve en question, somme toute assez dérisoire, constitue effectivement une prouesse significative au terme de laquelle on devrait qualifier les machines d'« intelligentes ». Symétriquement, il apparaît tout aussi utile de relire les articles d'Alan Turing pour comprendre ce qui le motivait lorsqu'il conçut ces épreuves et pour évaluer le prix qu'il attachait lui-même à leur franchissement. Or, une lecture attentive des deux articles de 1948 et de 1950 montre que l'ambition de ce que l'on a appelé depuis le « test de Turing » demeurait modeste. Il ne s'agissait pour lui que d'expliquer en quoi les machines pourraient un jour reproduire certaines activités intellectuelles humaines. La principale difficulté à laquelle il se heurtait à l'époque tenait à leur apparence extérieure qui faisait – et qui fait toujours – qu'en les voyant on ne peut les confondre avec des êtres humains, ce qui les discrédite quelque peu à nos yeux. Une autre difficulté venait de ce que l'intrication entre notre corps sensible et le monde

n'est pas aisée à reproduire. Il fallait donc trouver un subterfuge qui affranchisse la machine de tout lien « corporel » avec le monde. Dès 1948, Alan Turing imagina cinq types d'épreuves qui ne recourent qu'à des organes élémentaires de vision, de parole et d'écoute, sans nécessiter un corps : (1) les jeux, comme le jeu d'échec, de morpion, de bridge ou de poker, (2) l'apprentissage des langues (3) la traduction d'une langue dans une autre, (4) la cryptographie, (5) les mathématiques. On conçoit dès lors qu'il ait recouru à la première épreuve, à savoir un jeu, en l'occurrence le jeu d'échec, en 1948, puis qu'il l'ait raffiné en recourant à un nouveau jeu, le jeu de l'imitation, fondé sur l'apprentissage des langues par une machine en 1950. C'est dans ce contexte que se place son fameux test qu'il décrit merveilleusement, avec un exemple déconcertant de conversation entre un ordinateur et un homme. Sa prophétie n'apparaît alors qu'incidemment, comme pour montrer que le défi lancé est empirique et que sa réalisation n'a rien d'inimaginable.

Contrairement à ce que beaucoup affirmèrent par la suite, cela ne délimite pas ce qu'Alan Turing entendait par intelligence des machines, car il n'avait jamais ignoré les dimensions perceptives de l'intelligence, loin s'en faut, mais il les jugeait trop ardues à reproduire à court terme ; et contrairement aussi à ce que d'autres imaginèrent par la suite, Turing n'a jamais évoqué une indiscernabilité totale entre l'humain et la machine. En effet, la portée de cette épreuve demeure empirique au sens où elle doit pouvoir se soumettre à une expérimentation, c'est d'ailleurs ce que traduit la quantification précise énoncée dans sa prévision, quantification qui autorisa la mise en scène d'Huma Shah et de Kevin Warwick pour commémorer le 60e anniversaire de la mort d'Alan Turing…

Bref, que l'on ait fabriqué des machines qui réalisent la prédiction d'Alan Turing formulée en 1950 ne clôt pas le débat autour de l'intelligence des machines, car tant pour Turing que pour ses successeurs immédiats, comme John McCarthy qui inventa le terme d'intelligence artificielle, doter les machines de pensée constitue l'horizon régulateur d'une science à venir, et non un objectif final à atteindre dans un terme défini.

« L'intelligence artificielle n'est pas une science. »

Je ne nie pas que nous pourrions créer une intelligence artificielle et qu'une forme de base existerait déjà actuellement. Je ne sais pas de quand date vraiment l'apparition de l'intelligence dans l'évolution des espèces et combien de temps il a fallu pour qu'elle atteigne notre stade. Sans doute l'homme de Cromagnon était-il aussi intelligent que nous. Ce serait bien un contexte, des techniques, des progrès qui nous feraient croire que notre intelligence est supérieure. J'en viens donc à croire qu'une intelligence artificielle est possible.

Citation de David Strainchamps, par Jean-Michel Truong sur son site

Si vous rencontriez un banquier, vous viendrait-il à l'idée de lui demander : « La banque, vous y croyez ? » Cela n'aurait guère de sens car la banque apparaît à tous comme une institution tangible que nous respectons tous, à défaut de lui faire confiance. Songeriez-vous, en croisant un mathématicien, à lui demander : « Les mathématiques, vous y croyez ? » Imagineriez-vous interpeller un physicien avec ces mots : « La physique, vous y croyez ? » Ces disciplines instituées depuis longtemps s'imposent à tous, même si l'on ne saisit pas toujours le sens des propositions qu'elles émettent. Certes, on sait que beaucoup de conjectures mathématiques n'ont pas été démontrées, que les mathématiciens ne s'accordent pas tous sur le sens à attribuer à telle proposition ou à telle notion, qu'une grande partie de leur activité consiste à corriger les erreurs de leurs collègues, que les théories

physiques nouvelles conduisent souvent à remettre en cause les théories anciennes etc., mais cela n'entame pas notre confiance dans les mathématiques, ni dans la physique, ni dans les sciences vénérables comme la chimie ou les sciences de la terre. Et l'ancienneté, à elle seule, ne constitue pas un gage de crédibilité. Aucun individu sérieux n'oserait remettre en question les travaux actuels sur le séquençage du génome et harceler les biologistes avec des interrogations du type : « La biologie moléculaire, vous y croyez ? »

Et pourtant, en dépit de l'emprise croissante qu'exercent les machines sur notre vie quotidienne, il m'est souvent arrivé, à l'énoncé de ma profession, professeur d'informatique, et de mon intérêt pour l'intelligence artificielle, de voir mes interlocuteurs s'exclamer : « L'intelligence artificielle, vous y croyez ? » Je constate ainsi que mon domaine de recherche se trouve ravalé au rang des pratiques douteuses comme l'iridologie ou la sophrologie, voire même à celui de sciences occultes comme l'astrologie ! Outre ce que cela peut avoir de blessant pour un esprit épris de positivité, cette question repose sur une perception erronée : l'intelligence artificielle n'est aucunement affaire de croyance, encore moins de foi. Ce champ de l'informatique recourt aux apports des mathématiques les plus ardues, en particulier de la logique formelle, des statistiques théoriques et de l'algèbre, pour modéliser et simuler des facultés intellectuelles comme le raisonnement, la compréhension du langage naturel, la perception, etc. De plus, et c'est là ce qui fait de l'intelligence artificielle une science, les modèles conçus par les chercheurs font l'objet de deux types de validation, une validation formelle et une validation empirique.

La validation formelle porte sur la cohérence des modèles et sur leur aptitude à faire l'objet d'une simulation informatique. En effet, en intelligence artificielle, les modèles prennent la forme d'algorithmes, c'est-à-dire de séquences d'opérations logiques parfaitement définies et non ambiguës. Pour qu'un ordinateur exécute ces algorithmes, il faut que ceux-ci se terminent en un temps fini, pas trop long. De plus, il faut que le résultat de l'exécution de ces algorithmes se conforme à ce que l'on espérait. La validation formelle des algorithmes vérifie que les opérations logiques évoquées dans les modèles sont bien définies, que l'enchaînement de ces opérations se termine au bout d'un temps fini, que le nombre d'opérations, ce que l'on appelle, en termes techniques, la complexité algorithmique, est « raisonnable », quelle que soit la taille des données et enfin que le résultat obtenu correspond à ce que l'on attend. Ce travail de validation fait appel aux ressources de l'informatique théorique, de la logique et de bien d'autres domaines des mathématiques.

Une fois que l'on a vérifié la cohérence des modèles et leur aptitude à faire l'objet de simulations informatiques, on exécute ces simulations sur des données réelles, puis on confronte les résultats obtenus avec des données d'expériences. Plus précisément, l'intelligence artificielle porte sur la reproduction, au moyen d'ordinateurs, de nos capacités mentales, par exemple de notre faculté à raisonner, à comprendre des textes, à démontrer des théorèmes, à percevoir des formes, etc. La validation empirique porte donc sur la comparaison de nos facultés intellectuelles et des résultats construits par des machines. Par exemple, on circonscrit l'ensemble des théorèmes démontrés par une machine, on

fait jouer un homme contre un ordinateur au jeu d'échecs, on pose des questions à un ordinateur qui doit les traduire, sans ambiguïté, dans un langage logique, etc. Mais là encore, deux types de validations empiriques existent, certaines portent sur les résultats, d'autres sur les processus.

À titre d'illustration, prenons un exemple : la démonstration automatique de théorèmes. D'un côté, on s'intéresse à l'ensemble des théorèmes qu'une machine est capable de prouver. Cela donne une idée de la puissance des algorithmes mis en œuvre par l'intelligence artificielle. D'un autre côté, l'attention se porte sur la comparaison entre la façon dont la machine prouve des théorèmes et la méthode employée par des hommes. On examine les différents pas de la démonstration d'un théorème par la machine et on essaie de les mettre en parallèle avec les étapes de la démonstration du même théorème par les hommes, y compris leurs hésitations, leurs errements, leurs retours en arrière, etc., toutes choses que recensent des expériences de psychologie cognitive.

Pour préciser les idées, prenons un second exemple : le jeu d'échecs. Depuis qu'une machine a battu le champion du monde en titre, nous sommes tous convaincus de la puissance des ordinateurs. Cependant, la validation des programmes qui affrontent les hommes au jeu d'échecs ne passe pas uniquement par cette épreuve. On s'attache aussi aux analogies entre ce que font les hommes et ce que fait une machine : comment les hommes choisissent-ils leurs coups ? Comment mémorisent-ils l'échiquier ? Comment évaluent-ils une position ? Est-ce qu'une machine procède de la même façon ? Si oui, on dispose, là encore, d'un modèle de l'activité humaine.

En résumé, nous voyons que, loin d'être uniquement un objet de croyance, l'intelligence artificielle se constitue en une discipline scientifique dont les méthodes sont explicitées et discutées dans la communauté des chercheurs, et dont les résultats sont validés par des expériences rigoureuses.

D'où vient donc l'idée selon laquelle l'intelligence artificielle serait un article de foi ? Certainement du fait que l'on se trompe sur ses objectifs. En effet, beaucoup croient que le projet de l'intelligence artificielle va bien au-delà de la simple simulation de comportements intelligents et qu'il porte sur la réification d'une conscience.

Cela tient à l'ambiguïté du terme d'intelligence artificielle qui, pris littéralement, désigne soit la discipline scientifique qui étudie et simule les différentes facettes de l'intelligence avec les ressources des techniques du traitement de l'information, soit la reconstitution artificielle d'une intelligence. La première éventualité correspond à ce dont nous parlons ici. Elle s'est constituée en objet de science depuis plus de soixante ans ; elle existe indubitablement et ne saurait être remise en cause. Quant à la seconde, elle répond à des mythes ancrés dans l'imagination humaine depuis l'aube des temps, par exemple à celui du Golem de Prague ou de Pygmalion. On l'évoque avec un article indéfini, pour parler d'« une intelligence artificielle », puisqu'il pourrait y en avoir plusieurs, car il y a plusieurs mythes. Ce sont ces mythes qui font éventuellement l'objet de croyance, mais certainement pas l'intelligence artificielle, qui se présente aujourd'hui comme une réalité tangible pour tous et presque aussi, voire peut-être plus, rémunératrice que l'institution bancaire…

« L'intelligence artificielle est une idée neuve. »

Eh bien ! Puisque cette femme vous est si chère…
JE VAIS LUI RAVIR SA PROPRE PRÉSENCE.
Je vais vous démontrer, mathématiquement et à l'instant même, comment, avec les formidables ressources actuelles de la Science – et ceci d'une manière glaçante peut-être, mais indubitable – comment je puis, dis-je, me saisir de la grâce même de son geste, des plénitudes de son corps, de la senteur de sa chair, du timbre de sa voix, du ployé de sa taille, de la lumière de ses yeux, du reconnu de ses mouvements et de sa démarche, de la personnalité de son regard, de ses traits, de son ombre sur le sol, de son apparaître, du reflet de son Identité, enfin.

Villiers de l'Isle-Adam, *L'Ève future*, 1886

L'intelligence artificielle a déjà plus d'un demi-siècle. Ce n'est pas une science tout à fait neuve, même si, comparée à la physique, à la géologie, aux mathématiques ou à la chimie, elle fait figure de jeunette. Quant à l'idée d'intelligence artificielle, c'est-à-dire de reproduction, au moyen de machines, des différentes fonctions de notre entendement, comme la capacité de parler, de lire, de comprendre, de calculer, de raisonner, etc., les prémices en vinrent très tôt, bien avant qu'il y eut des ordinateurs et que l'on sache les programmer. L'idée germa dans la tête des poètes et des philosophes, dès l'Antiquité.

Ainsi, Homère décrivit, dans le chant XVIII de l'*Iliade*, des servantes en or que « la raison habite ». Fabriquées par Héphaïstos, le dieu forgeron, elles ont, selon le poète, voix et force ; elles vaquent aux occupations quotidiennes à la

perfection, car les immortels leur ont appris à travailler. Ce sont donc des robots, au sens étymologique de travailleurs artificiels. D'autres mentionnent l'existence de têtes parlantes en Grèce ancienne, par exemple, d'un masque d'Orphée qui rendait des oracles à Lesbos.

De même, en Égypte, il existait des statues articulées, animées par la vapeur et par le feu, qui hochaient la tête et bougeaient les bras ; les prêtres interprétaient ces mouvements comme autant de prophéties. Et Héron d'Alexandrie est réputé avoir réalisé des oiseaux mécaniques qui pépiaient, volaient, buvaient…

Plus proche de nous, Descartes proposa, avec ses « animaux-machines », de reproduire artificiellement les fonctions biologiques essentielles à la vie animale comme l'alimentation, la communication parlée, la locomotion, etc. Et Leibniz, pour qui tout dans la nature se produit par calcul, voulut aller plus loin et simuler, sur des machines à calculer, le raisonnement. Un philosophe français, Julien de La Mettrie, imagina même que l'homme tout entier, âme et corps, se réduirait à une machine que l'on créerait un jour, grâce aux progrès de la technique.

En parallèle, les fabricants d'automates s'essayèrent à la réalisation matérielle de physiologies artificielles et à la simulation effective des fonctions biologiques sur ces machines. Citons, à titre d'illustration, les têtes parlantes fabriquées, presque simultanément, à la fin du XVIII[e] siècle, par trois inventeurs de génie, l'abbé Mical, en 1778, le baron Wolfgang von Kempelen, toujours en 1778, et Kratzenstein en 1780.

Cependant, si l'on parvenait bien là à reproduire ce que Descartes appelait les « esprits animaux », c'est-à-dire ce qui anime et assure la subsistance animale, nous nous trouvions

dans l'incapacité d'approcher la pensée et les phénomènes mentaux. Et jusqu'à ce que la technologie des relais téléphoniques ne se perfectionne, dans les années 1930, jusqu'à ce que l'on s'essaie à mettre bout à bout les éléments dont ces relais étaient faits, pour exécuter les fastidieux calculs exigés par la balistique, science du jet des projectiles, et par le décryptage des messages interceptés chez l'ennemi, rien de cet ordre n'avait vraiment abouti.

Au cours de la Seconde Guerre mondiale, sous le grondement sourd des bombes, quelques physiologistes et quelques mathématiciens crurent voir dans les circuits réticulaires aux mailles de métal que formait l'ordonnancement savant des composants électroniques des calculateurs, une image grossière de la substance grise de notre encéphale. Ce fut par exemple le cas lorsque, en 1943, Warren McCulloch, logicien et neuropsychiatre, et Walter Pitts, mathématicien âgé à l'époque d'à peine 20 ans, établirent un parallèle entre l'activité nerveuse et un assemblage d'automates électriques mis bout à bout. Des cellules nerveuses, les neurones, aux composants électroniques, et de leurs connexions, les synapses, aux conducteurs électriques qui relient ces composants, il y avait une analogie structurelle qu'une transposition audacieuse établit rapidement. Toutes ces tentatives donnèrent naissance au mouvement dit cybernétique*, qui se développa à partir de 1946, sous l'impulsion de Norbert Wiener, et se poursuivit pendant une dizaine d'années.

Plus exactement, McCulloch et Pitts montrèrent que toute fonction logique qui prend ses valeurs dans un ensemble à deux éléments comme vrai et faux ou noir et blanc, quelle que soit sa complexité, peut être réalisée par un réseau d'automates semblable à ceux que nous venons de décrire. Ils sont même

La machine à parler de Kempelen

En 1779, l'Académie des sciences de Saint-Pétersbourg offrit un prix pour qui réaliserait une mécanique capable de prononcer cinq voyelles. Kratzenstein gagna le prix, car il montra que toutes les voyelles pouvaient être prononcées en insufflant de l'air dans des tuyaux. Toutefois, ses travaux demeuraient essentiellement théoriques.

En revanche, la machine de Kempelen était opérationnelle. Elle se fondait sur un modèle des organes vocaux humains, et s'inspirait de la cornemuse avec un soufflet.

– les doigts de la main droite produisaient les consonnes par des battements semblables à ceux des lèvres ;

– les doigts de la main gauche donnaient les voyelles en bouchant plus ou moins un orifice.

Le répertoire de cette machine à parler contenait les mots : « opéra, astronomie, Constantinople » et quelques messages comme « vous êtes mon ami, je vous aime de tout mon cœur, venez avec moi à Paris », etc.

Goethe eut le loisir de l'entendre et raconte qu'elle était capable de prononcer très gentiment quelques mots enfantins. Aujourd'hui, cet instrument est la propriété du Deutsches Museum de Munich.

allés plus loin et ont prouvé que des réseaux très simples, qui contiennent deux couches d'automates en plus de la couche d'entrée, peuvent réaliser toutes les fonctions booléennes*, à condition que le nombre d'automates sur la couche dite intermédiaire soit suffisant. Dès lors, puisqu'un réseau d'automates peut s'interpréter simultanément au plan logique, neurobiologique et technologique, le lien entre les machines, ce qui relève du cerveau et ce qui porte sur les lois de la pensée et du jugement, autrement dit du vrai et du faux, semblait à portée de main.

Cependant, il apparut rapidement que la réalisation de machines douées effectivement de capacités cognitives posait des problèmes insurmontables, car il fallait être en mesure d'agencer correctement les automates matériels dont elles étaient constituées pour réaliser des fonctions mathématiques complexes imitant les performances des sujets humains. Or, la disposition manuelle des automates s'avérait trop complexe pour être menée à bien. Il fallait donc que les machines les agencent d'elles-mêmes. Et cette auto-organisation spontanée des réseaux d'automates posait des problèmes mathématiques difficiles à résoudre. Les échecs rencontrés conduisirent certains chercheurs à explorer de nouvelles pistes de recherche.

C'est ce qui motiva, en 1956, les promoteurs de la fameuse école d'été du Dartmouth College. Celle-ci réunit, entre autres, des logiciens, des ingénieurs, des psychologues, des économistes, des cybernéticiens, qui envisagèrent de créer une nouvelle science destinée à modéliser sur ordinateur la prise de décisions et qui donnèrent à cette nouvelle discipline le titre provocateur d'intelligence artificielle. La rupture avec l'esprit de la cybernétique était patente : d'une simulation de mécanismes biologiques d'auto-organisation, on passait à une simulation de phénomènes d'ordre psychologique. De la physiologie, c'est-à-dire de l'étude des cellules et de leurs constituants, on allait au raisonnement ; du genre animal, de l'« hominidé », on en venait au genre humain, à l'« homo sapiens ».

Autrement dit, selon ce que l'on entend par intelligence artificielle, l'origine en est plus ou moins lointaine. Si l'on considère que l'intelligence artificielle vise à reproduire, au moyen de machines, des êtres animés plus ou moins semblables à nous, l'idée en remonte à la plus haute Antiquité. Si

l'intelligence artificielle est vue comme la simulation, sur des automates à états discrets, autrement dit, sur des ordinateurs, des supports physiologiques sur lesquels s'ancrent nos facultés cognitives, par exemple du cerveau, cela commence au milieu de la Seconde Guerre mondiale, en 1943. Et si l'on envisage l'intelligence artificielle en son sens restreint de simulation informatique de nos capacités cognitives, la discipline n'a pas beaucoup plus d'un demi-siècle. Bref, la nouveauté – ou plus exactement l'ancienneté – de l'intelligence artificielle est toute relative.

« Les Japonais sont les champions de l'intelligence artificielle. »

Après le robot Asimo, qui ne sait pas faire grand-chose d'autre que marcher et se déhancher, voilà le HRP-2W. Mais lui sait se rendre utile : il est capable d'attraper un thermos de thé, de verser le breuvage dans une tasse et de l'amener à son maître. Lors de la démonstration réalisée mercredi par une unité de recherche de l'université de Tokyo, il est même allé rincer le mug dans l'évier, sans rien casser. Et surtout sans se plaindre. Le plus intéressant dans cette démonstration, c'est qu'il n'y avait aucun scénario de mouvements prédéfinis, a assuré le professeur Tomomasa Sato. [...] Au Japon, de nombreuses équipes travaillent sur le développement de robots d'assistance domestique. Avec le vieillissement de la population, le marché pourrait exploser dans les années à venir.

20 Minutes.fr, 1er mars 2007

Un projet d'ordinateur dit de cinquième génération a été lancé à grand fracas au Japon par le MITI – ministère de l'Industrie et des Échanges internationaux (Ministry of International Trade and Industry) – au début des années 1980. Munies de moteurs d'inférences – autrement dit de techniques d'intelligence artificielle – et fondées sur le langage Prolog, ces machines étaient censées supplanter les ordinateurs traditionnels. L'industrie électronique japonaise ayant acquis, depuis l'avènement du transistor, une position clef dans le secteur des technologies du traitement de l'information, le monde entier regarda avec attention et respect ce projet. Et puisque les Japonais misaient autant

sur le développement de l'intelligence artificielle, beaucoup crurent qu'ils en étaient les « champions ». Qui plus est, les succès de l'industrie japonaise dans le secteur des consoles de jeux vidéo ainsi que les progrès de la robotique dite de compagnie, par exemple les petits chiens robots fabriqués par la firme Sony, accroissent d'autant ce sentiment de l'excellence japonaise en matière d'intelligence artificielle. Or, à y regarder de près, aucun des arguments allégués ne justifie une telle réputation.

En tout premier lieu, la notion d'ordinateur de cinquième génération ne s'est jamais vraiment imposée et laissa sceptique bien des spécialistes dès que le MITI en fit part au monde. Pour comprendre les doutes que suscitèrent les prétentions des Japonais, rappelons à quoi correspondent les différentes générations d'ordinateurs. On considère généralement que les ordinateurs de première génération apparaissent avec l'emploi de dispositifs électriques de traitement de l'information. Il faut savoir qu'auparavant, par exemple au XVII^e^ siècle avec ladite « pascaline », c'est-à-dire avec la machine à calculer réalisée par Blaise Pascal, ou plus tard, au XIX^e^, avec les premières tentatives de Charles Babbage pour réaliser un ordinateur, c'est-à-dire une machine qui exécute des séquences d'opérations logiques, autrement dit des algorithmes, et pas uniquement des calculs numériques, ou encore, toujours au XIX^e^ siècle, avec la machine à raisonner fabriquée par Stanley Jevons, on recourait à l'emploi de pièces mécaniques pour fabriquer des automates matériels réalisant des calculs. Ce n'est qu'à la fin des années 1930 que l'on fit appel à des dispositifs électriques, comme des relais téléphoniques et des tubes à vides, pour construire les premiers calculateurs électro-

mécaniques puis, à partir de 1946, les premiers ordinateurs dignes de ce nom.

La seconde génération d'ordinateurs apparaît avec l'invention du transistor en 1947 puis avec le remplacement progressif des lampes, fragiles et encombrantes, par ces nouveaux composants, plus compacts, plus fiables et plus solides. On réduisit ainsi le poids et le coût des ordinateurs. À cela s'ajoutent la notion de microprogrammation qui fit son apparition en 1956, l'introduction de disques magnétiques pour stocker les informations et l'invention du langage Fortran – toujours en 1956 – ce qui facilita grandement la programmation des ordinateurs. On fait donc remonter le début de la seconde génération d'ordinateurs en 1956.

En 1958, un ingénieur de la société Texas Instrument, Jack Kilby, conçut le premier circuit intégré qui juxtaposait, sur un seul composant physique, de nombreux composants électroniques. Cela réduisit considérablement les coûts de fabrication des dispositifs électroniques et leur encombrement, puisqu'il n'y avait plus à assembler manuellement des composants. À partir de 1963, on fabriqua des ordinateurs dits de troisième génération, avec des circuits intégrés. Avec le temps, les techniques de fabrication ne cessèrent de se complexifier, mais le principe de la miniaturisation était acquis ; et c'est sur lui que reposèrent les progrès ultérieurs des ordinateurs. On assiste depuis à une baisse continue des coûts et à un accroissement tout aussi continu des performances selon des lois exponentielles qualifiées de « lois de Moore », du nom d'un ingénieur de la société Intel qui les a formulées en 1964.

On date des années 1970 et de l'invention des microprocesseurs, l'apparition des ordinateurs de quatrième

génération, dont on ne sait pas très bien s'ils se caractérisent par une innovation matérielle ou par des transformations du logiciel liées aux langages de programmation dits évolués comme Algol et Pascal, ou structurés, comme les langages à objets, ou encore aux systèmes de gestion de bases de données. Quoi qu'il en soit, la quatrième génération d'ordinateurs ne répond pas à une révolution technologique clairement identifiable, comme les trois premières, mais plutôt à une évolution difficile à cerner.

Dans ce contexte, le lancement de la cinquième génération d'ordinateurs par le MITI japonais, au tout début des années 1980, alors que la quatrième génération d'ordinateurs ne se distinguait pas vraiment de la troisième, suscita la perplexité chez les spécialistes. Et trente-cinq ans plus tard, cette perplexité n'a fait que s'accroître au regard de la faiblesse des résultats atteints. Néanmoins, aux yeux du grand public, les images s'imposent : les Japonais apparurent et continuent d'apparaître comme les avocats de l'intelligence artificielle et, en conséquence, comme les plus compétents dans ce secteur de l'informatique. Pourtant, à bien examiner les choses, rien ne justifie une telle assertion.

Rappelons-le, l'intelligence artificielle est née aux États-Unis en 1956. Les chercheurs qui contribuèrent à son développement furent essentiellement d'origine nord-américaine et européenne, qu'ils soient américain de souche comme Alan Newell, anglais comme Alan Turing, ou fraîchement immigrés d'Allemagne comme Herbert Simon, d'Irlande et de Lituanie comme John McCarthy, voire de Russie comme Marvin Minsky. Rien donc, au plan historique, ne légitime la prétendue prédominance des chercheurs japonais dans ce secteur. Et la participation

du Japon dans les conférences scientifiques internationales ne l'explique pas non plus.

Enfin, la présence au Japon d'industries de hautes technologies comme les consoles de jeux vidéo ou la robotique de compagnie ne change rien à notre constat. En effet, contrairement à ce que l'on imagine et à ce que les publicités laissent entendre, les progrès dans ces secteurs tiennent assez peu à l'emploi de l'intelligence artificielle, mais surtout à l'image de synthèse et à la réalité virtuelle.

Bref, la supposée prépondérance du Japon en matière d'intelligence artificielle ne repose sur aucun argument tangible, qu'il soit d'ordre historique, culturel, économique ou intellectuel. À cet égard, notons que la firme japonaise Sony possède depuis vingt ans un centre de recherche en intelligence artificielle à Paris, dans le V^e^ arrondissement, qui emploie beaucoup de chercheurs français et européens. De même, une société française de construction de robots androïdes, Aldebaran Robotics, qui fabriqua entre autres le robot NAO, racheta une autre société française, Gostai, spécialisée dans la conception de robots de télé-présence, avant d'être elle-même rachetée en 2012 par la société japonaise Softbank… Nous n'avons donc rien à envier aux Japonais dans ce secteur. Sans compter que ledit projet de cinquième génération mentionné plus haut recourait à l'emploi du langage de programmation logique Prolog, inventé en France au tout début des années 1970 par Alain Colmerauer. Ce chercheur, resté ignoré de ses compatriotes pendant de nombreuses années, dut sa reconnaissance au projet japonais. Et là comme ailleurs, la vieille maxime d'origine biblique se vérifie : « Nul n'est prophète dans son pays et dans sa maison » Évangile selon Saint-Marc, VI, 4 ; Saint-Matthieu XIII, 57.

« La recherche en intelligence artificielle est menée par les GAFA. »

Les GAFA en pleine course à l'intelligence artificielle. Apple, Facebook et Google rachètent des start-up à tour de bras pour garder leur position de précurseurs dans le domaine de l'IA. Dernière acquisition en date, celle d'Emotient par Apple.

Aude Fredouelle, Journal du Net, 8 janvier 2016

Le sigle GAFA signifie « Google, Amazon, Facebook, Apple ». Dans l'idée de ceux qui l'emploient, il désigne, de façon générique, les grands acteurs de l'Internet dont le pouvoir s'étend au rythme de la numérisation de la société, si bien que l'on parle parfois de GAFAM en ajoutant un M qui renvoie à la société Microsoft. Plus récent, et construit sur un principe analogue, le terme NATU évoque quatre grandes entreprises emblématiques de la « disruption » numérique, Netflix, Airbnb, Tesla et Uber. Et, pour compléter le tableau, on pourrait aussi y adjoindre les sociétés Twitter ou Paypal, et même, pourquoi pas, Yahoo et IBM, et bien d'autres encore qui jouent un rôle important dans les industries du numérique et dont on craint qu'elles ne se joignent elles-aussi à une sorte de complot planétaire des industriels de l'hyper-modernité. Il faut toutefois savoir que le terme GAFA est surtout utilisé en France et qu'il demanderait à être discuté plus avant, car un examen attentif des relations entre ces grandes compagnies montre qu'elles rivalisent plus souvent les unes contre les autres

qu'elles ne conspirent toutes ensembles contre le reste du monde.

Cependant, indépendamment de leur volonté de puissance, réelle ou supposée, et de leurs rivalités, l'intelligence artificielle joue un rôle important dans leurs stratégies. De multiples déclarations le confirment. À cela on doit ajouter des investissements conséquents qui se traduisent par l'ouverture de centres de recherche et par le recrutement d'un grand nombre d'ingénieurs spécialisés en intelligence artificielle. Parallèlement à cet engouement pour l'intelligence artificielle, on note, chose étrange, une aspiration à rendre publiques les technologies mettant en œuvre les principes d'intelligence artificielle. C'est ainsi que Google a annoncé en novembre 2015 qu'elle mettait en accès libre le système *TensorFlow* qui comprend des technologies d'apprentissage machine recourant à de l'apprentissage profond *(Deep Learning)* utilisées pour l'analyse d'images, pour la reconnaissance de la parole ou pour la conception de logiciels dits de questions-réponses parce qu'ils répondent automatiquement à des questions posées en langage naturel. Un mois plus tard, Facebook dévoilait son serveur dédié à l'intelligence artificielle, *Big Sur*, et le mettait aussi en libre accès. Quant à Elon Musk, le créateur des firmes Paypal, Tesla et SpaceX, il fonda en décembre 2015 la société OpenAI destinée à offrir à chacun l'usufruit de tous les logiciels d'intelligence artificielle.

Nous pouvons donc faire deux constats qui paraissent à certains égards contradictoires : d'un côté, l'intelligence artificielle apparaît cruciale pour tous ces acteurs et, d'un autre côté, ces entreprises se proposent de partager gratuitement les techniques d'intelligence artificielle qu'elles

développent. Sans doute, ces deux attitudes semblent de prime abord antagoniques car, si l'intelligence artificielle est stratégique, cela signifie qu'elle donne un avantage compétitif sur les autres, et donc qu'il serait préférable de ne pas dévoiler son savoir-faire en la matière.

Pour bien comprendre ce qui est en jeu ici et montrer en quoi il n'y a rien là d'antinomique, en dépit des apparences, il faut d'abord expliquer pourquoi l'intelligence artificielle devient un enjeu majeur, puis montrer les facteurs clefs du développement des techniques d'intelligence artificielle.

L'importance de l'intelligence artificielle tient à l'économie spécifique de la toile qui s'est mise en place à partir de 2004, avec l'avènement du web 2.0. Pour saisir cette spécificité et l'apport de l'intelligence artificielle, rappelons qu'au début du développement de la toile, dans le courant des années 1990, beaucoup pensèrent que l'économie du web se développerait sur le modèle classique des révolutions industrielles du passé, à savoir que les premiers arrivants sur les nouveaux marchés acquerraient rapidement des situations de domination quasi-monopolistique. Cela explique la bulle spéculative de la fin des années 1990, lorsque les investisseurs misaient sur tous les marchés en germes susceptibles de se développer, sans examiner le détail des technologies mises en œuvre ; cela explique aussi l'éclatement de cette bulle, au tournant des années 2000, car elle tenait à une piètre évaluation du potentiel réel des nouvelles entreprises et surtout à une mauvaise analyse de la structure du tissu économique de l'Internet. Quelques années plus tard, en 2004, un certain nombre d'industriels qui, en dépit de la crise, croyaient toujours dans la viabilité économique du web, essayèrent de comprendre les clefs du succès de grandes compagnies

de l'Internet comme Amazon et Google. Ils constatèrent que, dans le monde numérique, la réputation se défaisait aussi vite qu'elle se faisait et que, pour se maintenir, une entreprise devait exploiter toutes les sources d'information qui permettent d'évaluer et de comprendre l'opinion des consommateurs afin de remédier au plus tôt à d'éventuelles insatisfactions, d'améliorer ses produits en fonction de la demande et de répondre à des rumeurs malveillantes.

Cela conduisit à la mise en place du web 2.0, c'est-à-dire du web participatif où l'on implique l'utilisateur dans la conception des produits et, surtout, où l'on récupère ses appréciations et ses réactions. Toutes ces informations constituent d'immenses bases de données qui traduisent les goûts, les opinions et les critiques des utilisateurs. Il convient de les exploiter pour améliorer la qualité des produits, les configurer au regard des besoins de chacun et définir la stratégie des entreprises. Dans ce but, l'intelligence artificielle et, plus particulièrement, l'apprentissage machine jouent un rôle essentiel ; en effet, cela permet d'extraire des connaissances à partir de très grandes masses de données, ce que l'on appelle les *Big Data*. Ainsi, c'est pour interpréter ces immenses quantités d'informations recueillies sur le web que les grands acteurs de l'Internet manifestent le besoin d'intelligence artificielle et d'apprentissage machine.

Ces techniques sont connues depuis longtemps ; elles ne constituent pas, en tant que telles, des nouveautés. Ce qui importe, ce sont d'une part les « tours de main » et les astuces nécessaires à leur mise en œuvre dans différents contextes, et d'autre part les masses de données sur lesquelles les algorithmes d'apprentissage fonctionnent. Or, en rendant publiques les programmes informatiques

qui implémentent les algorithmes d'apprentissage machine, les grandes sociétés ne dévoilent pas tout leur savoir-faire ; qui plus est, en contrepartie, elles espèrent s'attirer la sympathie du public, améliorer leur réputation et séduire des développeurs d'applications qui renverront, à leur tour, des retours d'usage aidant à améliorer ces algorithmes d'apprentissage. Ils seront aussi susceptibles d'indiquer la façon dont, eux-mêmes, tirent parti de ces techniques, ce qui peut toujours servir. Enfin, en invitant les utilisateurs à lancer les procédures d'apprentissage sur leurs propres machines, ils pourront aussi récupérer des données susceptibles d'être utiles. En effet, il faut comprendre qu'aujourd'hui la source de la richesse dans le monde numérique ne tient ni à ces algorithmes, dont les principes sont anciens, ni à leur programmation informatique, mais aux masses de données sur lesquelles ils sont mis en œuvre ; ce sont ces données que Google accroît chaque fois que vous soumettez une nouvelle requête à son moteur de recherche ou que Facebook engrange à mesure que les réseaux d'amis s'accroissent, ou encore qu'Apple accumule lorsque vous utilisez ses produits. Or, ces masses de données demeurent privées… Et, en dépit du grand mouvement contemporain d'Open Science qui vise à partager les résultats de la recherche publique, elles demeurent la propriété exclusive des grandes sociétés de l'Internet qui n'ont jamais évoqué leur mise à disposition du grand public…

« L'intelligence artificielle pallie les défaillances de notre intelligence. »

Des machines remplacent nos jambes (bateau, bicyclette, automobile, avion), des prothèses assistent nos sens (lunettes, appareils acoustiques, téléphones, télévision). L'élevage et l'agriculture pratiquent la manipulation génétique, depuis le néolithique, par la sélection des espèces. La bionique, l'intelligence artificielle ne font que s'ajouter aujourd'hui au catalogue des prothèses qui assistent nos activités physiques ou mentales.

« L'Ordinateur et l'intelligence », site de Michel Volle

Beaucoup d'entre nous aimeraient qu'un petit ange gardien prenne soin de notre personne et nous assiste dans les activités quotidiennes. Il viendrait corriger nos erreurs d'inattention, nous souffler la réponse juste au bon moment, rappeler les noms de nos interlocuteurs quand notre mémoire flanche, prévenir nos bévues avant qu'elles ne nous échappent, etc. Avec l'intelligence artificielle, on fabrique, dès à présent, des agents dits intelligents, qui nous aident à gérer nos agendas, à organiser notre travail, à prendre nos billets d'avion, à filtrer et classer nos courriers électroniques, à faire nos courses sur Internet, etc. On appelle ces petits automates bienveillants et débrouillards des elfes ; ces espèces de lutins modernes suppléent à nos déficiences comme autant de domestiques fidèles et zélés. D'autres techniques du traitement de l'information, comme les correcteurs d'orthographe, les logiciels de reconnaissance de la parole et de l'écriture, les calculatrices, les dispositifs d'assistance au freinage (ABS), les régulateurs

de vitesse ou encore les assistances au stationnement, pallient depuis plus ou moins longtemps nos déficiences cognitives. L'intelligence artificielle appartient donc au magasin des prothèses intellectuelles dont la réputation est déjà bien établie. Elle y côtoie d'autres articles en vente depuis bien longtemps comme le boulier, la règle à calculer, les abaques, les bâtons de Napier, les astrolabes, les sphères armillaires, etc.

Néanmoins, à force de prendre en charge certaines activités, ces techniques nous rendent dépendants. L'histoire de l'écriture et du livre en offre une illustration patente. Grâce à un système de signes, les mémoires individuelles surmontèrent les déficiences liées à l'âge, à la maladie ou à la mort, et participèrent à la transmission collective du patrimoine intellectuel commun. Sans aucun doute, l'humanité bénéficia grandement du développement des supports matériels de nos mémoires. C'est grâce aux parois des grottes et aux cailloux que l'art pariétal et rupestre se transmit. Les tablettes d'argile des Mésopotamiens ainsi que les papyrus égyptiens jouèrent un grand rôle pour notre connaissance des civilisations passées, les livres manuscrits puis imprimés aussi. Et de nos jours, les disques optiques et autres dispositifs numériques de stockage engrangent l'intégralité de notre mémoire collective.

Cependant, les mémoires individuelles pâtirent de ces développements. Nous ne mémorisons plus les conversations, car nous n'en avons plus besoin ; nous n'apprenons plus de poésies par cœur, car tous les poèmes nous demeurent accessibles immédiatement sur Internet. Nous disposons chacun de toute la connaissance de l'humanité à portée d'un clic de souris, et parallèlement, nos mémoires individuelles en contiennent de moins en moins. À défaut d'être bien faites, nos têtes deviennent de moins en moins pleines…

De même, la prolifération des machines à calculer à bas coût nous rend de moins en moins aptes au calcul mental et l'emploi généralisé du correcteur orthographique autorise nos incompétences, etc. Bref, en même temps qu'elle remédie à nos manquements et qu'elle accroît notre efficacité, la technologie nous fait perdre certaines compétences auxquelles elle supplée automatiquement à notre place. Ainsi, si nous devenons plus efficaces avec le concours de l'intelligence artificielle, nous risquons aussi d'en devenir plus bêtes... À force de se faire remplacer par les machines, n'assisterons-nous pas à une prise de pouvoir passive des machines, par simple démission des hommes ?

Ces questions, formulées depuis bien longtemps reçoivent, avec les technologies de l'information et de la communication, un nouvel éclairage. En effet, d'un côté, comme nous venons de le voir, le risque de dépossession d'un certain nombre de facultés intellectuelles devient de plus en plus grand, à mesure que les tâches réalisées au moyen de ces facultés s'automatisent. Mais d'un autre côté, les technologies de l'information et de la communication sollicitent de plus en plus notre intelligence. Avec leur concours, nous acquérons de nouvelles compétences. La mécanisation des tâches les plus répétitives et pénibles de nos activités professionnelles dégage un temps précieux, pendant lequel nous pouvons nous consacrer à l'étude et à la réflexion. Qui plus est, des champs d'investigation neufs, qui jusque-là nous étaient demeurés fermés du fait de nos limitations cognitives, deviennent praticables grâce à la mécanisation des procédures intellectuelles. Ainsi en va-t-il dans le secteur des mathématiques, où de nouveaux procédés de démonstration faisant appel aux ordinateurs ont permis de résoudre des conjectures demeurées problématiques

jusque-là – par exemple le théorème des quatre couleurs –, en explorant de façon systématique toute une multitude de combinaisons qu'il eût été trop fastidieux d'énumérer à la main. Plus généralement, on peut démontrer rigoureusement, avec des machines, des théorèmes qui demandent des preuves très longues, de plusieurs pages, là où la probabilité qu'une erreur se glisse dans une démonstration purement manuelle devient très grande. De même, en biologie moléculaire ou en linguistique de corpus ou encore dans le domaine du commerce en ligne, seule l'intelligence artificielle traite aujourd'hui les énormes quantités d'information disponibles que l'on appelle désormais les *Big Data*, ou en français les masses de données. Enfin, sur la toile, les moteurs de recherche comme Google sont fondés sur des techniques d'intelligence artificielle ; ces moteurs autorisent des recherches d'informations neuves et originales, en aspirant l'ensemble des pages disponibles sur Internet, puis en établissant le graphe des liens qui relient ces pages et en déterminant, à l'aide d'heuristiques judicieuses issues de travaux d'intelligence artificielle, les sites les plus populaires.

Ainsi, l'intelligence artificielle transforme la nature des sciences les plus traditionnelles : elle suscite des découvertes et des inventions, elle étend l'empire de nos connaissances, elle modifie la pensée. Elle nous rend donc globalement plus intelligents, au sens étymologique de ce mot, puisqu'elle établit des passerelles – plus exactement des liens, intelligence venant de *inter legere* en latin, littéralement « lier ensemble », autrement dit « comprendre », « jeter des ponts » – entre des données très éloignées les unes des autres, parce que perdues dans la masse des informations disponibles.

« Nous passerons bientôt de l'intelligence artificielle faible à l'intelligence artificielle forte. »

L'affrontement récent entre le programme informatique AlphaGo de DeepMind, intelligence artificielle de Google, et le champion sud-coréen et numéro trois mondial du jeu de go, Lee Sedol, a logiquement été décrit comme une étape supplémentaire du même drame menant inexorablement à l'avènement d'une intelligence artificielle (IA) forte susceptible d'égaler l'être humain et de le dépasser dans toutes ses activités.

« Vers une intelligence artificielle forte ? », sur le site Intelligence artificielle et transhumanisme, juin 2016

Aujourd'hui, il est courant d'affirmer qu'il existe deux types d'intelligence artificielle, une intelligence artificielle faible (*Weak Artificial Intelligence* en anglais) ou étroite *(Narrow Artificial Intelligence)* qui simulerait des facultés cognitives spécifiques comme la reconnaissance de la parole, la compréhension du langage naturel ou la conduite automobile, et une intelligence artificielle dite générale *(Artificial General Intelligence)* ou forte *(Strong Artificial Intelligence)* qui reproduirait un esprit, voire une conscience, sur une machine, et dont certains disent qu'elle adviendra bientôt et qu'elle aura des répercussions majeures, tout à la fois positives et négatives, sur le devenir de l'espèce humaine. Cependant, pour être en mesure d'apprécier le bien fondé des perspectives inquiétantes qui se profilent avec l'intelligence artificielle forte, il convient d'abord de faire l'archéo-

logie de la distinction entre intelligence artificielle forte et intelligence artificielle au sens classique, qualifiée aussi d'intelligence artificielle faible.

L'origine de cette différenciation remonte au début des années 1980, avec les travaux d'un philosophe américain, John Searle, qui mettait en cause les théories des philosophes dits cognitivistes selon lesquels le fonctionnement de l'esprit serait en tous points analogue à celui d'un ordinateur et pourrait être reproduit intégralement à l'aide de techniques d'intelligence artificielle. Comme ce philosophe professait une grande admiration pour les réalisations pratiques de l'intelligence artificielle et qu'il souhaitait pourtant en montrer les limites intrinsèques, il distingua deux formes d'intelligence artificielle, celle des ingénieurs susceptible de reproduire un grand nombre de fonctions cognitives avec une manipulation symbolique d'information, et celle des philosophes qui prétendaient restituer avec les techniques d'intelligence artificielle, l'esprit dans son intégralité, et en particulier la conscience.

Il appela la première l'« intelligence artificielle faible » et lui concédait tout d'un bloc : selon lui, elle parviendrait à des réalisations techniques inouïes. En revanche, la seconde, celle des philosophes, il la qualifia d'« intelligence artificielle forte » tout en la discréditant, car selon lui, elle serait incapable d'atteindre ses objectifs. Pour le montrer, il fit appel à l'expérience de pensée dite de « la chambre chinoise » car elle se déroule dans une geôle localisée en Chine, ou tout au moins dans un pays où personne ne parle l'anglais et où l'on dispose d'une écriture idéographique. Il y emprisonna un Américain qui, fidèle à sa réputation d'Américain, ne parle qu'une seule langue, l'anglais, et y plaça un panier,

qui contenait toutes sortes de carreaux de céramique sur lesquels étaient dessinés des idéogrammes impénétrables. Sur l'un des murs, il y avait un petit œilleton, par lequel le prisonnier voyait un bout d'extérieur, et une lucarne devant laquelle il pouvait présenter des carreaux de céramique. Il y avait aussi un grand livre avec des règles du type : si tel et tel caractères ont été observés dehors, il faut présenter à la lucarne tel et tel caractères du panier. Dernier point, on signifiait à l'Américain que s'il voulait manger, il fallait qu'il obéisse avec diligence aux injonctions du grand livre. Plaçons nous maintenant à l'extérieur quelques années plus tard, après que le prisonnier ait appris à manipuler parfaitement les carreaux de céramique et supposons que nous soyons Chinois : nous écrivons des messages sur des banderoles et le prisonnier répond de façon tout à fait pertinente à nos questions en présentant des caractères à sa lucarne. Cela nous induit à penser qu'il comprend parfaitement le chinois. Comment en douter ? Or, Searle nous affirme le contraire : selon lui, même après avoir manipulé pendant des années et sans erreur des carreaux de céramique pour dialoguer de façon pertinente avec ses interlocuteurs, cet Américain enfermé seul dans sa prison ne comprendra jamais un traitre mot de chinois. Son activité demeure d'ordre mécanique, ou, pour reprendre la dénomination de Searle, syntaxique, puisqu'elle obéit à des règles bien définies, mais il n'accèdera jamais au sens, à savoir à ce que Searle, en tant que linguiste, appelle la sémantique.

Toujours d'après John Searle, la manipulation de carreaux de céramique conformément à des règles inscrites sur un grand registre correspond très précisément à ce que fait un programme d'intelligence artificielle. Cela peut éventuelle-

ment produire une illusion de compréhension, comme lorsqu'on imagine qu'une machine passe le test de Turing, mais cela ne donnera jamais accès à la signification ou *a fortiori* à la conscience… En d'autres termes, l'intelligence artificielle faible, celle qui imite les facultés cognitives humaines, est susceptible de se réaliser, tandis que l'intelligence artificielle forte, celle qui vise à restituer un esprit et une conscience sur une machine avec des techniques d'intelligence artificielle, n'est qu'une bévue.

Dans les années qui suivirent son introduction par John Searle, la notion d'intelligence artificielle forte eut tant de succès qu'on lui assimila souvent l'intelligence artificielle dans son ensemble. Ceci se produisit tout particulièrement chez des philosophes peu soucieux de philologie qui inventèrent la notion de « bonne vieille IA » (GOFAI – *Good Old Fashioned Artificial Intelligence* en anglais) pour caractériser ce qu'ils imaginaient avoir été l'ambition démiurgique de l'intelligence artificielle des origines et ses méthodes élémentaires fondées exclusivement sur la manipulation symbolique d'information. Ils voyaient alors le texte de John Searle comme une critique des ambitions excessives de l'intelligence artificielle elle-même.

De façon assez paradoxale, des scientifiques, comme le roboticien Hans Moravec, leur emboitèrent le pas vers la fin des années 1980 en reprenant à leur compte l'intelligence artificielle forte pour affirmer que les méthodes de l'intelligence artificielle un peu renouvelées – ce qu'ils appelèrent la « nouvelle IA » en référence à la « nouvelle cuisine » – conduiraient à la construction de machines totalement intelligentes faisant écho aux machines ultra-intelligentes de la science fiction.

Quelques années plus tard, au début du XXIe siècle, se fit jour un autre courant qualifié d'intelligence artificielle générale (AGI, *Artificial General Intelligence* en anglais) qu'il ne faut surtout pas confondre avec l'intelligence artificielle (IA) des origines née il y a soixante ans. Ses promoteurs, parmi lesquels on peut citer entre autres Ben Goertzel, Marcus Hutter ou Jürgen Schmidhuber, désirent refonder l'intelligence artificielle sur des bases mathématiques solides, équivalentes en certitude à celles sur lesquelles s'appuie la physique. Dans ce but, certains d'entre eux recourent à la notion théorique de complexité de Kolmogorov et à la théorie de l'inférence inductive de Ray Solomonov. Grâce à ce qu'ils considèrent comme la pierre philosophale de l'intelligence artificielle, ils aspirent à formaliser toutes les formes d'apprentissage machine en les ramenant à la contraction ultime des observations, au sens de la complexité de Kolmogorov, autrement dit à un rasoir d'Occam parfait. Ils assurent avoir ainsi jeté les bases d'une science générale de l'intelligence. D'autres fondent leurs affirmations sur des principes différents d'apprentissage machine, par exemple sur l'apprentissage dans les réseaux de neurones formels, que l'on appelle l'apprentissage profond *(Deep Learning)*, ou sur l'apprentissage par renforcement. Comme, d'après les tenants de l'intelligence artificielle générale, tous les théorèmes mathématiques au fondement de cette science générale de l'intelligence s'avèrent démontrés, il s'ensuit que la réalisation d'une intelligence artificielle totale ne souffre d'aucune limitation et dépend uniquement de la capacité de calcul et de stockage des machines.

Sans nous étendre sur les fondements et les justifications génériques de l'intelligence artificielle générale, indiquons

que tandis que l'intelligence artificielle forte trouve son origine dans les travaux de philosophes, l'intelligence artificielle générale vient de travaux de physiciens théoriciens reconvertis. Même si elle reprend à son compte le projet de l'intelligence artificielle forte, elle s'appuie sur des travaux mathématiques plutôt fumeux et, parfois, sur des réalisations informatiques, alors qu'initialement l'intelligence artificielle forte reposait uniquement sur une justification d'ordre discursif.

Ajoutons que, quoique, de par leur dénomination, l'intelligence artificielle dite forte ou générale d'un côté et l'intelligence artificielle au sens premier – que l'on a rebaptisée intelligence artificielle faible – de l'autre apparaissent sœurs, tant la finalité que les méthodes de l'une et de l'autre diffèrent radicalement : là où nous avions une discipline scientifique fondée sur des simulations informatiques et sur leur validation expérimentale, nous trouvons des approches philosophiques ou mathématiques fondées uniquement sur des argumentations théoriques, sans vraie contrepartie empirique ; de plus, on insistait pour l'une sur la décomposition de l'intelligence en fonctions élémentaires reproductibles sur des ordinateurs, alors qu'on s'appuie pour l'autre sur la recomposition totale d'un esprit et d'une conscience à partir de fonctions cognitives élémentaires.

Reconsidérons maintenant, à la lumière de ce qui vient d'être dit, les pronostics sur la réalisation d'une intelligence artificielle forte qui viendrait supplanter l'intelligence artificielle faible : là où l'on était capable de mesurer, pas à pas, les progrès d'une discipline scientifique, avec des évaluations empiriques, l'intelligence artificielle forte et/ou générale s'imposent, l'une et l'autre, comme des professions de foi

auxquelles on adhère plus par conviction que par raison… En conséquence, le danger ou l'espoir que représente l'intelligence artificielle forte apparaît plus imaginaire que réel.

Au reste, notons que le terme d'intelligence artificielle forte qui avait été initialement introduit par John Searle pour montrer les absurdités auxquelles conduirait une extrapolation immodérée des possibilités de l'intelligence artificielle dans l'investigation de la conscience, au-delà de ses ambitions légitimes dans le domaine scientifique et technologique, est maintenant repris par des ingénieurs qui l'emploient à leur compte pour promouvoir leur technique et son pouvoir illimité, sans disposer d'autres arguments que l'auto-affirmation de leur pouvoir.

L'INTELLIGENCE ARTIFICIELLE, COMMENT ÇA FONCTIONNE ?

« Il n'y a rien à craindre avec les ordinateurs, il suffit de les débrancher. »

En tant que matérialiste, je considère que le cerveau et ce qui le caractérise, la pensée et la conscience, sont des phénomènes électriques et chimiques, d'une très grande complexité certes, mais des phénomènes tout ce qu'il y a de plus matériel. Il suffit de « débrancher l'appareil » pour qu'il s'arrête.

Pierre Tourev, « Du matérialisme à la révolte », site La Toupie, 5 novembre 2006

Les ordinateurs électroniques fonctionnent à l'électricité ; il suffit de cesser de les alimenter pour qu'ils arrêtent instantanément leur course. Il n'y a donc aucunement lieu de craindre qu'ils prennent le pouvoir, ni qu'ils nous réduisent en esclavage ou pire, qu'ils nous éliminent de la surface de la Terre. Bref, à moins que nous ne consentions délibérément à notre perte, il sera toujours temps de les mettre hors circuit et de reprendre la possession pleine et entière de toutes nos prérogatives. Ces préliminaires étant posés, quelques commentaires additionnels viennent moduler quelque peu ces propos apparemment rassurants.

En tout premier lieu, précisons qu'une telle conclusion ne vaut que pour les ordinateurs électroniques actuels fonctionnant à l'électricité, sur les principes physiques que nous connaissons et maîtrisons pleinement aujourd'hui. Il se pourrait qu'un jour des automates biologiques, fondés sur l'emploi de macromolécules recombinantes ou sur la greffe de neurones animaux sur des supports de silicium, viennent

supplanter les machines actuelles ; on parle aussi d'ordinateurs quantiques et du développement des nanotechnologies qui bouleverseraient la conception des calculateurs. Dans toutes ces éventualités, les principes matériels différeraient tant de ceux sur lesquels reposent les ordinateurs électroniques actuels que l'on ne peut se prononcer. En effet, il se pourrait que la diminution de la consommation énergétique et son changement de nature procurent une autonomie de fonctionnement totale aux nouvelles machines susceptibles, par exemple, de s'alimenter d'elles-mêmes en ingérant des herbes ou des racines.

En deuxième lieu, avec les progrès récents dans la conception des batteries et des capteurs solaires, les machines devenues partiellement autonomes subviennent, de plus en plus, à leurs propres besoins énergétiques. Ainsi, les robots que l'on envoie dans l'espace sur la Lune ou sur Mars, alimentent leurs batteries par conversion de la lumière solaire. Si l'un d'entre eux s'animait soudain d'intentions belliqueuses à notre égard, il faudrait procéder à une opération chirurgicale délicate afin de déconnecter ses batteries et/ou ses capteurs solaires… S'il suffit toujours de débrancher ces machines pour les stopper et couper court à leur conduite, un tel débranchement pourrait éventuellement se révéler assez acrobatique…

En troisième lieu, songeons à l'interconnexion actuelle de tous les ordinateurs sur le réseau Internet et à la part croissante qu'elle prend dans la régulation sociale. Une panne ou un dysfonctionnement des serveurs racines du réseau Internet, en particulier de ceux qui convertissent les noms de domaine en adresses d'ordinateurs et que l'on appelle les serveurs du DNS (*Domain Name System*, en français

« système des noms de domaines »), perturberaient grandement la vie collective, car on serait incapable de consulter les réseaux sociaux comme Facebook et les courriers électroniques ne parviendraient plus à destination. Il en irait de même si les ordinateurs qui passent des ordres en bourse se détérioraient. Quant à notre mémoire collective stockée sur des ordinateurs ou sur le nuage (*cloud* en anglais), comment s'en séparer ? L'administration se trouverait soudain totalement désorganisée car l'état civil disparaîtrait. Dès lors, nous ne parviendrions plus à prouver notre identité. Nous ne pourrions plus payer avec notre carte de crédit ni retirer de l'argent dans les guichets automatiques… Les opérations bancaires s'arrêteraient immédiatement. Les patients ne disposeraient plus de leurs dossiers médicaux et les médecins ne sauraient comment les reconstituer. Dans les hôpitaux, les régulateurs de pression des respirateurs artificiels qui recourent à des ordinateurs ne fonctionneraient plus ; il en irait identiquement des systèmes de suivi de l'électrocardiogramme ou de l'électroencéphalogramme des patients. D'ici peu, l'Internet des objets accroîtra plus encore notre dépendance à l'égard de cette interconnexion des réseaux informatiques : le réfrigérateur sera en contact avec le garde-manger et ils se concerteront pour établir la liste des denrées manquantes avant de passer des ordres d'achat chez l'épicier sur Internet. L'aspirateur, l'automobile et tous les objets usuels seront eux aussi connectés au réseau Internet. Le jour où tout cela se déréglera, on pourra s'attendre au pire…

Plus les ordinateurs prennent prise sur l'organisation de la société, moins leur déconnexion, ou pire encore, leur débranchement, deviendra possible. D'ailleurs, les nouvelles formes de délinquance ou de guerre sur le cyberespace tirent

grandement parti de cette nécessité de rester branchés, à tout prix. Ainsi en va-t-il des modalités actuelles de racket : il suffit de menacer des grandes sociétés d'attaques virales massives sur leurs réseaux informatiques pour obtenir la rançon exigée…

Aujourd'hui, le matérialisme s'impose comme la religion dominante. La plupart des humains de nos sociétés développées pensent que l'âme humaine s'éteint lorsque aucun influx nerveux ne parcourt notre cerveau. Et personne ne doute qu'un ordinateur s'arrête lorsqu'on coupe son alimentation électrique. Il suffit donc de le débrancher pour qu'il ne nuise plus. Tous en conviennent ! Cependant, il se pourrait que la société humaine rendue dépendante des machines refuse de s'en passer quoi qu'il advienne, parce qu'elle y perdrait sa cohésion et son sens. Il n'y aurait alors plus aucun moyen de débrancher les ordinateurs au risque, si on le faisait, de créer une crise sociale majeure… Nous devenons donc insensiblement de plus en plus vulnérables à une déconnexion du réseau et à un arrêt des ordinateurs.

L'Internet des objets

Avec la miniaturisation des processeurs, on dit parfois que l'ordinateur disparaît. Cela ne signifie pas qu'il s'absente, loin de là, mais il devient si petit qu'on ne le voit plus et qu'il se glisse à notre insu dans tous les objets du quotidien : montres, téléphones, voitures, vélos, livres, fours, réfrigérateurs, lunettes, stimulateurs cardiaques, etc. Bref, il disparaît parce qu'il n'est plus apparent et qu'il en profite pour proliférer à un rythme accéléré, puis se disséminer et faire irruption partout et à tout propos. Certains se demanderont sans doute : à quoi bon ? Les techniciens répondront qu'ils se couplent à toutes sortes de capteurs, capteurs de pression, de tension, de température, de mouvement, de vitesse, de position, etc., qu'ils enregistrent les mesures prises par ces capteurs, les stockent, puis les utilisent pour faire des suggestions, voire, dans certains cas, prendre des décisions. Avec eux, le monde se transforme. La forme extérieure des choses demeure, mais leurs fonctions évoluent. Ainsi, une montre reste une montre, mais elle se connecte et permet de téléphoner, d'avoir la météo, de lire ses textos, etc. Un réfrigérateur reste un réfrigérateur, mais il attire notre attention sur les denrées périmées et il passe éventuellement commande tout seul. De même, la voiture reste la voiture, en ce qu'elle nous transporte, mais elle échange désormais toutes sortes d'information avec son environnement, pour assurer notre sécurité. Cela suppose que les objets puissent communiquer avec leur environnement. À cette fin, on a étendu le réseau internet en donnant une adresse spécifique à chacun des objets que l'on souhaite y connecter : c'est ce que l'on appelle l'Internet des objets, ou l'Internet des choses, ou encore, en anglais, *Internet of Things* (IOT en abréviation). On imagine aisément tout le bénéfice que le consommateur tirera de ces technologies. On doit toutefois mettre en garde contre les risques pour la vie privée, car toutes nos actions et tous nos déplacements seront numérisés et transmis sur le réseau. À cela s'ajoutent les vulnérabilités induites par le réseau, lorsque d'habiles programmeurs mal intentionnés franchiront les barrières de protection et feront irruption chez vous dans vos objets familiers, par exemple dans votre voiture, dans votre montre, dans votre réfrigérateur, ce qui donne froid dans le dos, voire, si vous en avez un, dans votre stimulateur cardiaque !

« L'intelligence artificielle reproduit l'activité de notre cerveau. »

Longtemps, la robotique a navigué en eau trouble parce que, jusqu'au milieu des années quatre-vingt, inspirés par Turing, les tenants de l'intelligence artificielle dite « traditionnelle » n'ont cessé de vouloir copier le cerveau humain, d'essayer de reproduire des systèmes cognitifs complexes (langage, calcul, etc.), tout en essayant de donner un maximum de savoir à leurs créations.

Bruno D. Cot, « Les Robots darwiniens », *L'Express*, rubrique Technologie, 21 juin 2001,

Rappelons d'abord que la découverte de la structure fine de la matière cérébrale à la fin du XIX^e^ siècle par deux physiologistes, Camillo Golgi et Santiago Ramón Y Cajal, fut couronnée du prix Nobel en 1906. Dès 1870, Camillo Golgi savait visualiser, avec des techniques de coloration au nitrate d'argent, la composition des tissus cérébraux mettant ainsi en évidence des cellules que l'on appela les neurones. Quelques années plus tard, en reprenant la technique de Camillo Golgi, Santiago Ramón Y Cajal dessina la forme des neurones avec un « corps » d'où partent un ensemble de protubérances en saillie appelées les dendrites et un long filament, l'axone, le long duquel se propage l'influx nerveux. Ce dernier se termine par des excroissances en boutons appelées les synapses, qui se connectent avec les terminaisons dendritiques d'autres neurones.

En 1943, Warren McCulloch, logicien et neuro-psychiatre, et Walter Pitts, mathématicien âgé à l'époque d'à peine

20 ans, établissent un parallèle entre des réseaux d'automates élémentaires et ce que l'on connaissait à l'époque de la structure du tissu cérébral, à savoir d'un côté les neurones, qu'ils assimilent à des automates élémentaires, et d'un autre côté les liaisons dites synaptiques entre les boutons synaptiques des neurones et les terminaisons dendritiques d'autres neurones qu'ils rapprochent des connexions entres ces automates. Les automates dont il est question là sont des composants électroniques idéalisés susceptibles d'être reliés les uns aux autres, en accrochant les sorties des uns aux entrées des autres par des liaisons dites synaptiques en référence à la métaphore cérébrale. Chacun de ces automates possède deux états, disons noir ou blanc, dont la valeur est calculée par une procédure simple, dépendant uniquement de la valeur des entrées.

Il est loisible d'établir une analogie entre chaque automate et une proposition logique qui correspond à un énoncé susceptible d'être vrai ou faux, puis entre l'état des automates – noir ou blanc – et la véracité ou la fausseté des propositions qu'ils représentent. Ainsi, si l'on prend deux propositions logiques : « Le ciel est couvert » et « Il pleut », on peut leur associer deux automates, que nous appellerons A1 pour le premier et A2 pour le second. Et selon que l'automate A1 est dans l'état noir ou blanc, la proposition : « Le ciel est couvert » est vraie ou fausse ; de même, selon que l'automate A2 est noir ou blanc, la proposition : « Il pleut » est vraie ou fausse.

Les automates sont ensuite reliés entre eux, de façon à établir des liens entre les propositions qu'ils représentent. Par exemple, on peut imaginer que la sortie de l'automate A2 pointe vers l'entrée de l'automate A1. Pour simplifier,

McCulloch et Pitts ont supposé que les entrées et les sorties des automates étaient elles aussi binaires, noires ou blanches. Ainsi, si l'état de A2 est noir, alors la sortie de A2 sera noire et l'entrée de A1 le sera aussi ; en conséquence, à l'étape suivante, l'état de A1 sera noir. Ceci signifie que si la proposition représentée par A2, à savoir : « Il pleut », est vraie, alors il s'ensuit que la proposition représentée par A1, c'est-à-dire : « Le ciel est couvert », l'est aussi.

McCulloch et Pitts montrèrent que toute fonction logique qui prend ses valeurs dans un ensemble binaire comprenant deux éléments comme vrai et faux ou noir et blanc, quelle que soit sa complexité, peut être réalisée par un réseau d'automates semblables à ceux que nous venons de décrire. Ils sont même allés plus loin et ont prouvé que des réseaux très simples, qui contiennent deux couches d'automates en plus de la couche d'entrée, sont capables de réaliser toutes les fonctions booléennes*, à condition que le nombre d'automates sur la couche dite intermédiaire soit suffisant. Dès lors, puisqu'un réseau d'automates peut s'interpréter simultanément au plan logique, neurobiologique et technologique, le lien entre ce qui relève du cerveau et ce qui relève des lois de la pensée et du jugement, autrement dit du vrai et du faux, semblait à portée de main. Cette simulation du fonctionnement du cerveau à l'aide de réseaux d'automates enthousiasma les esprits ; cela donna naissance à la cybernétique* en 1946, puis au connexionnisme*, à partir du début des années 1980.

Est-ce à dire que l'intelligence artificielle reproduit le cerveau humain ? Non et ce pour deux raisons. En premier lieu, l'intelligence artificielle fut créée en 1956, au cours de l'école d'été du Dartmouth College, pour tenter de

surmonter les obstacles que rencontrait la cybernétique dans ses essais de simulation du comportement du cerveau. Plus exactement, dès les premières tentatives de modélisation du système nerveux, il apparut qu'il était nécessaire d'établir automatiquement des liaisons entre automates, afin de reproduire des fonctions complexes analogues à celles que produit notre cerveau. En d'autres termes, il fallait reproduire des mécanismes dits de plasticité synaptique, par lesquels les neurones s'auto-organisent en établissant dynamiquement des connexions entre eux. Or, l'auto-organisation des réseaux d'automates pose des problèmes mathématiques extrêmement difficiles à résoudre qui conduisirent certains scientifiques, dont les pionniers de l'intelligence artificielle, à envisager des solutions alternatives. Et dès son origine, l'intelligence artificielle suggéra d'autres pistes. Elle ne se restreignit pas à une modélisation de la physiologie du système nerveux central ; la prise de décision, autrement dit les mouvements du psychisme, la dynamique sociale et l'évolution des espèces prirent aussi le rôle de modèles.

En second lieu, les dimensions du cerveau et de la machine diffèrent considérablement. Songeons que le cerveau humain se compose d'environ 300 milliards de cellules, dont 100 milliards de neurones, qui participent principalement à la communication de l'influx nerveux, et des cellules gliales, deux fois plus nombreuses, mais moins volumineuses, et dont on ne connaît pas encore la fonction exacte, même si beaucoup pressentent qu'elle n'est pas négligeable. Les neurones se composent d'un corps cellulaire, ou soma, de forme sphérique ou ovoïde, d'une chevelure d'excroissances plus ou moins drues et nombreuses que l'on appelle les dendrites et d'un long filament, l'axone, qui conduit l'influx nerveux

sur une longue distance, depuis le soma. Chaque neurone comporte un seul corps et un seul axone, mais un écheveau de dendrites dont la forme et la taille varient considérablement. Depuis les extrémités de l'axone s'établissent des connexions, les liaisons synaptiques, vers les dendrites d'autres neurones. Ces liaisons véhiculent de l'information soit par l'intermédiaire de molécules chimiques, les neuromédiateurs – on parle alors de synapses chimiques – soit par potentiels électriques – ce qui forme des synapses électriques. Les synapses établissent un réseau très dense : on en compte de 1 000 à 10 000 par neurone. Ces quelques chiffres donnent une idée de l'intrication et des dimensions de l'édifice. Rappelons-les : environ 100 milliards de neurones reliés, chacun, à plus de 1 000 de leurs congénères…

Et si la taille du cerveau varie considérablement d'une espèce à une autre : 1,3 à 1,4 kg pour l'homme, 30 g pour un chat, 2 g pour un rat, le nombre de cellules demeure très élevé, même chez des animaux réputés peu dotés comme la pieuvre – 300 millions de neurones. Malgré les prouesses de la technologie des semi-conducteurs, nous sommes donc loin d'approcher cette complexité avec les ordinateurs contemporains et *a fortiori* d'émuler informatiquement la dynamique neuronale, même pour des espèces très primitives. En effet, les réseaux de neurones formels que l'on simule actuellement sur ordinateur comportent rarement plus de quelques milliers de neurones. Et même si, exceptionnellement, on atteint plusieurs dizaines de milliers de cellules avec des centaines de milliers, voire des millions de connexions, on ne parviendra jamais à simuler la dynamique d'un million de cellules et *a fortiori*, de cent millions ou de cent milliards de neurones. Bref, la conclusion est

sans appel : l'intelligence artificielle ne reproduit pas l'activité de notre cerveau !

Il existe cependant des projets scientifiques qui relèvent ce défi ; c'est en particulier le cas du projet Blue Brain de l'École polytechnique fédérale de Lausanne qui visait d'abord, à partir de 2005, à reproduire des fragments de cerveaux de rats, puis à partir de 2008, à simuler une colonne corticale, c'est-à-dire la base du cortex, avec des réseaux de neurones formels contenant plusieurs milliers de neurones. Depuis 2013, l'équipe du projet Blue Brain coordonne un grand projet européen, le *Human Brain Project*, financé à hauteur d'un peu plus d'un milliard d'euros (1,19 M€ exactement) en grande partie par la Communauté européenne et auquel sont affiliés plus de 90 instituts de recherche qui s'engageaient à simuler, d'ici 2024, le fonctionnement du cerveau humain avec un superordinateur.

Néanmoins, en 2014, à peine un an après le début du projet, parut dans la presse grand public une lettre ouverte signée par plus d'une centaine de chercheurs de grande renommée et adressée à la Communauté européenne. Elle mettait en cause à la fois la gouvernance du projet, son coût prohibitif et, surtout, la pertinence de ses objectifs. Cela montre, s'il en était besoin, que le projet de simulation du cerveau par un ordinateur fait débat dans la communauté scientifique, alors que la réalité de l'intelligence artificielle dans la vie courante est attestée.

« L'intelligence artificielle n'est pas naturelle. »

Quand bien même l'intelligence artificielle se développe de plus en plus par l'étude du comportement d'une machine, elle reste une imitation de l'intelligence naturelle de l'homme. Et dans ce cas-là, l'intelligence demeure spécifique à l'être humain.

« L'intelligence, faculté spécifiquement humaine »,
exposé sur le site Docs school, 2006

L'intelligence, qu'elle soit humaine ou animale, est naturelle. Beaucoup l'affirment. D'ailleurs, comment ne pas souscrire à une telle proposition lorsqu'on observe la nature dans la variété de ses manifestations spontanées, et que l'on constate les prodiges d'ingéniosité déployés par les êtres vivants dans leurs activités quotidiennes de chasse, de préservation d'eux-mêmes et de leurs espèces ou de sociabilité. Entendue comme capacité à surmonter les obstacles dressés par la contingence, l'intelligence appartient à toutes les espèces vivantes ; à ce titre, elle est éminemment naturelle. La vie elle-même se définit par la présence de désirs – autrement dit de buts ou de volontés – et par le déploiement de stratégies, plus ou moins complexes, destinées à les réaliser, c'est-à-dire par l'intelligence. En cela, l'intelligence caractérise la vie. Or, il n'y a aucune objection de principe à ce que le vivant, dans toutes ses dimensions, fasse l'objet d'une investigation rationnelle. En conséquence, ce que l'on appelle intelligence, et qui recouvre les facultés dont les

êtres vivants usent pour réaliser leurs objectifs, n'échappe pas au projet général d'élucidation rationnelle de la nature. Du moins, aucun argument solide ne permet d'exclure *a priori* l'étude de l'intelligence du champ des investigations de la science.

Or, le projet de l'intelligence artificielle porte justement sur ce point : il suppose que les différents aspects de l'intelligence, une fois observés et décrits avec précision, peuvent être simulés sur des ordinateurs. Il s'inscrit donc dans la perspective d'une investigation scientifique de l'intelligence telle qu'elle est énoncée plus haut ; et il propose de la mener en ayant recours aux ressources de l'informatique. On conçoit que des activités élémentaires, comme la mémorisation, la reconnaissance de visages, le calcul arithmétique ou le raisonnement logique, puissent être reproduites sur des automates déterministes à nombre fini d'états, c'est-à-dire sur des ordinateurs, car elles ont un caractère répétitif. Cependant, certains pensent que ce qui touche à l'art, à la spiritualité et au génie humain, autrement dit à ce que Blaise Pascal appelle « l'esprit de finesse », y échappera toujours. Peut-être ont-ils raison, peu importe, car tous les autres aspects de l'intelligence qui relèvent de ce que Pascal caractérise comme « l'esprit de géométrie » – et ils sont nombreux ! – demeurent à la portée de l'intelligence artificielle. Et grâce à leur simulation informatique, nous en comprenons mieux les ressorts ; sans compter que cette mécanisation de l'intelligence autorise une automatisation de tâches qui, autrement, auraient requis la présence d'un être humain. En somme, l'intelligence produite par l'intelligence artificielle ressemble, à certains égards, à l'intelligence naturelle, ou tout au moins, elle y aspire. Cette conclusion s'impose, indubitablement !

Ceci étant dit, une question demeure : l'intelligence artificielle, en tant que discipline de l'esprit, est-elle naturelle ? Plus généralement, la science et la technique sont-elles naturelles ? La puissance qu'elles octroient n'outrepasse-t-elle pas celle que procurent les facultés naturelles de l'espèce humaine et ne nous incite-t-elle pas à nous détourner de nos devoirs premiers ? Le pape Benoît XVI le déplore lorsqu'il affirme que « la vie contemporaine donne une place de choix à l'intelligence artificielle, toujours plus assujettie aux techniques expérimentales, oubliant dès lors que toute science devrait protéger l'humanité et promouvoir son élan vers l'authentique perfection. » (discours prononcé le 23 octobre 2006, lors de la rentrée académique à l'université pontificale du Latran). Devons-nous le suivre ?

Et avec cette première interrogation s'ouvre une seconde question relative aux abstractions formelles et/ou mathématiques qui nous aident à décrire le monde. Grâce à elles, nous établissons des ponts entre des réalités concrètes différentes, ce qui permet d'assujettir toujours plus la matière à nos besoins. Ce faisant, le risque est grand de nous asservir nous-mêmes, en réduisant nos aspirations à des désirs matériels. Qui plus est, ces abstractions n'ont rien de naturel : elles dénaturent notre intuition, elles nous éloignent du réel immédiatement sensible, elles nous séparent du monde auquel nous appartenons de moins en moins. Or ces abstractions formelles, que certains condamnent au nom de l'humanisme, relèvent de l'intelligence, au sens propre. En effet, si nous nous fions à l'étymologie, le mot « intelligence » vient de la racine indo-européenne *leg* – « cueillir », « choisir », « rassembler ». Cela donne en grec *legein* et ses dérivés qui signifient « rassembler » puis, par dérivation,

« dire », c'est-à-dire « rassembler les paroles », et, en latin *legere*, « cueillir, choisir, rassembler », d'où « lire », autrement dit assembler les lettres. Ainsi, originellement, l'intelligence tient à la réunion de choses supposées différentes, et à l'établissement de liens entre et au-delà des choses. L'intelligence correspond donc à ce travail d'abstraction du réel auquel nous nous livrons pour le maîtriser. Elle est propre à l'homme puisqu'elle tient, plus ou moins, à l'assemblage des paroles et des lettres, à savoir au dire et au lire. Et ce faisant, elle éloigne de la nature puisqu'elle est le fruit du travail, du savoir-faire et du génie humain. Bref, l'intelligence, en ce sens second d'établissement de relations d'analogie entre les choses, n'est pas naturelle. Et, *a fortiori*, l'intelligence artificielle qui reproduit l'intelligence au moyen de machines, n'a rien de naturel. Mais cela doit-il être mis au crédit ou au débit de l'intelligence en général ou de l'intelligence artificielle en particulier ? La question reste ouverte.

« Les ordinateurs raisonnent de façon binaire. »

Argument issu de la continuité du système nerveux :
Le système nerveux n'est certainement pas un automate à états discrets. Une petite erreur d'information sur la taille d'une impulsion nerveuse entrant dans un neurone peut faire une grande différence sur la taille de l'impulsion de sortie. Comme il en va ainsi, on peut affirmer qu'il n'est pas possible d'espérer simuler le comportement du système nerveux avec un système à états discrets.

Alan Turing, « Computing Machinery and Intelligence »,
Mind, 59, pp. 433-460, 1950

Certains se souviennent peut-être des frayeurs millénaristes qui inquiétèrent beaucoup de nos contemporains lors du passage à l'an 2000. Celles-ci ne tenaient qu'à cette fichue habitude de compter sur nos dix doigts, ce qui a imposé la numération en base dix. Convenons-en, à l'époque moderne, dans la société numérique, cela n'a plus de sens : hommes ou machines, archaïsme ou progrès, il faut choisir ! Et si l'on veut vraiment entrer dans la modernité, bannissons de telles conventions imprégnées d'un anthropomorphisme désuet !

Pour expurger tous ces résidus, un remède radical, simple et peu coûteux existe. Il est de mon devoir de vous l'indiquer, même s'il apparaîtra incongru aux yeux de certains : on devrait, à la naissance, couper les neuf doigts superfétatoires, pour que tous apprennent, dès le plus jeune âge, à compter en base 2 avec le doigt restant. Plié, tendu, plié,

tendu, plié, plié… 0,1,0,1,0,0, rien de plus facile. Et quoi de plus propice à l'avènement d'une vraie civilisation numérique dépouillé de tous ces inutiles oripeaux anthropomorphiques ! Qui plus est, on en profiterait pour instituer un rituel baptismal laïque qui serait du meilleur effet. Je vous vois hésiter. Vous reculez ? Pourtant, la modernité est à ce prix, sachez-le !

À moins que l'on ne s'engage sur la voie aride d'une conciliation entre les facultés humaines et les nécessités imposées par les techniques nouvelles. Mais la route toute semée d'embûches s'avère si longue à parcourir et si tortueuse que peu s'y aventurent. En accroissant l'aptitude des machines à lire, à calculer, à raisonner, à mémoriser, à maintenir leur attention, à percevoir, à agir ou à réagir, puis en fabriquant des machines qui répondent effectivement aux besoins et aux possibilités des hommes, sans les mutiler, l'intelligence artificielle nous engage sur ce chemin…

Or, à ce point, on adresse parfois une objection de principe à l'intelligence artificielle, qui tient justement à ce que les machines opèrent des calculs sur une représentation binaire. Selon certains, cela serait à l'origine d'une irréductibilité consubstantielle de notre cerveau et, en conséquence, de notre esprit, à ces machines.

En 1950, Turing répondit par anticipation à cette objection adressée au projet de l'intelligence artificielle. Pour bien comprendre sa réponse, il faut se reporter au contexte. À l'époque, certains craignaient que le discontinu n'atteigne jamais la fluidité du continu. Aujourd'hui, l'expérience quotidienne nous convainc tous du contraire : le son numérique l'emporte en fidélité sur n'importe quel son analogique ; et il en va de même pour l'image fixe ou animée…

En termes mathématiques, ceci signifie que toute fonction continue peut être « approximée » aussi finement qu'on le souhaite par une fonction numérique. D'ailleurs, un théorème démontré par Claude Shannon établit les caractéristiques de la fonction numérique, ou ce qu'on appelle, en termes techniques son échantillonnage, en fonction du degré d'approximation qu'on souhaite atteindre. En somme, l'objection fondée sur l'opposition entre le caractère discret, ou binaire, mais cela revient au même, des machines et la continuité supposée de notre système nerveux central est réfutée. On la réfute d'autant plus volontiers qu'il n'est pas certain que notre cerveau fonctionne sur un mode continu.

Mais une fois cette première objection invalidée, une seconde se fait jour : comment un ordinateur, qui est un automate à nombre fini d'états, simule une machine à nombre infini d'états ? Rien, en effet, n'assure que notre cerveau possède un nombre fini d'états ; loin de là, les résultats de la neurophysiologie contemporaine tendent à prouver le contraire. Sans compter que même si le nombre d'états de notre cerveau se révélait fini, il serait si grand qu'aucun ordinateur actuel ne serait en mesure de le simuler. À cette seconde objection, il convient d'en ajouter une troisième : les ordinateurs sont des automates déterministes, ce qui signifie qu'à une entrée donnée ne correspond qu'une seule sortie, toujours identique à elle-même. Et il se peut que le fonctionnement de notre cerveau échappe à un déterminisme parfait. En conséquence, l'ordinateur électronique contemporain, conçu comme un automate déterministe à nombre fini d'états, sera peut-être à jamais incapable de simuler notre cerveau. Ce n'est bien sûr qu'une éventualité ; cependant, on ne sait pas aujourd'hui la démontrer, l'infirmer

ou démontrer son contraire, à savoir qu'il serait éventuellement possible, un jour, de simuler le comportement de notre système nerveux central à l'aide d'ordinateurs.

Ces deux objections invalident-elles le projet de l'intelligence artificielle ? Une réponse négative s'impose et ce, pour deux raisons : d'une part, l'intelligence artificielle ne vise pas à reproduire le comportement du cerveau, mais simplement à simuler notre psychisme à l'aide d'une machine. De ce point de vue, il n'est pas nécessaire de reproduire une conscience pour faire une intelligence artificielle. Un agent matériel animé par un programme informatique assez simple, et avec lequel on dialogue comme avec un homme peut être considéré comme intelligent, même s'il ne saisit pas le sens de ce qui lui est dit.

Au reste, si les ordinateurs électroniques actuels sont toujours des automates déterministes à nombre d'états finis, il se peut qu'une nouvelle génération de machines non déterministes et/ou à nombre infini d'états apparaisse un jour. Les ordinateurs dits biologiques, fondés sur la recombinaison d'ADN ou sur la greffe de neurones de rats sur du silicium, constituent de telles tentatives ; les projets d'ordinateurs quantiques ou chimiques en sont d'autres. Beaucoup de chercheurs s'engagent aujourd'hui dans une telle voie, susceptible de repousser plus loin les capacités de simulation des machines, au point de reproduire, éventuellement, le comportement de notre cerveau. Mais répétons-le encore, ce projet de simulation va bien au-delà de celui de l'intelligence artificielle.

« Les ordinateurs ne se trompent jamais. »

Un jour viendra où quelque mécanicien de génie construira une machine toute semblable à l'homme, ce qui constitue peut-être le but inconscient de toute science. Je ne doute pas qu'une telle machine puisse désirer, marcher, rire et pleurer, parler, se souvenir et même oublier, mais je doute qu'elle puisse « se tromper » et non pas « faire des erreurs », ce qui est à la portée de tout le monde. Elle aurait alors acquis la fonction d'un Sujet et la capacité d'un transfert, sinon elle ne douterait de rien. C'est là que j'attends nos « ingénieurs ».

Michel Neyraut, *Le Transfert, le fil rouge*, 1974

Que ce soit un ordinateur, une broyeuse à chocolat ou un moteur à explosion, une machine ne saurait, à proprement parler, se tromper. Si elle doit refaire la même opération, dans les mêmes conditions, elle l'exécutera identiquement. Ainsi, une calculatrice donne toujours le même résultat.

Il en va de même avec une montre d'autrefois qui possède des aiguilles, des engrenages et des ressorts à spirale. Si on fait tourner l'aiguille des minutes d'un tour complet, l'aiguille des heures avance d'une heure exactement. En poursuivant et en ajoutant onze tours de l'aiguille des minutes, l'aiguille des heures fait un tour complet depuis sa position initiale… Il y a un lien direct entre le mouvement de la grande aiguille et celui de la petite. Et personne ne songe à s'émerveiller de cette concordance. Pourtant, c'est sur un principe tout semblable que sont construites les calculatrices depuis que Pascal, le mathématicien et philosophe français, les a inventées au XVII^e siècle.

Pour bien comprendre ce qui se produit, prenons un compteur kilométrique à roues, comme ceux qu'on trouve dans les anciennes voitures. En général, ce compteur comprend six chiffres, chacun porté sur une roue, un pour les centaines de mètres, un pour les kilomètres, un pour les dizaines de kilomètres, un pour les centaines de kilomètres, un pour les milliers de kilomètres et, enfin, un pour les dizaines de milliers de kilomètres. Personne ne s'étonne que, chaque fois qu'une roue revient à zéro, après un tour complet, la suivante avance d'un pas. Ainsi, si le compteur marque 129998, ce qui fait douze mille neuf cent quatre-vingt-dix-neuf kilomètres et huit cents mètres, cent mètres plus loin, il marquera 129999, soit douze mille neuf cent quatre-vingt-dix-neuf kilomètres et neuf cents mètres, et cent mètres plus loin encore, il marquera 130000, car la dernière roue, la plus à droite, passera de 9 à 0, entraînant sa voisine de gauche d'un pas, ce qui la fera à son tour passer de 9 à 0, la retenue se propagera ainsi sur les deux suivantes, qui passeront toutes deux à zéro, jusqu'à la deuxième roue sur la gauche, qui de deux passera à trois.

Ainsi, le compteur kilométrique ajoute un à chaque centaine de mètres. Mais selon un principe tout semblable, il peut ajouter n'importe quel nombre, il suffit de tourner les roues correspondant aux unités, aux dizaines, aux centaines, etc. Par exemple, si on veut ajouter 53 à 164, on commence d'abord par charger l'un des deux nombres à ajouter sur le compteur, par exemple 164. Pour cela, on met la roue la plus à droite sur la position 4, celle qui suit sur la gauche à 6, et enfin, celle encore plus à gauche sur 1, toutes les autres roues plus à gauche restant à zéro. Maintenant, pour ajouter 53, on fait avancer la roue la plus à droite, celle des unités

qui était à quatre, de trois pas. Elle se retrouvera donc au sept. Ensuite, on fait avancer la roue suivante de cinq pas, ce qui fait qu'on passe de 6 à 1 (il suffit de compter : sept, huit, neuf, dix, onze). Au cours de son mouvement, cette roue étant passée par zéro, elle fera avancer de 1 la roue immédiatement placée sur sa gauche. Celle-ci étant positionnée à un, elle passera donc à deux. Maintenant que l'opération est terminée, il suffit de lire le résultat sur les six roues à partir de la gauche : 0, 0, 0, 2, 1, 7, ce qui fait deux cent dix-sept. On vérifie bien qu'il s'agit du résultat exact de l'addition de 164 et de 53. Et on comprendra qu'il en aille de même pour l'addition de n'importe quelle paire de nombres.

De la même façon, on sait ramener les multiplications à des séries d'additions et de soustractions. Et dans tous les cas, les opérations s'exécutent parfaitement, sans qu'aucune erreur ne puisse se glisser dans les calculs. Aujourd'hui, les calculettes fonctionnent sur le même principe, les roues crantées ayant été remplacées par des circuits électroniques qui enchaînent les opérations et stockent les résultats. Et les ordinateurs ordonnent à des calculettes d'exécuter des séries d'opérations élémentaires de ce type, et ces calculettes renvoient toutes des résultats exacts.

En revanche, il se peut qu'on se trompe dans la programmation, et que les instructions que l'on donne aux ordinateurs ne correspondent pas exactement à ce qu'on souhaitait faire… C'est la source d'erreurs la plus commune. Donc, pour reprendre un vieil adage latin, *errare humanum est*, l'erreur est humaine ; la machine est parfaite, seul l'homme commet des erreurs. Mais seul l'homme est capable de corriger ses fautes… et seul l'homme sait déceler les insuffisances des machines et évaluer la pertinence ou l'inadéquation de leur utilisation…

« Les ordinateurs sont invincibles aux échecs et au go. »

Les progrès en informatique et en robotique ont été fulgurants. Ainsi, en 1997, l'ordinateur Deep Blue battait aux échecs le champion du monde Garry Kasparov. Il y a quelques années déjà, une équipe de Sony lançait un programme dont l'objectif ultime est de créer, pour 2050, onze androïdes capables… de battre l'équipe championne du monde de football ! Au rythme où va le progrès en la matière, qui peut dire de quoi seront capables les androïdes dans quelques années ? Et qui peut dire ce qu'ils pourront ressentir ?

Série de bandes dessinées *Pandora Box*, Éditions Dupuis

Les jeux mobilisent bien des énergies humaines. Un empereur chinois déplorait que les meilleurs de ses sujets passent tant de temps à étudier les différentes stratégies au go, au lieu d'enrichir et de défendre leur pays. Les fabricants de machines dépensèrent, eux aussi, beaucoup pour défier les humains. Emblématique, le jeu d'échecs nous en dit long sur l'histoire erratique de ces tentatives : le 11 mai 1997, un ordinateur a battu le champion du monde en titre aux échecs, Garry Kasparov. Plus récemment, en mars 2016, le programme AlphaGo conçu par la société DeepMind l'a emporté sur l'un des meilleurs joueurs au monde, Lee Sedol. Et, en janvier 2017, le programme d'intelligence artificielle Libratus, conçu par des chercheurs de l'université Carnegie Mellon, a dominé quatre joueurs professionnels au cours d'un tournoi organisé sur 20 jours aux États-Unis.

Épisodes remarquables dans la longue rivalité de l'homme et des machines, c'est, dans chacun de ces trois cas, l'aboutissement d'efforts notables et de bien des paris perdus.

Déjà, au XVIII[e] siècle, un homme, le baron Johann Wolfang Von Kempelen, prétendit avoir fabriqué une machine jouant aux échecs. Il la présenta dans la plupart des grandes cours d'Europe aux princes, à leurs ministres, aux seigneurs et à leurs domestiques, qui tous en furent éblouis. Plus tard, Edgar Poe, dans une merveilleuse nouvelle, *Le Joueur d'échecs de Maelzel*, a décrit le subterfuge par l'entremise duquel un nain, glissé à l'insu de tous dans l'intérieur de la machine, déplaçait les pièces de l'échiquier.

Au début du XX[e] siècle, un ingénieur espagnol, Torres Quevedo, construisit un automate électromécanique qui animait un bras mécanique et faisait glisser les pièces sur un échiquier. Cependant, le spectre d'action de cet automate était limité à des fins de parties du type roi contre roi plus tour. Nous étions encore loin de voir une machine l'emporter sur l'homme dans un tournoi.

Dans un article écrit en 1947, et intitulé « Intelligent Machinery », le défi a été relancé par Alan Turing. Trois ans plus tard, en 1950, il écrivit le premier programme informatique capable de jouer aux échecs, mais compte tenu de la lenteur des machines de l'époque, il en a simplement simulé le déroulement avec du papier et des crayons. La même année, un autre mathématicien légendaire, le père de la théorie de l'information, Claude Shannon, a conçu un plan d'action pour que les ordinateurs jouent aux échecs… Mais les obstacles techniques s'opposaient encore à tous ces efforts.

Le travail ne cessa pas pour autant, et huit ans plus tard, en 1958, un ordinateur a battu un homme au jeu d'échecs. Plus exactement, cet homme était une femme, la secrétaire de l'équipe de programmeurs qui avait conçu le programme de jeu. Et auparavant, cette femme n'avait jamais joué aux échecs. On lui avait simplement appris les règles une heure avant son tournoi et sa défaite contre l'ordinateur... Tout cela nous fait sourire aujourd'hui ; pourtant, à l'époque, c'était déjà considéré comme une victoire de la machine : un ordinateur était capable de vaincre un homme dans l'exercice d'une tâche complexe.

Certains se sont alors permis de faire des prédictions un peu inconsidérées : ainsi Herbert Simon, qui deviendra plus tard prix Nobel d'économie et Turing Award, a affirmé publiquement, en 1958, qu'avant dix ans, des ordinateurs battraient le champion du monde au jeu d'échecs, du moins si les règles ne leur interdisaient pas l'accès aux compétitions... Et depuis, des informaticiens n'ont eu de cesse de fabriquer des machines capables de vaincre les meilleurs d'entre les hommes au jeu d'échecs. Mais les choses sont allées plus lentement que prévu.

En 1966, alors que le terme des prédictions d'Herbert Simon approchait, un programme d'ordinateur fut battu par un enfant de dix ans. Et à la fin des années 1960, tandis que Spassky était le champion du monde en titre au jeu d'échecs, les programmes informatiques jouaient tout au plus comme d'honnêtes lycéens. Le philosophe Hubert Dreyfus, critique officiel de l'intelligence artificielle depuis 1965, en tira argument pour affirmer que l'intelligence artificielle ne surpasserait jamais l'homme.

Cela n'a pas suffi pour faire taire les prophètes modernes ; les prédications se firent entendre à nouveau dans le courant

des années 1970 avec un pari lancé entre David Lévy, classé « maître international » par la Fédération internationale d'échecs, et John McCarthy, l'un des pionniers de l'intelligence artificielle. L'enjeu était modeste : il affirmait qu'un ordinateur devait être en mesure de battre David Lévy dans un match avant la fin 1978. En 1978, le meilleur programme d'ordinateur de l'époque, « CHESS 4.7 » fut battu dans un match en cinq parties par David Lévy, avec trois victoires pour l'homme, une partie nulle et une victoire pour la machine. Les prédictions étaient encore démenties.

Mais les informaticiens ne désarmèrent pas. Et un peu plus de dix ans plus tard, en 1989, une machine, « Deep Thought », conçue par M. Hsu, l'emporta sur David Lévy. Les prédictions semblaient maintenant être réalisables, avec seulement un petit retard. Encouragée par cette victoire, l'équipe qui réalisa « Deep Thought » prétendit que le titre de champion du monde serait à sa portée d'ici deux ou trois ans. La même année eut lieu un match spectacle entre le champion du monde Garry Kasparov et « Deep Though », et la machine fut encore vaincue.

Hsu, le concepteur de « Deep Thought », fut alors embauché par la firme IBM. Son programme changea de nom, il s'appelait désormais « Deep Blue », et il était exécuté sur des machines plus puissantes. À Philadelphie, en février 1996, un match en six parties opposa Garry Kasparov et « Deep Blue ». Le champion du monde gagna trois parties sur six, la machine l'emportant uniquement sur une partie, la première, et laissant deux parties nulles. La machine était toujours battue.

L'année suivante, du 3 au 11 mai, Garry Kasparov affronta à nouveau « Deep Blue », toujours sur six parties. Il gagna

la première. « Deep Blue » emporta la deuxième, puis les trois suivantes furent nulles. Enfin, au cours de la dernière, Garry Kasparov déclara forfait après le 19e mouvement… Il s'avoua alors vaincu et troublé. Il dit avoir eu la sensation que la machine « lisait en lui ». « Deep Blue » avait battu le champion du monde en titre avec 3,5 points contre 2,5 pour l'homme. Les prédictions d'Herbert Simon avaient été atteintes avec un retard de moins de trente ans, ce qui apparaît bien faible au regard de l'histoire de l'humanité.

Mais quelle signification faut-il attribuer à ces victoires ? Les machines sont-elles devenues invincibles ? En toute rigueur non, car une marge d'incertitude demeure et il peut toujours se trouver un esprit supérieur qui parvienne à les vaincre. En effet, pour l'emporter contre tout partenaire imaginable, une machine devrait être en mesure d'énumérer exhaustivement toutes les parties. Or cette énumération est à jamais impossible, même à l'issue d'un accroissement inouï de la capacité des ordinateurs, car le nombre de parties possibles, évalué à 10^{120} par le mathématicien Claude Shannon, est supérieur au nombre de particules dans notre galaxie…

Si la machine qui l'a emporté sur Kasparov n'est pas invincible dans l'absolu, dépasse-t-elle à jamais les hommes, dans l'ordre de l'intelligence ? Peut-on imaginer que des hommes qui la fréquenteront quotidiennement apprendront, à son contact, à la dépasser ? Nul ne le saura, car les concepteurs de « Deep Blue » ont détruit leur machine, juste après sa victoire.

Qui plus est, à l'issue de cette victoire, d'autres jeux, en particulier le go, semblaient encore hors de portée

des ordinateurs. En effet, les programmes qui battent les hommes au jeu d'échecs sont fondés sur une anticipation des conséquences d'une petite proportion des coups possibles en fonction des réactions les plus probables de leurs adversaires. Même s'il n'est pas envisageable d'explorer toutes les positions, du fait de leur nombre, évalué à 10^{50} par Claude Shannon, la machine énumère tout de même un nombre assez conséquent de possibilités dont elle évalue les vertus en comptant le nombre de pièces présentes sur l'échiquier et en appréciant leur position. Cependant, dans le cas du go, cela s'avère inimaginable car il y en a trop, beaucoup trop. Songeons que le nombre de positions légales possibles est de l'ordre de 10^{170} et le nombre de parties de 10^{600}, c'est-à-dire 1 avec 600 zéros! À cela on doit ajouter qu'il est très difficile d'apprécier automatiquement la position des pièces, car cela repose sur des configurations de pions. Les techniques employées pour construire une machine qui joue au jeu d'échecs ne saurait donc suffire pour jouer au go.

Ces dernières années, on a d'abord développé des méthodes statistique dites de Monte-Carlo pour le go afin d'aider à évaluer les positions sur le *goban*, le tablier quadrillé sur lequel se joue la partie. Cela permit de faire des progrès considérables. Toutefois, il a fallu attendre 2008 pour que les programmes résultants battent des joueurs très bien classés. Ainsi, en 2008, le système MoGo gagne-t-il une partie face à Kim Myung Wan 8e dan, puis, en 2013, le système Crazy Stone bat Ishida Yoshio, 9e dan. Cependant, à l'époque, on était loin d'imaginer l'emporter sur les meilleurs joueurs…

Les choses se sont ensuite considérablement accélérées lorsque la petite société britannique DeepMind, rachetée

par Google, conçut le programme AlphaGo. Ce dernier utilise les techniques d'apprentissage machine, en particulier le couplage de l'apprentissage profond et de l'apprentissage par renforcement, pour analyser toutes les parties passées et se perfectionner en jouant contre lui-même, jusqu'à défier les meilleurs joueurs au monde. C'est ainsi qu'en octobre 2015, AlphaGo a défait Fan Hui, le plus fort joueur d'Europe, puis qu'en mars 2016, il l'a emporté, au terme d'un match historique, sur Lee Sedol, alors considéré comme un des meilleurs joueurs mondiaux. On peut donc désormais dire qu'après qu'une machine a été sacrée championne du monde au jeu d'échecs, une machine a été championne du monde au jeu de go.

Cependant, aussi puissant que soit les ordinateurs, on ne saura jamais explorer toutes les parties possibles que ce soit aux échecs ou *a fortiori* au go. Il n'est donc pas impossible qu'un homme un jour batte encore une machine à l'un de ces jeux… même si cela demeure aujourd'hui fort improbable. Enfin, il existe des jeux de stratégie où les hommes l'emportent encore, pour quelques années au moins, sur les machines…

« Le "*Deep Learning*" révolutionne l'intelligence artificielle. »

« Je n'ai jamais vu une révolution aussi rapide. On est passé d'un système un peu obscur à un système utilisé par des millions de personnes en seulement deux ans. » Yann LeCun, un des pionniers du « deep learning », n'en revient toujours pas. Après une longue traversée du désert, « l'apprentissage profond », qu'il a contribué à inventer, est désormais la méthode phare de l'intelligence artificielle.

Morgane Tual, *Le Monde*, 27 juillet 2015

Deep Learning, réseaux de neurones, apprentissage supervisé, apprentissage non supervisé, apprentissage par renforcement… ces notions semblent correspondre à un renouveau, voire à une révolution de l'intelligence artificielle. Qu'en est-il vraiment ?

Pour bien comprendre la signification de ces termes et mesurer leur portée, il faut se référer à l'histoire de leur apparition qui advint très tôt, au XX^e^ siècle, dans les années 1940 et 1950. Plus précisément, tout débute en 1943, avant même la construction des premiers ordinateurs électroniques, alors que l'on commence seulement à fabriquer des calculateurs électromécaniques, c'est-à-dire des machines capables de réaliser des opérations arithmétiques très rapidement, au moyen de relais téléphoniques. L'idée, assez naturelle somme toute, de dresser un parallèle entre ces machines et nos cerveaux traverse les pensées d'un brillant mathématicien, Walter Pitts, qui, âgé d'à peine vingt ans à l'époque,

écrit avec un neurologue, Warren McCulloch, un article intitulé *Un calculateur logique des idées immanentes dans l'activité nerveuse.* Ils y établissent une analogie entre d'un côté ces automates et les cellules de notre cerveau, les neurones, et d'un autre côté, les connexions de ces automates, et les liaisons dites synaptiques qui relient les neurones entre eux. Pour résumer ces analogies, on appelle « neurones formels » ces automates, et « synapses formelles », leurs connexions que l'on module d'un poids plus ou moins grand, selon la conductivité de la connexion. Outre la description de cette analogie, McCulloch et Pitts démontrent qu'en organisant ces neurones formels en trois couches, et en connectant les neurones formels de chaque couche avec des neurones formels de la couche suivante par des synapses formelles dont on ajuste correctement les poids, on peut réaliser n'importe quelle fonction logique. Avec ces réseaux de neurones formels, il existe donc un pont entre ce que les neurosciences disent du cerveau, la logique, c'est-à-dire les lois de la pensée, et l'ingénierie. Ce résultat majeur va donner naissance à une discipline nouvelle, la cybernétique*, qui ne se restreint pas à l'étude du cerveau, mais qui étend l'analogie à l'étude des lois de la régulation de tous les phénomènes complexes comprenant des entités communiquant entre elles, telles les lois qui régissent le vivant, par exemple la régulation des hormones, les interactions humaines dans les groupes ou, plus généralement, la société.

Toutefois, même si de tels réseaux de neurones formels organisés en trois couches permettent de réaliser n'importe quelle fonction logique, il convient de configurer les liaisons synaptiques entre les neurones formels, autrement dit

d'associer à chacune de ces liaisons un nombre, ce qui est extrêmement fastidieux, voire inextricable manuellement. On recherche donc, dès le début des années 1950, des procédures pour établir automatiquement les pondérations des liaisons entre les synapses formelles.

On imagine alors donner à une machine des exemples étiquetés pour qu'elle ajuste automatiquement les poids des synapses formelles afin de retrouver automatiquement les étiquettes des exemples : c'est ce qu'on appelle l'*apprentissage machine*. Plus précisément, cet apprentissage est dit *supervisé* car un « professeur » donne des exemples avec leurs étiquettes. À titre d'illustration, si l'on donne des formes géométriques comme des losanges, des polygones, des cercles ou des carrés, on indique à la machine que ce sont des losanges, des carrés, des pentagones, des cercles, des ellipses etc. Et on espère quelle sera en mesure de distinguer automatiquement cercles, ellipses, pentagones, quadrilatères si on lui donne suffisamment d'exemples ainsi étiquetés. Cela s'oppose à l'apprentissage dit *non supervisé* où les exemples sont donnés à la machine sans que l'on précise la catégorie dont ils relèvent. On souhaite alors construire des procédures capables de regrouper ensemble des exemples similaires et de dégager automatiquement des caractéristiques de ces groupes, comme les hommes le font dans la nature, par exemple lorsqu'ils distinguent, parmi les figures géométriques planes, les coniques et les polygones, sans que l'on ait évoqué ces notions auparavant. C'est là, à l'évidence, une tâche extrêmement délicate, bien plus que l'apprentissage supervisé, puisqu'elle relève de l'invention et pas uniquement de la description.

Dans les années 1940, puis au début des années 1950, même l'apprentissage supervisé apparaît quasiment insupportable, et les nombreuses tentatives qui ont été faites

se sont soldées par des échecs. Heureusement, à la fin des années 1950, deux événements vinrent faire évoluer la situation, sans toutefois la débloquer totalement.

En 1957, un psychologue américain, Frank Rosenblatt, met au point un algorithme d'apprentissage pour des réseaux de neurones formels à deux couches qu'il appelle des perceptrons, car ils reproduisent selon lui les capacités de perception des rétines. Deux ans plus tard, en 1959, un ingénieur, Arthur Samuel, dote un programme informatique conçu pour jouer au jeu de dames de capacités d'*apprentissage par renforcement.* Il ne s'agissait là ni d'apprentissage supervisé ni d'apprentissage non supervisé, mais de permettre à la machine d'améliorer progressivement ses performances grâce à l'expérience. Le principe de cette forme d'apprentissage repose sur l'existence de récompenses et de punitions que renvoie l'environnement. L'apprentissage essaie de maximiser la somme des récompenses espérées tout en minimisant les punitions, en supposant que les récompenses et les punitions se distribueront de façon statistiquement identique dans le futur à ce qu'elles étaient dans le passé. Dans le cas du jeu de dames, les récompenses correspondent aux gains, par exemple au nombre de pièces adverses prises, et les punitions, de façon symétrique, au nombre de pièces perdues. Ce principe se généralise aisément. Ainsi, imaginons un aspirateur. Pour qu'il soit efficace, on lui donnera une « récompense » lorsqu'il aspirera de la poussière, autrement dit on l'imagine amateur de poussière, c'est-à-dire « poussiérophile ». Et, grâce à l'apprentissage par renforcement, après un temps d'adaptation, il ira visiter en priorité les lieux où son expérience lui a appris qu'il y a plus de poussière.

Bref, les notions essentielles de l'apprentissage machine, qu'il soit supervisé, non supervisé ou par renforcement,

existent depuis plus d'un demi-siècle. On peut donc se demander pourquoi il a fallu attendre autant de temps avant que le grand public ne mesure l'importance qu'elles ont pour la société. Cela tient à trois facteurs : deux facteurs techniques liés, l'un, à la puissance de calcul des machines et, l'autre, à leur capacité à générer et à engranger de grandes quantités de données, ce que l'on appelle les *Big Data* (« masses de données » en français), en particulier grâce au web ; ainsi qu'à un facteur d'ordre scientifique dû à l'amélioration des algorithmes d'apprentissage.

Plus précisément, au plan scientifique, la procédure d'apprentissage mise au point en 1957 par Frank Rosenblatt pour les perceptrons ne marchait que pour des réseaux de neurones formels à deux couches. Or, si Walter Pitts avait bien montré que les réseaux de neurones à trois couches pouvaient réaliser n'importe quelle fonction logique, il n'en va pas de même pour les réseaux à deux couches, tant s'en faut. En 1969, Marvin Minsky démontra que la procédure d'apprentissage décrite par Frank Rosenblatt n'apprend que des fonctions très simples, dites linéairement séparables. Il a fallu attendre le début des années 1980 pour que des mathématiciens conçoivent une procédure d'apprentissage capable d'apprendre sur des réseaux de neurones à plusieurs couches. En termes techniques, on appelle cette procédure la *rétro-propagation du gradient*. Quelques années plus tard, d'autres mathématiciens cherchèrent à trouver les fondements théoriques de cet apprentissage. Cela les conduisit à développer d'autres techniques d'apprentissage supervisé inspirées des principes mathématiques sur lesquels reposent l'apprentissage dans les réseaux de neurones formels comme ce que l'on appelle les machines à vecteurs de support *(Support*

Vector Machine) et les machines à noyaux *(Kernel Machine),* qui furent bien souvent utilisées dans les années 1990 et au début des années 2000. Puis, à partir des années 2010, on se remit à faire de l'apprentissage avec des réseaux de neurones classiques, mais on put recourir, grâce à la grande puissance de calcul des machines, à un très grand nombre de connexions, de l'ordre de plusieurs centaines de milliers, voire de quelques millions, et beaucoup de couches – entre 10 et 15 – dont certaines restent « figées » en ce sens que les poids synaptiques y demeurent fixes, tandis que d'autres évoluent par apprentissage. En raison de cette multiplicité de couches, on caractérise ces techniques comme de l'apprentissage profond *(Deep Learning).* La comparaison des capacités d'apprentissage du *Deep Learning* avec celles des autres techniques d'apprentissage supervisé sur des tâches de reconnaissance d'images, a montré que les techniques de l'apprentissage profond apprennent de façon efficace sur de très grandes quantités d'exemples tout en surpassant les performances des autres techniques. Ainsi, en entraînant par apprentissage profond des algorithmes de reconnaissance faciale sur 200 millions d'images de visages, le système FaceNet de la société Google obtient un taux d'identification correcte de 99,63 %. C'est la raison pour laquelle elles se sont imposées ces dernières années.

Cela étant, bien qu'elles apparaissent neuves, les techniques d'apprentissage profond reprennent les principes anciens d'apprentissage par rétro-propagation du gradient, introduits dans les années 1980, qui, eux-mêmes, reprennent les principes d'apprentissage des perceptrons apparus dans les années 1950 qu'ils généralisent et mettent en œuvre sur ces masses de données que l'on qualifie, à défaut de trouver d'autres noms pour les caractériser que leur taille, de *Big Data,* littéralement « grosses données ».

DES MACHINES ET DES HOMMES

« Une machine ne peut pas être créative. »

La machine analytique n'a aucunement la prétention de créer quoi que ce soit. Elle peut faire tout ce que nous savons lui ordonner d'exécuter. Elle peut accomplir une analyse ; mais elle n'a pas le pouvoir d'anticiper une relation analytique ou des vérités. Sa fonction est de nous assister en mettant à notre disposition ce que nous connaissons déjà.

Ada Lovelace, notes de la traductrice du mémoire
de Luigi Menabrea sur la machine analytique inventée
par Charles Babbage (1842)

En va-t-il de la créativité comme du bon sens qui, selon Descartes, serait la chose du monde la mieux partagée ? L'expérience montre que partout et toujours, dans leur langage, dans leurs inventions, dans leurs organisations sociales, dans le mal comme dans le bien, les hommes ont été spontanément créatifs. Cela signifie-t-il que la créativité est le propre de l'homme ? Si oui, qu'y aurait-il de si magique dans la créativité ? Si non, une machine pourrait-elle faire preuve de créativité ? Mais dans cette éventualité, qu'en serait-il de la singularité individuelle de chaque créateur qui semble si étroitement liée à son génie propre ?

Cette question liminaire introduit tout naturellement les apories auxquelles se confrontent les spécialistes d'intelligence artificielle lorsqu'ils se risquent à aborder la créativité et sa simulation sur ordinateur. D'ailleurs, la possibilité même d'une originalité pour une machine posa problème bien avant l'existence des ordinateurs, au début

du XIX^e^ siècle, et cette question fut reprise au plan théorique par Alan Turing au milieu du XX^e^ siècle, qui critiqua l'argument dit d'Ada Lovelace selon lequel une machine semblable à la machine analytique de Babbage, autrement dit à un ordinateur, ne pourrait pas faire preuve de créativité. Et c'est là l'une des objections les plus couramment retenues contre l'intelligence artificielle. En dépit de cette objection, les spécialistes d'intelligence artificielle n'ont pas renoncé.

Pour eux, la recherche, la découverte scientifique, la créativité, l'originalité se présentent comme autant d'aspects de la vie sociale que l'on ne saurait détacher des autres : il existe des métiers universitaires, des enseignements de préparation à la recherche, et l'on reconnaît certains professionnels de la science pour leurs compétences propres. Bref, en dépit du prestige qui les auréole, aucune raison *a priori* ne permet d'exclure les activités intellectuelles et artistiques du champ des activités ordinaires et, en conséquence, du domaine de ce qui est susceptible d'être simulé par une machine.

L'intelligence artificielle a en premier lieu abordé l'activité intellectuelle des chercheurs et inventeurs à partir des modèles classiques de la résolution de problèmes qu'elle avait élaborés auparavant et qui se fondent sur la métaphore d'un labyrinthe à explorer. Bien évidemment, les découvreurs ne se contentent pas d'explorer des labyrinthes avec succès, du moins si l'on conçoit un labyrinthe comme un simple dédale de couloirs avec une entrée et une sortie. Toutefois, si le labyrinthe devient une figure abstraite, autrement dit s'il désigne métaphoriquement une recherche dans un écheveau de possibles, alors on imagine aisément que toute résolution de problèmes s'assimile à un parcours dans un labyrinthe et, par suite, que l'activité scientifique

soit vue comme une simple résolution de problèmes, ou comme une succession de résolutions de problèmes. L'argument principal à l'appui de cette thèse tient à ce que les grands découvreurs possèdent toujours une forte aptitude à résoudre des problèmes.

Depuis cinquante ans, cette hypothèse, à savoir que la découverte et l'activité scientifiques sont réductibles à de la résolution de problèmes, a été étayée par les travaux d'un certain nombre de chercheurs œuvrant en intelligence artificielle, dans le champ qu'ils ont convenu d'appeler la « découverte scientifique ». Les premiers travaux conduits dans cette direction l'ont été en 1971, avec le système Dendral qui analyse automatiquement la sortie d'un spectrographe de masse, à la façon d'un chimiste expert. En l'occurrence, l'expert était un prix Nobel de chimie… Dans un ordre d'idée tout différent, le programme AM de Douglas Lenat construit des concepts mathématiques à partir de notions pré-numériques de sac, de séquence, etc., issues des travaux de Piaget sur la psychologie du nombre chez l'enfant, puis il engendre des conjectures portant sur ces concepts. Après avoir « découvert » les nombres, puis les nombres premiers, ce programme émit de lui-même la conjecture dite de Goldbach, selon laquelle tout nombre pair est somme de deux nombres premiers. Sans nous étendre sur ce programme, indiquons que là encore la découverte procède de parcours de graphes, c'est-à-dire d'exploration de labyrinthes : des notions élémentaires – comme les notions de sac, de séquence, etc. – sont combinées entre elles pour donner naissance à des concepts nouveaux, eux-mêmes combinés entre eux et ainsi de suite.

Et sur le même modèle, quatre programmes ont été conçus par Langley, Zytkow, Simon et Bradshaw, et nommés en

hommage à quatre grands scientifiques du passé : Bacon, Glauber, Stahl et Dalton. Leur but est de reconstituer artificiellement, sur des ordinateurs, les démarches scientifiques de ces chercheurs.

Mais les spécialistes d'intelligence artificielle ne se contentent pas de reconstruire des inventions et des découvertes anciennes, autrement dit des choses qui ont déjà été faites ; ils ont aussi fabriqué des machines douées de créativité, c'est-à-dire capables de créer d'elles-mêmes, au sens étymologique de faire croître du nouveau.

Cette question fut abordée de deux façons différentes. Tout d'abord, une approche, que l'on peut qualifier de « logico-mathématique », assimile toutes les facultés mentales à la démonstration de théorèmes mathématiques. Dans ce cadre, la capacité créative est vue comme la faculté d'inventer des notions nouvelles, qui ne sont pas fournies explicitement par les hommes qui conçoivent et réalisent les machines. Un exemple particulièrement illustratif est donné par l'énigme de l'échiquier tronqué : une case étant supprimée à chacune des extrémités d'une diagonale d'un échiquier de huit cases sur huit, on se demande s'il est possible de couvrir toutes les 62 cases de l'échiquier ainsi tronqué avec des dominos qui couvrent chacun exactement deux cases. La solution, ou plutôt l'absence de solution, apparaît évidente si l'on colore les cases de l'échiquier. On se rend alors immédiatement compte que le nombre de cases blanches est différent du nombre de cases noires, ce qui interdit une couverture de l'échiquier par des dominos bicolores. Si maintenant nous posons le même problème à une machine, sans indiquer que les cases sont colorées, celle-ci serait-elle capable d'inventer une notion analogue

à celle de couleur ? Si oui, nous pouvons dire que nous avons affaire à une machine créative… Or les techniques modernes de démonstration automatique de théorèmes montrent qu'il est possible de construire une telle machine sans explorer tous les possibles.

Il existe aussi des approches psychologiques de la créativité qui ont cours dans le courant dit sémantique de l'intelligence artificielle. Dans ce cadre, toute imagination est vue comme la recombinaison d'éléments de mémoire préexistants. La création, expression concrète de l'imagination, fait alors appel à la mémoire. Et la simulation de nos capacités créatives passe par une modélisation de nos mémoires et des processus de réminiscence associés, sans que le terme de mémoire employé ici se résume à de simples dispositifs de stockage d'informations. Plus exactement, cette modélisation fait d'abord appel à des processus de récupération des matériaux stockés, puis à des mécanismes de « rapiéçage » de ces morceaux. À titre d'illustration, la licorne, produit par excellence de notre imagination, se présente comme la combinaison de deux êtres réels dont nous conservons en nous le souvenir : le cheval et le narval. Le travail de l'imagination qui crée un être fictif comme la licorne passe par la récupération des souvenirs anciens et par leur « raboutage ». Ce principe a été généralisé et simulé sur ordinateur. C'est ainsi que les capacités créatives d'un bassiste de jazz qui improvise dans un trio rythmique ont pu être mimées par l'emploi de rythmes ou de mélodies connus, puis par une combinaison judicieuse de ces éléments, dans le contexte du jeu.

Bref, la création, du moins certaines formes de création scientifique ou artistique, sont simulées sur ordinateur. Cela

veut dire qu'il n'y a rien de magique là, et que toutes nos capacités, même les plus étonnantes, peuvent faire l'objet d'études scientifiques et de simulations au moyen d'ordinateurs. Pourtant, cela ne signifie aucunement que nous avons réduit l'homme à une simple mécanique ; l'ordinateur nous permet au contraire d'en saisir la richesse et de mesurer l'écart entre ce que nous comprenons aujourd'hui et tout ce qu'il y aurait à comprendre.

Ainsi, la simulation de la découverte et de la créativité sur ordinateur ne doit pas être vue comme une fin en soi. Il n'y a là ni fermeture ni clôture, mais au contraire un aiguillon, c'est-à-dire, au sens étymologique, une stimulation dont il reste à recueillir les fruits.

L'argument de Lady Ada Lovelace

Au milieu du XIX^e siècle, un homme, Charles Babbage, imagina une machine capable d'enchaîner automatiquement des opérations logiques et mathématiques. Il en dessina les plans puis il passa de nombreuses années à essayer de la fabriquer. Faute d'argent, il n'y parvient pas. Plus tard, dans la seconde moitié du XX^e siècle, les plans en ont été repris et cette machine, que Babbage appelait le « moulin », a été construite ; elle fonctionne désormais parfaitement ; c'est un ordinateur mécanique…

L'assistante de Charles Babbage, Lady Ada Lovelace, s'essaya à rédiger quelques programmes capables de faire fonctionner ce « moulin » au cas où il serait réalisé. En réponse aux sollicitations des curieux qui spéculaient sur les pouvoirs de cette machine extraordinaire, elle expliqua que le « moulin » de Babbage ne faisait qu'exécuter mécaniquement les instructions qui lui étaient fournies. Ainsi, d'après elle, il n'était doué d'aucune spontanéité et ne serait donc pas en mesure de faire preuve de créativité.

Lorsqu'en 1950, Alan Turing rédigea son fameux article sur la capacité des machines à penser, il répondit à certaines des objections couramment adressées à ceux qui prétendaient qu'une machine pourrait penser. Selon l'une d'entre elles, une machine ne fait jamais qu'exécuter les opérations qu'un homme a programmées. En conséquence, elle n'est pas créative, puisqu'elle est réduite à exécuter fidèlement les ordres donnés. Alan Turing rapprocha cette objection de la position de Lady Ada Lovelace. Or, d'après lui, celle-ci n'est pas tout à fait exacte. En effet, même si une machine est parfaitement déterministe et si elle se plie scrupuleusement aux instructions qu'on lui a fournies, elle surprend ceux qui l'ont programmée. Et, de cet effet de surprise, peuvent naître des comportements étonnants, analogues aux comportements créatifs d'enfants, de scientifiques ou d'artistes.

« Les machines n'ont pas d'émotions ni de conscience. »

On ne conviendra pas qu'une machine équivaut à un cerveau avant qu'une machine écrive un sonnet ou compose un concerto à partir de pensées et d'émotions ressenties, et non par le hasard d'une combinaison de symboles, c'est-à-dire qu'elle n'écrive pas seulement, mais qu'elle sache ce qu'elle a écrit. Aucun mécanisme ne pourrait ressentir (et pas seulement le signaler artificiellement par un stratagème facile) ***le plaisir de la réussite, la douleur d'un spasme, le réconfort suscité par la flatterie, la confusion provoquée par ses propres erreurs, le charme du sexe, la colère ou la déprime causée lorsque l'on n'obtient pas ce que l'on souhaite.***

Discours solennel prononcé par le professeur Jefferson Lister en 1949, cité par Alan Turing dans « Computing Machinery and Intelligence », *Mind*, 59, pp. 433-460, 1950

Déjà, en 1950, lorsque Turing écrivit son célèbre article sur l'intelligence des calculateurs cité en exergue, l'un des principaux arguments que l'on opposait à l'idée qu'une machine puisse penser portait sur leur absence d'émotions et de conscience. Cette question mérite donc que l'on s'y attarde quelque peu et qu'on la détaille. Sans compter qu'une réponse hâtive risquerait de fourvoyer le lecteur. En effet, cette question recèle en elle au moins trois interrogations d'ordres différents. La première porte sur la possibilité de fabriquer une machine douée de conscience. La deuxième est relative à la nécessité qu'une machine éprouve des émotions pour produire de l'intelligence artificielle.

Enfin, la troisième invite à réfléchir au rôle éventuel joué par les émotions en intelligence artificielle. Prenons ces questions l'une après l'autre.

Avant d'affirmer et de trancher, il faut parfois prendre les problèmes à leur racine, au risque de semer le trouble. En l'occurrence, avant de se prononcer sur l'éventuelle conscience des machines, demandons-nous comment nous savons qu'une machine éprouve ou n'éprouve pas d'émotions et possède ou ne possède pas de conscience ? Plus généralement, comment sait-on que les pierres précieuses ou les cailloux ordinaires, les plantes, les animaux, les femmes et les hommes autres que nous-mêmes, ont ou n'ont pas de conscience ? Un philosophe, Thomas Nagel, dans un article célèbre publié en 1974 et intitulé « What it is Like to Be a Bat ? » (C'est comment d'être une chauve-souris ?), a imaginé une expérience de pensée dans laquelle il se retrouverait dans la peau et dans le corps d'une chauve-souris. Pourrions-nous, en tant que femmes ou hommes, comprendre ce que pensent des chauves-souris qui n'ont jamais partagé notre expérience humaine ? Que ressentent-elles ? Comment perçoivent-elles le monde ? Que le lecteur ferme les yeux et qu'il songe à ce que signifie vivre toute la journée accroché par les pattes à la voûte d'une grotte, voler la nuit venue, traquer les moustiques pour se nourrir, être aveugle-né, mais percevoir le monde par l'écho des ultrasons que vous émettez, etc. Le rouge ou le vert existent-ils pour de tels êtres ? La musique a-t-elle un sens ? Pourtant, même sans avoir accès aux sensations qu'éprouvent les chauves-souris, beaucoup d'entre nous leur prêtent des émotions et une conscience.

Considérons maintenant une machine perfectionnée. Rien ne nous dit qu'elle ne dispose pas d'une forme singulière de

conscience à laquelle nous n'avons pas accès. Mais quoi qu'il en soit de l'existence de celle-ci, elle nous restera certainement assez étrangère et pour longtemps, du moins tant que nous ne nous transformerons pas nous-mêmes en machines. Ainsi, l'élucidation des émotions des machines et l'accès à leur(s) conscience(s) semblent actuellement difficiles, car ce ne sont pour nous que des assemblages de composants matériels qui ne se constituent pas en entités organiques capables de se percevoir comme des sujets. Toutefois, en dépit de tous ces arguments, nous ne sommes pas en mesure d'affirmer avec certitude qu'elles se trouvent dénuées de conscience.

Venons-en maintenant à la deuxième question, sur le rôle de la conscience et des émotions dans la fabrication des machines dites intelligentes. Rappelons, à ce propos, que le projet de l'intelligence artificielle vise la simulation des facultés cognitives de l'esprit en général, qu'il soit humain ou animal, par exemple de la capacité à parler, à démontrer des théorèmes, à jouer aux échecs, à reconnaître des visages, etc. Nous avons vu, à cet égard, que le test d'intelligence des machines imaginé par Alan Turing dans son fameux article écrit en 1950 se déroule dans un théâtre d'illusions avec le « jeu de l'imitation ». Là, une machine se fait passer pour ce qu'elle n'est pas, en répondant aux questions qui lui sont posées. Or, pour accomplir chacune des tâches que nous avons énumérées, par exemple, répondre à des questions ou jouer aux échecs, point n'est besoin d'éprouver des émotions, des sentiments ou de l'empathie. Il suffit simplement qu'à une entrée donnée corresponde une sortie appropriée. Ainsi, si l'on donne un théorème à démontrer à une machine, il suffit qu'elle fournisse la succession des mani-

pulations formelles d'expressions symboliques qui, partant des axiomes parvient au théorème, pour être réputée l'avoir démontré.

Bref, la réponse à la deuxième question s'impose : l'intelligence artificielle n'a pas besoin de doter les machines d'émotions ni de conscience pour parvenir à des résultats tangibles et démontrables. Dès que, dans le temps de l'action, une machine se comporte comme si un être intelligent l'animait, on parle d'intelligence artificielle. C'est là une approche expérimentale et pragmatique, qui vise à reproduire sur une machine des facultés intellectuelles. Le malentendu tient à ce que beaucoup croient que l'intelligence artificielle réduit nécessairement l'intégralité de l'intelligence, de la conscience, de la créativité à des mécanismes élémentaires qu'une machine peut exécuter. Dès lors, beaucoup affirment qu'il reste toujours quelque chose d'irréductible et donc que l'intelligence artificielle ne peut être qu'un fiasco ; peut-être ont-ils raison d'affirmer que certaines dimensions de l'esprit échappent à la machine, mais cela n'empêche pas de commencer à reproduire tout ce que nous pouvons. Nous verrons bien un jour si l'on est capable de déterminer, par avance, ce qui échappera à jamais à la reproduction mécanique. Pour l'heure, rien ne permet encore de le faire !

La troisième question touche au rôle des émotions en intelligence artificielle. Contrairement à une idée assez répandue, selon laquelle l'intelligence des machines serait froide, voire glacée et « métallique », le rôle des émotions est indubitable. En effet, l'intelligence d'une machine ne correspond pas à une propriété intrinsèque ; la seule intelligence des machines, c'est celle que nous leur attribuons.

L'informatique affective

Désormais, les ordinateurs prolifèrent et se miniaturisent, au point de se glisser dans nos objets les plus familiers et d'en changer quelque peu la fonction. Un agenda, un téléphone, une montre, une caméra contiennent des ordinateurs miniatures. Il en va de même avec les porte-monnaie, les albums photographiques ou les livres. Et il en ira peut-être bientôt ainsi avec notre frigidaire, qui se souviendra de ce qu'il contient, avec nos papiers d'identité qui stockeront toute notre histoire, avec nos stylos qui enregistreront nos écrits, avec nos habits équipés de puces et d'antennes RFID, afin que les plus désordonnés s'y retrouvent dans leur garde-robe, etc.

Pour faciliter les échanges quotidiens avec tous les automates qui peuplent de plus en plus nos vies, et avec les robots domestiques qui se mettent à notre service, il faut trouver un langage commun. Or, dans le temps de l'action, le langage des chiffres demeure trop abscons pour être accessible. De même, le langage naturel se révèle souvent difficile à interpréter, d'autant plus que la compréhension tient en général plus à l'information implicite et à la culture partagée entre l'émetteur et le récepteur qu'au message lui-même. En revanche, le langage des émotions apparaît beaucoup plus immédiat et plus universel. Ainsi, le froncement des sourcils, la forme des yeux ou de la bouche suscitent en nous une émotion. Nous comprenons immédiatement l'inquiétude, la déception, la satisfaction, l'interrogation, etc. Des études de psychologie ont montré l'efficacité de ce mode de communication. **L'informatique affective en tire parti pour construire des interfaces homme-machine ou, plus généralement, des robots qui échangent avec les hommes au moyen du langage des émotions. Ces dernières sont suscitées en nous par les couleurs, les sons ou les expressions imprimées sur les visages des êtres virtuels avec lesquels nous sommes censés échanger.**

Or, les mécanismes par l'entremise desquels nous projetons nos facultés sur les machines font intervenir nos sentiments. Si l'on savait ce qui suscite en nous l'impression qu'un homme ou qu'un animal éprouve des émotions ou pense, nous serions capables de reproduire ces manifestations sur une machine qui susciterait, à son tour, les mêmes émotions et les mêmes croyances chez nous. C'est sur ce principe que fonctionnent la plupart des logiciels d'intelligence artificielle. Cela a conduit, ces dernières années, à porter une attention de plus en plus grande aux manifestations des émotions, au point qu'il existe désormais un domaine spécialisé, l'informatique affective – *affective computing* en anglais – à la frontière de l'intelligence artificielle et des sciences cognitives, qui aborde ces questions.

En conclusion, résumons ce qui vient d'être dit : certes, il se peut que les machines n'éprouvent pas d'émotions et n'aient pas de conscience ; nous n'en savons rien. Mais elles n'en ont pas besoin pour devenir intelligentes, au sens qu'on attribue à ce mot en intelligence artificielle. En revanche, il faut comprendre les mécanismes émotionnels humains pour que l'on soit en mesure d'en doter les machines et qu'on leur attribue de l'intelligence.

« Les machines n'ont pas d'intuition. »

Puissance de calcul limitée, faiblesse psychologique… Mais alors, que reste-t-il à l'homme ? L'intuition, sans doute, dernier bastion de l'intelligence, qui se nourrit de l'expérience sensible et de la connaissance de l'autre.

« Quand la machine décide. Comment Deep Blue a battu Kasparov ? », Decisio Info, 27 mai 2004

Descartes donne une définition de l'intuition dans les *Règles pour la direction de l'esprit* : « Par intuition j'entends non le témoignage variable des sens, ni le jugement trompeur de l'imagination naturellement désordonnée, mais la conception d'un esprit attentif, si distincte et si claire qu'il ne lui reste aucun doute sur ce qu'il comprend » (René Descartes, *Règles pour la direction de l'esprit, Œuvres de Descartes*, Levrault, 1826, règle troisième, tome XI).

Selon cette définition très classique, l'intuition désigne une connaissance immédiate et évidente qui ne recourt ni à la perception, ni à l'imagination, ni même au raisonnement, c'est-à-dire à l'analyse. Les systèmes de traitement de l'information ne sauraient donc posséder d'intuition puisque, en dehors d'opérations déterministes sur des états binaires de composants électroniques, ils ne font rien et n'accèdent à rien, ou, tout au moins, à aucune connaissance immédiate… En vérité, ils ne procèdent qu'à des opérations mécaniques sur des dispositifs matériels, autrement dit à des calculs. Leur existence ne se définit que par des calculs ; leurs connaissances, pour autant qu'on puisse leur en attribuer, se

réduisent aux seules transformations des informations qu'ils stockent dans leurs mémoires.

Néanmoins, pour programmer effectivement des ordinateurs qui soient en mesure de simuler des raisonnements complexes, l'intelligence artificielle conçoit des mécanismes d'auto-évaluation qui dressent un bilan approché de l'état où se trouvent les ordinateurs, et qui indiquent les actions les plus appropriées à la résolution des buts qui leurs ont été fixés. Ces mécanismes recouvrent, à certains égards, ce que l'on entend par intuition, parce qu'ils fournissent des indications superficielles qui orientent les machines en leur attribuant l'équivalent de préférences. On les qualifie d'heuristiques.

Le terme apparaît un peu barbare. Mais si l'on songe à la légende d'Archimède dans sa baignoire lorsqu'il découvrit le fameux principe et qu'il s'exclama : « Eurêka ! », le sens en vient naturellement : du grec *heuristikê*, art de trouver, les heuristiques désignent les techniques d'aide à la découverte. Selon les disciplines, leur rôle varie. Pour le philosophe, le philologue ou l'historien, elles se réfèrent à des hypothèses émises à titre provisoire, en vue d'aider à la découverte de faits saillants qui valident ou invalident ces hypothèses. Pour le pédagogue, une méthode est qualifiée d'heuristique lorsqu'elle fait découvrir aux élèves ce qu'on souhaite leur enseigner. Pour le spécialiste d'intelligence artificielle, les heuristiques tentent d'abréger, par quelques suggestions judicieuses, la litanie fastidieuse des énumérations exhaustives, afin de couper court aux errements des machines, quitte parfois à couper trop court... Autrement dit, elles jouent pour la machine le même rôle que les intuitions pour nous : de même que quelques signes nous révèlent la vérité,

sans qu'il soit besoin de raisonner plus avant, les heuristiques fournissent à la machine une évaluation générale de la situation, qui guide ses actions ultérieures, sans avoir à envisager toutes les éventualités.

Les heuristiques jouent un rôle central en intelligence artificielle lorsque la formulation des problèmes demeure trop lâche pour qu'un algorithme, autrement dit une succession ordonnée d'instructions déterminées, trouve une solution en un temps raisonnable.

Démontrer des théorèmes de mathématiques, jouer aux échecs, planifier les actions d'un robot, traduire des textes et les comprendre… autant de tâches qu'aborde l'intelligence artificielle, mais qui ne sauraient être résolues de façon systématique à l'aide d'un algorithme simple. L'obstacle principal à une solution ne tient pas nécessairement à l'absence de formulation logique. Ainsi, une formalisation claire circonscrit parfaitement l'ensemble des théorèmes mathématiques démontrables à partir d'un système d'axiomes initial. De même, les parties d'échecs envisageables, en respectant les règles du jeu, sont toutes potentiellement contenues dans l'expression de ces règles. L'obstacle ne tient pas non plus aux capacités des machines actuelles. Il est plus fondamental ; il tient à la combinatoire, autrement dit, au nombre faramineux des possibles qu'il faudrait parcourir, si l'on devait procéder à une énumération exhaustive de tous les théorèmes ou de toutes les parties d'échecs ou *a fortiori* de toutes les parties de go dont nous avons vu que le nombre avoisine 10^{600}. Ce nombre défie toutes les volontés… et toutes les mémoires, fussent-elles de machines !

Imaginons qu'un chef d'État pris d'une folie maniaque décide de faire un inventaire parfait des ressources minières

et des trésors enfouis dans le sous-sol de son pays. N'ayant aucune imagination, supposons qu'il engage une armée de sondeurs chargés de passer au crible des foreuses, mètre par mètre, toute la surface du territoire, sans rien omettre. La méthode, imparable au demeurant, se révélerait fort coûteuse : couvrir le pays, s'il est grand comme la France, exigerait cinq cents milliards de forages... Et ce chiffre est dérisoire au regard du nombre de positions d'échecs, 10^{50}, qui correspond à peu près au nombre d'atomes de la Terre, nombre lui-même dérisoire face au nombre de positions de go accessibles, 10^{120} qui est très largement supérieur au nombre de particules dans l'Univers observable, environ 10^{85} !

On conçoit dès lors l'impuissance à laquelle condamne toute exploration exhaustive sans le recours à des heuristiques... Or les parties d'échecs ou de go ne sont jamais jouées au hasard ; les théorèmes ne sont pas engendrés aveuglément par des mathématiciens fous ; les livres ne sont pas écrits par des singes dactylographes frappant capricieusement sur des machines à écrire... Personne, sauf dans les romans métaphysiques, ne parcourt, sans but, l'océan des possibles ! Il faut absolument être guidé dans la traversée quotidienne de cet océan et porter des œillères, pour river son regard sur des objectifs fixés à l'avance, quitte à manquer parfois certaines opportunités. Bref, il faut disposer d'intuitions.

Usuellement, les prospecteurs de minerais ne procèdent à des forages que là où c'est utile et, pour le savoir, ils s'aident de raisonnements géologiques afin de reconstituer les processus de formation des roches. De même, les archéologues connaissent les us et coutumes de nos ancêtres, ce

qui oriente leurs investigations ; et les chercheurs de trésors savent qu'habituellement ceux-ci demeurent accessibles et à l'abri des intempéries, car ils n'ont été cachés que pour être préservés et, un jour, retrouvés... Les spécialistes d'intelligence artificielle se sont inspirés de quelques règles de bon sens, tirées de l'expérience d'hommes de métier, pour forger l'équivalent des intuitions et les introduire dans les ordinateurs, afin d'orienter leurs explorations sur les voies les plus prometteuses. L'utilisation de ces ersatz d'intuitions, qu'ils qualifient d'heuristiques, s'est progressivement imposée comme nécessaire à tous ceux qui voulaient étendre l'emploi des ordinateurs à la simulation des activités intellectuelles. C'est cet emploi qui est à l'origine de l'intelligence artificielle ; c'est lui qui la caractérise, en tant que champ autonome de l'informatique, au point qu'elle fut parfois décrite comme une programmation heuristique. À cet égard, notons que l'exploitation systématique des heuristiques, en référence au savoir des hommes de métier, a donné naissance à la notion de système expert* et à l'ingénierie des connaissances* qui en est le prolongement naturel.

En somme, si les machines ne disposent pas d'intuitions au sens propre, cette absence fut ressentie comme problématique dès les premières tentatives de simulation du raisonnement. L'intelligence artificielle dota alors les machines d'un palliatif de l'intuition, qu'elle range sous le vocable d'heuristiques et qui caractérisent son approche. Bref, l'intelligence artificielle n'ignore pas le rôle de l'intuition ; loin de là, elle place l'intuition et ses simulacres au cœur de ses préoccupations.

« Avec l'intelligence artificielle émotive, nous confierons bientôt les personnes âgées aux robots. »

Les robots qui arrivent sur le marché pourront non seulement nous aider dans notre vie quotidienne, mais aussi nous faire la conversation en comprenant nos émotions et nos intentions, et nous répondre avec des intonations et des mimiques adaptées. Et très vite, à force de les fréquenter, nous risquons de penser qu'ils sont des compagnons bien plus agréables et faciles à vivre que les humains…

Serge Tisseron, « Les robots empathiques »,
Culture Mobiles, octobre 2015

Les robots dits de compagnie assurent une présence, qui peut être distante avec une personne connectée à l'autre bout de la ligne, mais aussi virtuelle avec ce que l'on appelle les *chatbots*, mot-valise formé sur la contraction de « chatter » et de « bots », littéralement les robots bavards qui dialoguent, échangent et répondent avec pertinence et patience à un certain nombre de questions. Ces robots peuvent prendre différentes apparences : ils affichent l'image d'une personne de l'entourage sur un écran, ou une image d'êtres virtuels qui ont éventuellement la forme d'un avatar affectif comme un petit animal. Le principe sur lequel repose leur conception demeure toujours le même : on suscite, chez les personnes, des émotions en mimant, sur le robot, quelques traits associés à ces émotions, par exemple un mouvement de lèvres, le clignement des yeux, un haussement d'épaule, etc.

Un robot-peluche en forme de petit phoque, Paro, qui bouge la queue quand on le caresse a été développé il y a quelques années. Dans de nombreux pays, ce type de robot commence à être utilisé pour assister les seniors. Au Japon notamment, où il n'y a pas beaucoup de maisons de retraite, les personnes âgées se rendent en maison de journée, comme dans une maison de quartier, pour y pratiquer des activités en groupe. Elles sont ensuite ramenées chez elles le soir. Dans le meilleur des cas, un proche ou un voisin dort avec elles ; sinon, elles restent seules, accompagnées d'un petit robot…

Ces petits robots ont aussi été introduits dans des unités de soin, en France, à Grenoble, au gérontopôle, à savoir au centre de gérontologie consacré, comme son nom l'indique, à l'étude des pathologies du vieillissement humain. Ces faux animaux affectifs peuvent vraiment présenter un intérêt pour aider les patients atteints de démence sénile. Ils constituent une présence affective rassurante pour le patient. Quelques travaux ont montré qu'il s'agit quasiment de la seule intervention non médicamenteuse qui donne des résultats tangibles dans la maladie d'Alzheimer en diminuant les angoisses. Dans une étude menée en Nouvelle-Zélande par l'équipe du professeur Haylay Robinson, la comparaison entre un vrai chien et le robot-phoque Paro a montré que les deux interventions étaient bénéfiques pour les quarante résidents et que ceux-ci touchaient et parlaient plus au robot qu'au chien. De même on a constaté qu'un plus grand nombre de patients engageait des conversations à propos du robot qu'à propos du chien.

Comme le note le paléoanthropologue Pascal Picq, spécialiste de l'évolution de l'homme et des grands singes, en France, la diffusion des robots de compagnie se heurte à

des barrières d'ordre culturel que l'on ne retrouve pas dans d'autres pays, en particulier au Japon : le fait que l'animal soit virtuel et, surtout, le côté apparemment régressif de celui qui échange avec ce qui est considéré comme un « jouet » conduisent à déconsidérer ces utilisations thérapeutiques des robots de compagnie... Un psychanalyste, Serge Tisseron, évoque même dans son livre *Le Jour où mon robot m'aimera* une empathie qu'il qualifie d'artificielle parce qu'elle conduit à s'identifier à des robots. Il y voit à la fois un oxymore et un risque majeur de déshumanisation.

Pourtant, il n'y a rien d'absurde à cela. Ces dispositifs assurent une présence ; comme ils se déplacent, ils peuvent suivre la personne et répondre à ses questions, lorsqu'elles sont posées de vive voix ; ils permettent aussi aux proches de rester en contact à distance. C'est en tout cas l'un des arguments utilisé par les développeurs de robotique d'accompagnement. Rappelons que les patients atteints de démence sénile posent de façon répétitive dix fois, cinquante fois, cent fois la même question à leur entourage qu'ils exaspèrent. Un robot de compagnie doté d'un agent conversationnel répond, avec patience, au rabâchage, laissant ainsi le personnel soignant et la famille plus disponibles lorsque cela s'avère nécessaire.

Zora, un petit robot de forme humanoïde (modèle NAO) haut de 58 cm a été introduit en mai 2015 dans une maison de retraite d'Issy-les-Moulineaux. Ce petit robot peut assister le personnel lors des animations et ateliers thérapeutiques : exercices de kiné et tai-chi, lecture du journal, jeu du « Qui suis-je ? »... Selon la directrice de ce centre, il s'agit d'une façon supplémentaire de stimuler les résidents, « au même titre que l'art-thérapie ou la zoothérapie ».

Cependant, il est bien évident que le robot ne remplacera jamais le personnel soignant et les animateurs. Faut-il insister : ils ne prennent sens qu'au regard d'une organisation sociale au sein de laquelle ils s'insèrent. En cela on parle souvent de dispositifs « sociotechniques ». Ainsi, les robots de compagnie ne se substituent pas aux hommes, du moins ils ne le devraient pas, ils ne les dispensent pas non plus d'accomplir leurs devoirs d'assistance à leurs semblables, loin de là, mais ils peuvent les aider à assumer leurs multiples responsabilités.

Notons enfin qu'il ne s'agit pas nécessairement de robotique androïde. Certains pensent même qu'il faudrait éviter des robots qui adoptent des rôles humains dont l'effet sur les personnes fragilisées qu'ils doivent aider paraît incertain. C'est la raison pour laquelle on songe plus souvent à des animaux, comme des chiens ou des phoques en peluche.

« Les voitures autonomes sont programmées pour tuer leurs passagers. »

Imaginez que votre propre voiture autonome décide qu'un mort est préférable à deux – et que ce mort, c'est vous ?

Sarah Kaplan, *Washington Post*, 28 octobre 2015

Nous nous étions accoutumés sans trop de difficulté à l'idée de trains sans conducteur. Ils circulent désormais très régulièrement sur des lignes de métro sans que l'on y fasse attention. Dans les airs, les pilotes automatiques des avions prennent le relais des hommes depuis bien longtemps ; et il en va de même en mer, sur les cargos de tous ordres, pétroliers ou porte-containers, et aussi sur les voiliers. Les drones, qu'ils soient militaires ou civils, sous-marins ou aériens, sont à même de se repérer et de se diriger sans le concours des hommes. Là encore, personne n'y voit plus rien de surprenant. En revanche, jusqu'à peu, l'idée que des voitures se conduiraient toutes seules paraissait un tantinet incongrue, peut-être parce qu'il y a des arbres sur le bord des routes et que des piétons les traversent de façon intempestive, peut-être aussi du fait de la circulation un peu chaotique… Bref, cela relevait de la science fiction, mais pas de la réalité. Or désormais, grâce au progrès de l'intelligence artificielle, le scénario apparaît réaliste. D'ailleurs, dès aujourd'hui, des voitures du commerce font des créneaux pour se garer toutes seules ; des régulateurs de vitesse assurent que l'on ne dépasse pas les limites prescrites ; et il existe même des

automobiles qui se dirigent seules, à condition que l'on conserve les mains sur le volant. La société Google annonce qu'elle a fait rouler des véhicules sans conducteur sur des millions de kilomètres de route dans des zones quasi-désertiques. On annonce pour demain des voitures qui conduiront de façon totalement autonome dans des conditions particulières, par exemple sur les autoroutes, où il n'y a pas de piétons, ou encore lorsque la circulation est dense, et que l'on roule pare-chocs contre pare-chocs. On dit aussi que les enfants nés en 2016 et plus tard ne passeront jamais leur permis de conduire, car ils n'en auront plus besoin. La société Uber parle même de bientôt supprimer les chauffeurs dans ses véhicules.

Or, pour réaliser des voitures autonomes, il faut faire appel à beaucoup de techniques d'intelligence artificielle, d'abord de l'apprentissage machine pour traiter les informations que fournissent des capteurs de toutes sortes, en particulier des caméras, afin identifier la route, sa texture, les arbres, les obstacles, les autres véhicules, les piétons, les feux tricolores, les panneaux de circulation et le reste du décor. Et, une fois l'environnement reconnu, il faut encore de l'intelligence artificielle pour prendre en un clin d'œil la décision idoine : freiner ou accélérer, tourner à gauche ou à droite, etc. Ce sont donc les succès de l'intelligence artificielle qui rendent tangibles les projets de voitures autonomes. Et, ceux-ci suscitent de nombreux espoirs. En effet, comme le montre l'analyse rétrospective des accidents de la route, ceux-ci tiennent surtout à des erreurs humaines provoquées par l'inattention due à la simple distraction, par l'exaltation de la vitesse qui va jusqu'à la perte de contrôle, ou encore par l'ébriété. Or, comme on attend d'un agent

artificiel qui prend les rennes de la voiture qu'il décide de façon parfaitement rationnelle en quelques microsecondes sur la base de renseignements fiables, on suppute qu'il en résultera une diminution conséquente du nombre d'accidents et par là du nombre de victimes. Comment ne pas s'en réjouir ? Cependant, en même temps que certains se satisfont de cette perspective, d'autres s'inquiètent des choix douloureux que nous aurons à faire lorsque nous programmerons ces machines et des conséquences que cela aura.

En effet, supposons qu'une voiture roule et que soudain cinq personnes traversent juste devant elle au feu vert, sans qu'elle ait le temps de freiner. Elle peut alors soit aller tout droit et tuer les cinq personnes, soit tourner sur la gauche, heurter le couple de piétons qui marchent innocemment sur le trottoir et les condamner tous les deux à une mort certaine, soit encore, dans un geste héroïque, frapper le feu tricolore ce qui détruira la voiture et sacrifiera son passager. Le conducteur humain n'aura certainement pas le temps d'évaluer les conséquences de ses actions et tuera les cinq personnes qui traversent imprudemment devant lui. Une machine programmée pour faire des choix rationnels en se fondant sur un critère purement utilitariste, décidera de limiter le nombre de morts à un seul, le conducteur. Cela veut dire que, lorsque vous achèterez une voiture, elle aura été froidement programmée pour vous tuer si cela peut limiter le nombre de victimes d'un accident, sans prendre plus d'égards avec vous, son propriétaire, qu'avec les autres.

Est-ce là une attitude vraiment éthique à laquelle tous doivent adhérer, sans aucune contestation possible ? Cette question a fait couler beaucoup d'encre dans les journaux et chez les philosophes depuis 2015. Pour l'examiner revenons

sur les conséquences des choix : (a) aller tout droit et tuer les cinq personnes qui traversent au feu vert (b) renverser et massacrer les deux piétons innocents (c) sacrifier le conducteur. Du point de vue strictement comptable si l'on vise à limiter le nombre de victimes et donc à optimiser le bien-être de l'humanité, le choix (c) semble le plus approprié. Si maintenant le véhicule protège son maître, en adoptant l'équivalent des lois d'Asimov, il exécutera l'action (b) ce qui conduira au massacre du couple qui marche sur le trottoir. En procédant ainsi, on ne prend à aucun moment en compte l'inconscience du groupe de cinq personnes qui violent la loi en traversant au feu vert, ce qui pourrait, aux yeux de certains spécialistes d'éthique, infléchir la nature de la décision et amener au choix (a). Notons que, dans l'histoire, se produisirent des situations dramatiques bien plus traumatisantes encore, mais qui, au plan formel, apparaissent analogues. Ce fût par exemple le cas des *Judenrat* (conseils juifs) mis en place par les nazis dans les ghettos d'Europe de l'Est pour aider à la déportation des populations. Au plan purement comptable, les nazis essayaient de convaincre que cela limitait le nombre de victimes. Les choix des *Judenrat* paraissaient alors similaires à ceux de la voiture autonome : certains cédèrent aux demandes des nazis en adoptant une motivation utilitariste comparable à celle qui conduit à l'adoption de la stratégie (c), quelques uns en profitèrent pour infléchir les décisions à leur profit ou à celui de leurs proches, d'autres refusèrent tout compromis et se suicidèrent. On doit toutefois souligner que les choix qu'affronteront les programmeurs apparaissent bien loin de ceux des *Judenrat* car les voitures autonomes ne s'imposeront vraiment que si, grâce à leur mise en service, le nombre de victimes d'accidents de la route est susceptible de diminuer substantiellement.

« Le "*Big Data*" menace la démocratie. »

Vous le comprenez sûrement. À moins d'être un ermite vivant dans les montagnes toute votre vie, il y a des tonnes de données là-bas sur vous et il n'y a rien que vous puissiez faire pour les stopper ou les contrôler. Les masses de données proviennent d'un peu partout, et elles peuvent être manipulés par de bonnes (ou mauvaises) mains pour construire des connaissances qui font frissonner sur votre vie, vos motivations et vos habitudes.

« Big Data Or Big Brother? », *Forbes*, 3 mars 2015

Désormais, avec l'extension du courrier électronique, des paiements dématérialisés, des textos, des tweets, des posts Facebook, des profils sur les réseaux sociaux, de l'utilisation des moteurs de recherche, des Pass Navigo, des cartes Vitale etc. la plupart de nos activités quotidiennes, tant personnelles que professionnelles, que ce soient nos déplacements, nos propos, nos écrits, nos loisirs, nos résultats scolaires, nos peines de cœur, nos achats, nos investissements bancaires, nos maladies ou nos soins, laissent des traces qu'on enregistre et stocke quelque part, sans que l'on ne sache vraiment ni qui les enregistre, ni où on les stocke. Il s'ensuit que rien ne disparaît jamais totalement, ce qui fait qu'à la longue, ces informations s'accumulent et constituent un matelas très épais de données sur chacun d'entre nous qui nous suit partout, tout au long de notre vie. Et, même si nous ne nous en souvenons pas et si nous ne savons pas où retrouver ces marques qui témoignent de notre passé et nous trahissent à notre insu, d'autres sauront s'en charger pour nous le moment venu.

Ce sentiment désagréable d'intrusion et de danger nous pousse régulièrement à nous inquiéter, à frissonner et à nous exclamer avec fatalisme *Big Brother !* en référence au célèbre roman *1984* de George Orwell. Nous frémissons alors un instant en songeant à ce que le futur nous réserve, puis nous oublions rapidement, pris par d'autres occupations…

Que se cache-t-il derrière ces inquiétudes passagères ? Beaucoup d'entre nous craignent qu'un État totalitaire ne s'empare de ces données personnelles pour faire régner la terreur dans la population. On pense à des emprisonnements arbitraires et à des faux procès intentés par un pouvoir autoritaire qui ne supporterait pas la contestation, quelque forme qu'elle prenne, et qui fouillerait dans notre passé pour s'assurer de notre loyauté absolue. Sans doute, de tels régimes politiques dominèrent-ils de grands pays d'Europe et d'Asie au cours du XX^e^ siècle. Et, assurément, de tels gouvernements despotiques règnent encore dans des contrées éloignées et coupées du monde comme la Corée du Nord. On s'inquiète aussi, avec raison, de l'absence de liberté de la presse dans nombre de pays, comme la Russie ou la Chine, voire la Turquie. Pourtant, même si l'élection de certains hommes politiques laisse craindre des dérives, en particulier des attaques contre la liberté de la presse et l'indépendance de la justice, dans les nations démocratiques, il est difficile de se convaincre de la réalité de telles craintes, car il existe de nombreux contre-pouvoirs.

À cela s'ajoute la réalité technique de la surveillance électronique : il ne suffit pas d'accumuler des données sur les individus, il faut les filtrer et les interpréter pour leur donner sens et incriminer à bon escient. À défaut, on s'y noie ! Cela suppose la mise en œuvre de techniques d'intelligence

artificielle pour reconnaître les voix, l'origine des accents, les émotions, les paroles prononcées, les visages, les empreintes digitales, les intentions belliqueuses ou violentes, les réseaux de connivences, les conspirations, etc. Des progrès considérables ont été faits dans ces domaines afin d'analyser les communications téléphoniques, les enregistrements vidéos sur les caméras de surveillance, les requêtes sur les moteurs de recherche ou les parcours sur la toile. Or, une étude attentive montre que, si l'on doit redouter l'effet de ces technologies, on ne doit pas tant craindre la volonté de domination d'États autoritaires sur leurs propres ressortissants, que celle de grands groupes industriels qui possèdent les grandes masses de données et maîtrisent les technologies de leur traitement. En effet, aujourd'hui, dans les États de droit, les gouvernements démocratiques se soumettent au contrôle populaire qui les conduit bien souvent à renoncer à assumer leurs responsabilités tandis que, n'étant assujettis à aucun contrôle, car ils sont déterritorialisés, les acteurs de l'Internet dominent sans partage, n'ayant de comptes à rendre à personne d'autre qu'à leurs actionnaires.

À titre d'illustration, mentionnons la reconnaissance des visages. Indubitablement, elle joue un rôle central dans la sécurité, puisqu'avec elle, nous ne pouvons plus être anonyme lors de nos déplacements. Aujourd'hui, les sociétés privées disposent d'atouts majeurs avec les techniques actuelles qui font appel à de l'apprentissage profond *(Deep Learning)* sur d'immenses quantités de données. Ainsi, entraîné sur 200 millions d'images, le système FaceNet, développé par la société Google, obtient-il un taux de reconnaissance correct de 99,63 %. Or, dans les régimes démocratiques comme celui de la France et des autres États

européens, la puissance publique n'a pas le droit d'utiliser les images individuelles, et, quand bien même elle l'aurait, elle ne les possèderait pas. La supériorité des géants de l'Internet, en particulier des réseaux sociaux qui possèdent les photographies que nous leur laissons en gage, sur les États en matière de sécurité intérieure apparaît donc patente, au point que les États s'associent avec eux. Ainsi, depuis début 2017, pour entrer aux États-Unis, les étrangers doivent-ils déclarer, dans le formulaire ESTA, les réseaux sociaux sur lesquels ils sont actifs ainsi que les identifiants qu'ils utilisent pour accéder à leurs comptes. Dans un registre d'idées analogue, en France, la direction générale de la sécurité intérieure (DGSI) a signé un contrat avec la société américaine Palantir pour traiter les données nationales françaises, car, selon les dires du directeur de la DGSI, seule une entreprise privée financée par la CIA au États-Unis maîtriserait suffisamment les techniques de traitement de masses de données. Cela signifie que désormais, les États recourent aux grands acteurs du numérique pour assurer la sécurité intérieure.

Symétriquement, ces acteurs défient les États lorsqu'ils utilisent des techniques de cryptographie pour permettre à des individus de transmettre des messages à l'insu des autorités. On se souvient de la polémique qui opposa début 2016 le gouvernement fédéral américain à la société Apple qui se refusait à prêter concours au FBI pour décrypter le contenu des téléphones des auteurs de la tuerie de San Bernardino qui avait eu lieu le 2 décembre 2015. Le cas de l'application Telegram, qui permet aux terroristes et aux hommes politiques de communiquer sans que l'on puisse accéder au contenu de leur message, est lui aussi illustratif de cette opposition.

Souvenons nous du rôle que cela eut dans la mobilisation des djihadistes qui, durant l'été 2016, ont investi l'église de Saint-Étienne-du-Rouvray, proche de Rouen, et ont assassiné le prêtre Jacques Hamel qui s'y trouvait.

Face aux acteurs privés qui le narguent et veulent se substituer à lui, l'État démocratique assure de plus en plus difficilement la sécurité intérieure des nations à l'heure des *Big Data*. Comment dès lors l'assimiler à un *Big Brother*, c'est-à-dire à un monstre froid qui dévorerait ses enfants comme le Léviathan de Hobbes ? Bien au contraire, aujourd'hui, il se montre de plus en plus impuissant et, par là, de moins en moins capable d'assurer la sécurité des citoyens, tandis que les géants de l'Internet prennent le relais. Tout en affichant les meilleures intentions du monde, ces derniers aspirent à tout dominer à la place des États. Cette tâche leur apparaît d'autant plus aisée qu'ils n'ont besoin d'aucune caution démocratique, d'aucun vote, d'aucune légitimité autre que celle que leur offre la maîtrise des technologies qui régissent de nouvelles régions du cyberespace. Bref, avec les *Big Data,* nous n'avons pas à craindre qu'un *Big Brother* unique s'impose à tous, mais bien plutôt que les *Gentils Organisateurs* (GO) du monde moderne se substituent aux États et aux institutions républicaines pour vider l'idée démocratique de souveraineté populaire de toute substance.

« Les "robots tueurs" remplaceront bientôt les soldats. »

Les armes autonomes choisissent leurs cibles et engagent le feu sans intervention humaine. Cela inclut, par exemple, des hélicoptères quadri-rotors armés capables de rechercher et d'éliminer des personnes répondant à certains critères prédéfinis, ce qui les distingue des missiles de croisière ou des drones pilotés à distance pour lesquels les humains prennent toutes les décisions de ciblage. Les technologies de l'intelligence artificielle ont atteint un stade où le déploiement de tels systèmes sera pratiquement – sinon légalement – réalisable dans les années, et non les décennies, qui viennent, et les enjeux en sont majeurs : les armes autonomes ont été décrites comme la troisième révolution dans la guerre, après la poudre à canon et les armes nucléaire.

Institut du futur de la vie, « Autonomous Weapons: an Open Letter from AI & Robotics Researchers », juillet 2015

En juillet 2015 une lettre ouverte publiée à l'occasion de l'ouverture de la conférence internationale d'intelligence artificielle, l'IJCAI (International Joint Conference on Artificial Intelligence), et signée par plus de 3 000 spécialistes d'intelligence artificielle et de robotique ainsi que par plus de 17 000 citoyens, dont des personnalités prestigieuses comme l'astrophysicien Stephen Hawking, l'homme d'affaire Elon Musk, le fondateur de la société Apple Steve Wozniak, le prix Nobel de physique Frank Wilczek, le philosophe Daniel Dennett ou le linguiste Noam Chomsky, manifestait l'inquiétude des plus grands scientifiques contemporains face aux applications potentielles de l'intelligence

artificielle dans le domaine militaire. Selon eux, des armes autonomes vont bientôt voir le jour. Cela signifie que des dispositifs seront capables de sélectionner d'eux-mêmes leurs cibles et d'engager le tir, sans que n'intervienne plus aucun humain dans la chaîne de décision. Toujours selon les auteurs de cette lettre ouverte, nous aurions affaire à une révolution dans l'art de la guerre qui ne trouve de précédents qu'avec l'utilisation de la poudre à canon et de la bombe atomique dans les conflits armés. Comment ne pas frissonner à l'idée de nous retrouver à la merci d'une machine qui décidera d'elle-même de nous éliminer, sans nous laisser aucun échappatoire. Avec ces robots, la guerre deviendrait proprement inhumaine, puisque les humains d'une des armées n'y participeraient plus directement et en conséquence ne risqueraient plus leur vie, tandis que les autres se retrouveraient à la merci de machines impitoyables. Certains, comme le roboticien Ronald Arkin, trouvent pourtant qu'il y aurait plus d'éthique – et donc d'humanité – dans l'inhumanité des robots que chez les humains, pour autant qu'on les astreigne à obéir aux lois de la guerre juste, car ils ne perdraient jamais leur sang-froid, ne se mettraient pas en colère, ne seraient pas soumis aux passions et demeureraient rationnels en toutes circonstances. Cependant, quand bien même on nous prouverait que notre crainte des robots soldats s'avère irrationnelle, ceux-ci continueraient toujours à nous glacer le sang ! Ajoutons à cela qu'indépendamment de nos réactions émotives spontanées, le point de vue de Ronald Arkin selon qui ces robots-soldats seraient plus éthiques que les soldats humains se discute (voir encadré) avec des arguments parfaitement rationnels.

Mais, que signifie au juste cette notion d'arme autonome qui donne froid dans le dos, et en quoi l'intelligence artifi-

Les robots guerriers seront plus éthiques que les soldats humains

Une polémique sur la vertu morale des robots guerriers enfle depuis quelques années aux États-Unis et dans les pays anglo-saxons. Elle a été lancée par un roboticien américain, Ronald Arkin, selon qui, des robots qui obéiraient aux règles d'engagement de la guerre juste, se conduiraient mieux, au plan moral, que des soldats humains sur le champ de bataille, car ils ne perdraient jamais leur sang-froid et ne sauraient être mus par un esprit de vengeance, comme le sont si souvent les femmes et les hommes, car les robots n'éprouvent pas d'émotions. Rappelons que ces prises de position se firent entre 2004 et 2009, à l'époque où le monde eut vent, par WikiLeaks, de la conduite fort condamnable des militaires américains lors de la seconde guerre d'Irak. On conçoit que, dans ce contexte, certains aient pu désespérer des vertus morales humaines, au point de faire plus confiance à des robots. Et, il est vrai que les robots obéissent strictement aux instructions. À supposer qu'on intègre, dans leur programmation, des règles de conduite conformes aux exigences morales humaines, ils s'y soumettront nécessairement et se comporteront donc correctement, sans commettre d'exactions ni de crimes inutiles. En d'autres termes, ils ne tueront qu'à bon escient, lorsque les lois de la guerre juste les y autoriseront.

« Tuer à bon escient » au nom des lois de la « guerre juste » : on ne peut s'empêcher de frémir à l'écoute de cette proposition. Cela amène à se demander ce que l'on entend par là et, plus généralement, ce que sont les règles de la « guerre juste ». Rappelons que cela fait très longtemps que l'on tente de codifier la guerre et que cela a conduit à distinguer trois types de droits : le *jus ad bellum* (droit avant le conflit), qui définit les conditions de déclaration d'hostilités et les règles d'engagement, le *jus in bello*, qui précise ce qui est licite au combat, et le *jus post bellum* (droit après la guerre) qui prend place après les conflits, en précisant les règles d'indemnisation des victimes, les réparations de guerres, etc. Bien évidemment, aujourd'hui, avec les robots soldats, il est essentiellement question du droit au combat, c'est-à-dire du *jus in bello*, et partiellement du *jus ad bellum* lorsqu'il précise la nature de l'engagement. D'après la convention de Genève de 1949, le *jus in bello* stipule que l'on doit discriminer entre militaires

et civils, pour épargner ces derniers. Or aujourd'hui, il est quasiment impossible à une machine d'opérer la distinction entre civils et militaires, surtout dans les guerres asymétriques où les combattants ne portent plus d'uniforme. À cela on doit ajouter que ce principe dit de discrimination souffre deux exceptions. En effet, lorsqu'un civil participe au combat, il devient alors licite de l'attaquer ; symétriquement, lorsqu'un militaire est neutralisé, on n'a pas le droit de s'en prendre à lui. Bref, même s'il s'exprime sous forme d'une règle de conduite apparemment simple, le principe de discrimination apparaît si complexe dans sa mise en œuvre effective qu'il ne se réduit pas à un algorithme, loin s'en faut.

Quand aux règles d'engagement, elles recourent à un principe dit de proportionnalité selon lequel la riposte doit être en rapport avec l'attaque. Or, cette question relève du jugement et ne se laisse pas plus réduire à un algorithme que le principe de discrimination.

En conclusion, à l'analyse, l'idée que les robots soldats pourraient se conformer aux règles de la guerre juste et, en conséquence, se comporter de façon plus « éthique » que les femmes et les hommes au combat paraît fort douteuse.

cielle contribue-t-elle à son perfectionnement ? Les armées disposent depuis longtemps de dispositifs qualifiés d'autonomes parce qu'ils se pilotent eux-mêmes et atteignent des objectifs qui leur ont été fixés à l'avance. À titre d'illustration, une fois déclenché, un missile à guidage LASER « accroche » sa cible jusqu'à ce qu'il parvienne à l'« avaler ». Les drones eux aussi procèdent de la sorte : ils sont téléguidés à plus de 3 000 km du théâtre d'opérations par des pilotes qui déclenchent des tirs de missiles sur des objectifs précis. Un fois lancés, ces missiles prennent en chasse leurs cibles avec des systèmes de guidage jusqu'à les atteindre. Avec les technologies contemporaines, nous visons donc plus loin et avec plus de précision. Cependant, si l'innovation ne consistait qu'à ça, nous n'aurions là qu'un accroissement

dans la distance de tir, sans rupture majeure dans l'art de la guerre : toutes proportions gardées, cette rupture s'apparenterait à celle qu'apporta l'arbalète par rapport à l'arc ou le canon par rapport au fusil.

Or, avec les engins autonomes qualifiées parfois de « robots tueurs » ou encore, en termes techniques, de SALA (« Systèmes d'armes létales autonomes »), il en va autrement qu'avec les armes traditionnelles, car ces systèmes autonomes choisissent leurs cibles, puis engagent le tir sans intervention humaine : c'est ce qui fait leur particularité. La nouveauté ne tient ici ni à la puissance de feu, comme c'est le cas avec la poudre à canon et, surtout, avec la bombe atomique, ni au rayon d'action, puisque la portée n'est pas plus grande que celle des drones autonomes, mais à leur nature « logique ». Tandis qu'avec l'arc, avec l'arbalète, avec le fusil, avec le largage de bombes et même avec les drones, le soldat choisissait sa cible en la visant à distance, avec les armes autonomes il la sélectionne en la « spécifiant », c'est-à-dire en déterminant, par avance, ses caractéristiques logiques de façon abstraite. Pour le dire de façon plus savante, jusqu'ici la modalité sémiotique de désignation de la cible était indexicale en cela que l'on pointait, éventuellement avec un viseur, avant d'engager le tir, alors qu'avec les armes autonomes, elle apparaîtra strictement formelle et mathématique. Ceci signifie que l'intelligence artificielle introduit une distance entre le soldat et sa cible qui n'est pas seulement d'ordre physique, mais aussi, et surtout, d'ordre logique, puisqu'il ne la voit plus, ne la désigne plus, mais se contente de la décrire…

Cependant, cette transformation majeure dans la pratique de la guerre demeure hypothétique, car il apparaît très

difficile de caractériser mathématiquement un ennemi. S'agit-il d'un homme en uniforme ou en treillis ? Pourtant, ce n'est, comme nous l'avons vu, presque plus jamais le cas avec les guerres dites asymétriques où les soldats se confondent de plus en plus avec la population. Faut-il tirer sur tout ce qui bouge ? Parle-t-on de barbus, de « bronzés » ou d'individus aux yeux sombres ? Cela prêterait à rire s'il n'y allait pas là de la vie de femmes et d'hommes. Bref, cette notion d'arme autonome qui paralyse au premier abord paraît vite dérisoire dès que l'on analyse, en pratique, l'utilisation que l'on est susceptible d'en faire dans le domaine militaire. On conçoit dès lors la réserve que manifestent les états-majors sur le sujet et le moratoire partiel que se sont imposées les armées américaines. Faudrait-il alors interdire à jamais les armes autonomes ?

Plusieurs organisations non gouvernementales dont Human Right Watch se sont emparées de cette question pour imposer l'interdiction des armes autonomes au même titre qu'il existe déjà une interdiction des armes chimiques ou biologiques. Cette prohibition porterait à la fois sur l'emploi de ces armes dans des conflits armés, sur leur stockage et sur les recherches les concernant. Or, si les grandes puissances ne souhaitent pas, pour l'instant, développer ces armes au plan opérationnel, elles ne souhaitent pas non plus s'empêcher d'en maîtriser les technologies, au cas où d'autres s'en empareraient et s'en serviraient. Il s'ensuit des débats entre les partisans d'une interdiction et les opposants à cette même interdiction. Depuis 2014, cela fit l'objet de plusieurs réunions internationales organisées sous les auspices de l'ONU à Genève. Cependant la législation sur lesdits « robots tueurs » ou SALA risque fort de

se heurter à une opposition frontale entre des organisations non gouvernementales qui voient dans leur interdiction un enjeu majeur qu'elles cherchent à faire triompher pour attester de leur influence, et des États industriels beaucoup plus réservés, car ils souhaitent préserver toutes les options pour le futur.

« Il faut donner des droits aux robots. »

Le parlement européen [...] demande à la Commission, lorsqu'elle procèdera à l'analyse d'impact de son futur instrument législatif, d'examiner les conséquences de toutes les solutions juridiques envisageables, telles que [...] la création d'une personnalité juridique spécifique aux robots, pour qu'au moins les robots autonomes les plus sophistiqués puissent être considérés comme des personnes électroniques dotées de droits et de devoirs bien précis, y compris celui de réparer ***tout dommage causé à un tiers ; serait considéré comme une personne électronique tout robot qui prend des décisions autonomes de manière intelligente ou qui interagit de manière indépendante avec des tiers.***

Projet de rapport concernant des règles de droit civil sur la robotique, rapporteure Mady Delvaux, Parlement européen, commission des affaires juridiques, 31 mai 2016

Le « Projet Grands Singes » (GAP – « The Great Ape Project ») vise à attribuer les droits fondamentaux de la personne aux primates supérieurs (chimpanzés, gorilles, orangs outangs, etc.) à raison de leurs capacités cognitives considérées aujourd'hui comme étant plus grandes que celles de certains humains, en particulier que celles de personnes handicapées ou atteintes de démences. Cela s'inscrit dans un grand mouvement d'égalité qui après avoir lutté contre l'esclavage, puis contre le racisme et le sexisme, pour émanciper d'abord tous les hommes, quelle que soit leur race, et ensuite toutes les femmes, vise maintenant tous les êtres vivants. Symétriquement, il existe aujourd'hui, un projet d'extension des droits de la personne aux robots autonomes, à raison, là encore, de leurs facultés à prendre des

décisions. À cet égard, une fois n'est pas coutume, l'Europe est en pointe suite à la sortie du rapport d'une élue européenne, Mady Delvaux, sur ce sujet et à l'adoption d'une résolution du parlement européen proposant d'attribuer des droits et des devoirs aux robots et, plus généralement une personnalité juridique à ce que le texte désigne comme des « personnes électroniques ». Ce brouillage des frontières de l'humanité, revendiqué au nom de principes généreux d'équité et de respect de tout et de tous, ne manque pas de déconcerter au point que l'on s'interroge sur les motivations de telles revendications.

Dans le cas de l'extension des droits de la personne aux animaux, il y a le souci de la souffrance, avec la volonté compréhensible et légitime, si ce n'est de la supprimer, tout au moins de la diminuer, en particulier dans les abattoirs, et d'éviter les expérimentations animales lorsqu'elles ne sont pas indispensables. À première vue, les motivations ultimes de l'attribution de droits aux robots autonomes et aux personnes électroniques apparaissent plus obscures, car, pour l'instant, nul n'évoque les sensations des robots, encore moins leur douleur.

L'argument déployé par les tenants de l'attribution d'une personnalité juridique aux robots, que ce soit des avocats comme maître Alain Bensoussan en France, ou des parlementaires comme Mady Delvaux en Belgique, repose sur l'imputation de la responsabilité dans le cas d'accidents où des systèmes robotisés autonomes seraient impliqués. Cela tient à l'extrême complexité des dispositifs techniques actuels qui fait que, dans le temps de l'action, on éprouve tous des difficultés à établir l'enchaînement exact des relations de cause à effet à la source des actions des robots, ce

qui nous déroute. À défaut, on projette sur eux une entité abstraite que l'on appelle un agent, et dont on suppose qu'elle les meut. C'est à cet être, tout à la fois virtuel et fugitif, que l'on attribue la source des actions du robot. Mais, peut-on imputer la responsabilité des actions du robot à cet agent ? Peut-on même parler là de responsabilité ? En effet, la notion de responsabilité morale des robots n'a pas de sens, puisque ceux-ci ne sont pas des êtres libres en ce qu'ils ne possèdent pas de volonté propre et qu'ils sont assujettis à des buts qu'un autre leur a fixés.

L'attribution du statut de personne aux robots autonomes n'a donc pas de signification morale et il n'a pas la prétention d'en avoir une. En effet, d'après ses promoteurs, cela correspond à ce que l'on appelle une « fiction juridique » : cette personnalité est équivalente à la personnalité morale d'une entreprise. Une telle personnalité permettrait d'indemniser les victimes d'accidents impliquant des robots. Sur un registre analogue, ajoutons que, tant le rapport susmentionné de Mady Delvaux, que le programme de Benoît Hamon, candidat à la présidence de la République française aux élections de 2017, proposent tous deux de faire payer une taxe aux robots, à raison des emplois qu'ils font disparaître. L'idée commune partagée à la fois par le projet d'attribution d'une personnalité juridique aux robots et par la proposition de leur taxation revient à offrir une compensation à ceux que les robots auraient lésés, à hauteur des préjudices subis. On en comprend aisément la logique. Toutefois, cela pose un problème majeur : les robots n'ayant ni conscience, ni possession, comment les sanctionner d'une façon satisfaisante pour les victimes de leurs méfaits ? Il ne saurait être question de destruction ou de peine de prison ! Par

analogie à la personnalité morale des sociétés, qui indemnise les créanciers ou les victimes avec les fonds propres de la société, on doterait les robots d'un fonds propre destiné à indemniser les victimes de leurs inconséquences. Autrement dit, les fabricants et/ou les utilisateurs assureraient les robots contre les dégâts qu'ils risqueraient d'occasionner.

Quoique séduisante au premier abord, cette attribution d'un droit moral aux robots pose un certain nombre de questions. Nous en énumérerons ici trois qui nous apparaissent dissuasives.

En tout premier lieu, l'obligation d'abonder un fonds d'assurance destiné à indemniser les victimes de robots risque de détourner les nouveaux arrivants dans l'industrie robotique, car ils n'auront pas les moyens de payer pour des dégâts virtuels, avant même de faire leurs preuves. Il en va de même pour les utilisateurs potentiels qu'une telle assurance découragera.

En deuxième lieu, l'attribution de la responsabilité au robot en cas d'accident éludera l'enquête. On paiera pour s'assurer de l'absence de recours des victimes. Mais, outre que cela ne règle pas le cas des dommages susceptibles d'une poursuite pénale, cela ne forcera pas non plus à déterminer les causes exactes et, par là, la vraie responsabilité qui incombe parfois au fabricant, pour malfaçon, parfois au programmeur, parfois à l'utilisateur, parfois au propriétaire ou encore à celui qui a entraîné le robot lors de sa phase d'apprentissage... Or, cette détermination est essentielle pour faire progresser la technologie et éviter que des accidents semblables ne se reproduisent.

En troisième lieu, l'idée de taxes sur les robots aurait des effets délétères sur l'économie, car désormais, loin de causer

le chômage, les robots industriels accroissent la productivité des usines et donc la compétitivité, ce qui évite la désindustrialisation et permet de maintenir des emplois…

Bref, si elle frappe l'imagination et si, à ce titre, elle recueille beaucoup d'échos dans les médias, l'attribution d'une personnalité juridique aux robots pose bien plus de problèmes qu'elle n'en résout !

L'AVENIR DE L'INTELLIGENCE ARTIFICIELLE

« Nous ne sommes pas prêts pour le tsunami technologique qui advient. »

Le tsunami technologique auquel nous assistons demande des réponses économiques, morales, éthiques… et rien n'est prêt. C'est un choc à la fois technologique et psychologique.

Laurent Alexandre cité par Céline Lanusse,
La Tribune, 28 mai 2015

Sans doute, le terme « tsunami » conserve-t-il des connotations emphatiques qui déroutent et semblent, de ce fait, peu appropriées pour évoquer les conséquences de l'évolution des technologies contemporaines. Pourtant, à le considérer de près, il s'avère plus pertinent qu'il n'y paraît au premier regard.

Rappelons que le mot « tsunami » vient du japonais pour désigner une vague d'origine sismique qui pénètre profondément dans les terres, submergeant tout sur son passage. Cela correspond à ce que l'on appelle, en français, un raz-de-marée. Au sens figuré, ce terme caractérise aussi un bouleversement au plan moral, social ou politique de la société. Ainsi, on parle d'un raz-de-marée électoral, par exemple d'un raz-de-marée conservateur.

Dans le cas de la technologie, nous avons affaire à des transformations brusques qui par leur ampleur modifient considérablement les modes de communication entre les hommes : désormais, nous échangeons quasi-instantanément et quasi-gratuitement lettres, photographies, vidéos sur

toute la surface de la planète. À ces échanges massifs d'information qui font que la vie sociale et l'activité politique se déroulent en grande partie sur la toile et sont donc numérisées, font échos des techniques de traitement de très grandes masses d'information *(Big Data)* de plus en plus efficaces qui permettent par exemple de reconnaître un visage entre des centaines de milliers, voire des millions, ou de transcrire la parole, ou encore d'établir automatiquement l'état de l'opinion sur un sujet.

Ces évolutions très rapides des technologies de l'information et de la communication entraînent dans leur mouvement une vague de transformations sociales de très grande ampleur, un peu comme un tremblement de terre sous-marin provoque un raz-de-marée qui balaie tout sur son passage dans les terres émergées qu'il engloutit.

Cela surprend, car l'information possède un caractère intrinsèquement immatériel qui lui donne, à première vue, une allure bien inoffensive. Or, l'expérience de ces dernières années montre les effets dévastateurs du développement massif des technologies de l'information tant sur l'économie que sur le tissu social et sur l'organisation politique.

Au plan économique, la concomitance du déploiement des réseaux de communication et du développement des transports de marchandises à bas coût autorise une externalisation du travail et une délocalisation des entreprises. Cela modifie l'organisation de la production de biens manufacturés qui se déporte vers l'Asie et induit, en retour, un fort taux de chômage dans les pays où le coût du travail était élevé comme les pays européens. Par contrecoup, des crises sociales et politiques d'ampleur inégalée se produisent aujourd'hui dans ces pays.

Au plan social, la captation quasi-gratuite d'images et de sons ainsi que leur libre circulation sur les réseaux modifient les rapports d'autorité : le privilège accordé au savoir vacille, car tous ont accès à l'information, ce qui fait illusion au point que l'on en est très souvent conduit à confondre information et savoir. Les statuts d'autorité du médecin, du professeur, du journaliste ou de l'homme politique s'effondrent, laissant la place à quelques figures médiatiques dominantes qui tirent leur pouvoir de leur réputation plus que de leurs compétences.

Au plan politique, enfin, les réseaux mondiaux de communication rendent les frontières territoriales poreuses : où que l'on se trouve sur la planète, sauf peut-être en Corée du Nord, on accède à l'information produite dans n'importe quel autre pays du monde. Il en résulte une dissociation entre l'État et le territoire qui conduit à une dissolution de l'idée de souveraineté sur laquelle se fondaient les États de droit depuis plusieurs siècles. À cela s'ajoutent de nouvelles vulnérabilités sur le cyberespace où des acteurs non-étatiques interviennent – et s'imposent même parfois – en conduisant des agressions virtuelles et des chantages en direction soit d'États, soit de compagnies privées, soit même de particuliers.

Sommes-nous prêts pour faire face à ces évolutions ? Malheureusement, beaucoup de nos contemporains se contentent de jeter un œil fiévreux dans le rétroviseur de l'Histoire, sans prendre conscience des défis majeurs qui se dressent devant eux. Ils s'inquiètent de ce qui ne présente plus de danger, comme de la prééminence de l'État et de son emprise sur les citoyens, sans comprendre que les

nouveaux enjeux, les nouvelles luttes, les nouvelles régions et les nouveaux pouvoirs sont ailleurs et que cette vague de changements risque de nous emporter tous, si nous n'y prenons pas garde, comme le fait un tsunami, alors qu'il suffirait simplement de s'élever de quelques mètres pour y faire face…

« L'intelligence artificielle n'a pas tenu ses promesses. »

Cela fait un demi-siècle, depuis que les ordinateurs sont apparus au monde, que l'on promet de bientôt les programmer pour les rendre intelligents et que l'on promet aussi, ou plutôt que l'on a peur qu'ils parviennent bientôt à nous assimiler nous-mêmes comme des ordinateurs. En 1947, Alan Turing prédisait qu'il existerait un ordinateur intelligent d'ici la fin du siècle. Maintenant que le millénaire est dépassé de trois ans, il est temps de faire une évaluation rétrospective des tentatives faites pour programmer des ordinateurs intelligents comme HAL dans le film 2001.

Hubert L. Dreyfus, Stuart E. Dreyfus, « From Socrates to Expert Systems: The Limits and Dangers of Calculative Rationality », *Philosophy and Technology*, 2004

On attend toujours l'intelligence artificielle au tournant. D'années en années, elle déçoit toujours plus les espérances. HAL, l'ordinateur intelligent du film de Stanley Kubrick *2001, l'Odyssée de l'espace*, ne voit pas encore le jour, même si le millénaire est déjà passé depuis longtemps. Traduction automatique, compréhension du langage naturel, reconnaissance de la parole et des visages, vision, démonstration de théorème, résolution de problèmes, robotique… l'histoire récente accumulerait les échecs. Rien de vraiment tangible n'adviendrait dans ce secteur de la technologie… Autant de lieux communs bien répandus, que l'on retrouve depuis longtemps et dont un philosophe américain, Hubert Dreyfus, spécialisé dans la critique de l'intelligence

artificielle, s'est fait le champion depuis plus de cinquante ans aujourd'hui.

En regard de ces constatations désabusés que démentent quotidiennement les réalisations technologiques, on fait régulièrement état des promesses vertigineuses faites par les pionniers de la discipline, il y a une soixantaine d'années, en particulier celles d'Alan Turing, de Marvin Minsky, d'Herbert Simon, d'Alan Newell et de tant d'autres, et on se gausse !

Or curieusement, lorsqu'on tente de décrédibiliser l'intelligence artificielle en invoquant ces prédictions excessives, on omet toujours de citer les déclarations exactes que firent les précurseurs mentionnés ci-dessus. Faisons-le ici, pour tenter d'instruire le procès.

Commençons par le pionnier, Alan Turing, l'homme qui anticipe l'intelligence artificielle avant qu'elle ne naisse vraiment, en 1950, dans un article célèbre intitulé « Machines à calculer et intelligence ». Dans ce fameux texte, il essaie de préciser ce qu'on pourrait appeler « penser » pour une machine. D'après lui, la pensée et plus généralement l'intelligence n'ont à faire ni avec l'apparence physique, ni avec le grain de la voix, ni même avec les traits du visage. Cela n'a pas non plus directement à voir avec la conscience. Une machine sera dite intelligente si ce qu'on observe de son comportement semble émaner d'un être intelligent. Et pour préciser ce qu'il entend exactement par intelligence des machines, Alan Turing imagine un subterfuge, ledit « jeu de l'imitation ». Nous ne nous étendrons pas ici sur la pertinence de ce test d'intelligence qui a été beaucoup commenté et qu'on appelle le test de Turing. En revanche, intéressons-nous aux prédictions d'Alan Turing lui-même : d'après lui, on devrait être en mesure, d'ici 50 ans (rappelons

que nous étions en 1950), de concevoir un ordinateur capable de tromper un interrogateur dans plus de 30 % des cas, sur un échange de cinq minutes. Malgré tous les déboires et toutes les déconvenues de l'intelligence artificielle, on est effectivement capable depuis quelques années de concevoir des machines (appelées parfois *chatbots* en anglais, littéralement, « robots bavards » en français, ou plus communément « agents conversationnels ») avec lesquelles tous peuvent échanger et discuter sur Internet. Or ces agents conversationnels qui jouent au jeu de l'imitation obtiennent des performances très semblables à celles prévues par Turing. On prétendit d'ailleurs, en 2014, qu'un tel agent, nommé Eugene Goostman et développé en 2001, parvint 13 ans plus tard, grâce aux algorithmes d'apprentissage dont il se compose, à tenir tête à des hommes dans les conditions prévues par Alan Turing.

Dans un ordre d'idées analogue, Herbert Simon, qui devint plus tard prix Nobel d'économie et reçut la médaille Turing, fit avec son collègue Alan Newell des déclarations fracassantes. Selon eux, (nous étions en 1958) :

– dans 10 ans, les ordinateurs, s'ils ne sont pas interdits de participation aux compétitions internationales, devront indubitablement devenir les champions du monde au jeu d'échecs ;

– dans 10 ans, un ordinateur sera certainement capable de composer de la musique douée d'une indéniable valeur esthétique ;

– dans 10 ans, les ordinateurs démontreront des théorèmes mathématiques totalement originaux ;

– dans 10 ans, les ordinateurs simuleront le psychisme au point que toutes les théories psychologiques prendront la forme de programmes d'ordinateurs, etc.

Il va sans dire que ces prédictions se sont toutes révélées erronées. Ainsi, en 1965, un enfant de 10 ans battit l'un des premiers programmes d'ordinateur jouant aux échecs. Pourtant, en 1997, c'est-à-dire 40 ans plus tard, un ordinateur parvint à défier et à vaincre le champion du monde en titre au jeu d'échecs. Et, depuis, ils l'emportent même au jeu de go réputé beaucoup plus complexe, et au poker, plus lucratif encore, à défaut d'être plus difficile... Les ordinateurs sont beaucoup utilisés par les musiciens et ils contribuent à la création artistique contemporaine. Ils prennent une part importante dans l'activité des mathématiciens, pour démontrer des théorèmes. De même, les psychologues ont beaucoup fait appel à des modèles informatiques. Cela veut dire que la plupart de ces annonces n'étaient pas totalement absurdes, même si elles ont été quelques peu démenties, car les délais ne furent pas respectés. Mais qu'est-ce qu'un retard de quarante ans au regard de l'histoire de l'humanité ?

Toujours dans le domaine de l'intelligence artificielle, on s'enthousiasma au début des années 1980 pour ce qu'on appelait les « systèmes experts », c'est-à-dire pour des logiciels qui comprenaient, en lieu et place des programmes informatiques traditionnels, du savoir spécialisé se référant à des connaissances détenues par des hommes de métier. Là encore, les anticipations furent déçues : le développement industriel des systèmes experts ou des systèmes à base de connaissances fut beaucoup plus lent que ce que tous les spécialistes de prospective avaient imaginé. Pourtant, aujourd'hui, ces techniques se diffusent dans l'industrie, où elles sont couramment utilisées. Qui plus est, les progrès d'Internet leur ont donné une actualité neuve : ces technologies contribuent à réaliser ce que les chefs des gouver-

nements européens avaient désigné, en 2000, à Lisbonne, comme une « société de la connaissance », dans laquelle la production de richesses tient bien plus à la création de savoir qu'à la fabrication d'objets manufacturés.

Enfin, aujourd'hui, avec l'apprentissage machine, en particulier avec l'apprentissage profond *(Deep Learning)* et les masses de données *(Big Data)*, les possibilités d'applications de l'intelligence artificielle paraissent innombrables…

Contrairement donc à une idée couramment répandue, l'intelligence artificielle a tenu beaucoup de ses promesses ; et elle va même bien plus loin. Ainsi, qui aurait imaginé, il y a à peine vingt ans, une encyclopédie aussi riche et facile d'accès que celle que nous procurent les moteurs de recherche sur la toile ? Ou une voiture autonome ? Ou encore les performances actuelles des programmes de reconnaissance des visages et de transcription de la parole ? Dès lors, pourquoi accuse-t-on l'intelligence artificielle ? Peut-être parce que la foi que certains mettent en elle va bien au-delà de tout ce qui est réalisable, puisqu'il s'agit pour eux de réifier la conscience (c'est-à-dire d'en faire une chose matérielle) sur un ordinateur et de parvenir à ce que certains appellent l'intelligence artificielle forte ou générale… Dans ces perspectives, l'intelligence artificielle risque fort de décevoir et, d'une certaine façon, heureusement ! Mais si on relit attentivement les vraies promesses des pionniers, on constate qu'en dépit de leur enthousiasme, elles demeurent très mesurées à cet égard.

« Les robots nous mettront tous au chômage. »

Ne vous plaignez pas que le progrès technique détruise des emplois, il est fait pour ça.

Alfred Sauvy, *Informatisation et emploi*, 1981

Au début du XX^e siècle, l'apparition de la riveteuse sur les chantiers navals mit au chômage des confréries entières d'ouvriers spécialisés dans la pose des rivets sur les coques de navires. Voici une des conséquences tragiques, parmi tant d'autres, de l'automatisation. En va-t-il de même avec les robots intelligents que l'ont conçoit aujourd'hui dans le secret des laboratoires ? Mettront-ils au chômage toutes les confréries de travailleurs ? L'étymologie même du mot « robot » le suggère. Rappelons, à cet égard, que le terme est apparu pour la première fois en 1921, dans une pièce de théâtre écrite par l'écrivain tchèque Karel Čapek et intitulée RUR – *Rossum's Universal Robot.* Ce mot, forgé à partir du radical tchèque *robota* qui signifie « travail », y désignait des « travailleurs artificiels ». On ne s'étonnera donc pas que ces petits automates, réalisés grâce aux progrès des techniques contemporaines, remplacent les travailleurs humains dans les tâches les plus fastidieuses de leurs activités. Certains, comme le roboticien Moshe Vardi, directeur de l'Institute for Information Technology à l'université Rice au Texas, vont jusqu'à affirmer qu'avec les progrès de l'intelligence artificielle, « nous approchons du moment où les machines

pourront surpasser les humains dans presque toutes les tâches » (*in* « Les robots intelligents arrivent, menaçant des millions d'emplois », *Le Parisien*, 13 février 2016), ce qui exige de s'y accoutumer et de trouver les moyens d'y faire face. La concomitance du perfectionnement des robots et de l'apparition d'un chômage de masse endémique depuis une quarantaine d'années dans nos sociétés occidentales développées, en particulier en France, semble accréditer cette idée.

De là à craindre que la généralisation de l'emploi des machines mette tout le monde au chômage en raréfiant le travail dans tous les secteurs d'activité, il n'y a qu'un pas que beaucoup franchissent allègrement que ce soit dans notre pays ou dans d'autres. Ainsi, toujours selon Moshe Vardi, les robots nous voleront nos emplois ce qui fait qu'il y aura plus de 50 % de la population active au chômage dans les 30 prochaines années (voir Anne Dolhein, « Les robots auront capté la "plupart" des emplois d'ici à 2045 selon Moshe Vardi : vers 50 % de chômage et des loisirs infinis ? », Reinformation.tv, 15 février 2016). Loin d'être neuve, cette même idée justifia, en son temps, la réduction du temps de travail en France et l'apparition desdites 35 heures qui limitent la durée de travail hebdomadaire. Elle continue de faire son chemin, au moins en France, chez certaines personnalités politiques qui voudraient réduire le travail hebdomadaire à 32 heures, tout en instituant un revenu universel permettant à tous de survivre sans travailler.

Or le constat ne s'impose pas à l'évidence et la solution apportée non plus ! D'ailleurs, la France apparaît bien seule, dans le concert des pays développés, à avoir procédé à une diminution de la durée hebdomadaire du travail. En

réalité, derrière l'affirmation selon laquelle « les robots nous mettront tous au chômage », se cachent deux propositions toutes deux discutables. Selon la première, la robotisation croissante accroît le chômage de masse que nous connaissons aujourd'hui. Et, selon la seconde, nous serions tous susceptibles d'être victimes de la « captation du travail » par les robots, ce qui préluderait à leur hypothétique « prise de pouvoir »... Examinons ces deux propositions l'une après l'autre pour en prendre l'exacte mesure et en comprendre le bien fondé.

Indubitablement, certaines spécialités disparaissent du fait de l'automatisation des processus industriels de production. Le phénomène n'est pas neuf ! L'exemple de la riveteuse cité en introduction en atteste. Cependant, le chômage endémique que nous constatons en France depuis une quarantaine d'années tient plus à la globalisation des échanges et à la délocalisation des activités manufacturières qu'à l'automatisation. Plus que la robotique, la communication quasi-instantanée de l'information sur toute la surface du globe autorise une telle délocalisation des organes de production. On conçoit dès lors que les lois de l'offre et la demande s'appliquent à l'ensemble de la planète et, en conséquence, que des tâches qui requièrent peu de savoir-faire spécialisé s'exécutent là où le travail coûte le moins. D'ailleurs, des études montrent que le nombre des emplois perdus en Europe et en Amérique du Nord dans les industries manufacturières depuis le début des années 1970 correspond à celui des emplois créés en Asie du Sud-Est dans ces mêmes secteurs économiques.

Or, à l'évidence, les robots n'interviennent pas dans ces phénomènes. En revanche, la télécopie d'abord, le courrier

électronique ensuite, le web maintenant prennent une part déterminante dans cette mondialisation du marché du travail. Au demeurant, nous venons de voir que les travailleurs victimes de ce grand bouleversement sont d'abord et surtout les moins formés, ou plus précisément, ceux qui sont les plus interchangeables. De ce fait, la seconde des propositions selon laquelle nous serions tous victimes de cette grande raréfaction du travail, due à la modernité technologique, soulève quelque peu la perplexité. En effet, à l'avenir, la plupart des activités manuelles et routinières susceptibles d'être automatisées ou délocalisées le seront. En revanche, les travaux de conception qui demandent une formation intellectuelle poussée se développent de plus en plus. Et l'expérience montre qu'en dépit des affirmations de technoprophètes comme Moshe Vardi, ces activités ne sont pas vraiment susceptibles d'être délocalisées. À titre d'exemple, le nombre des emplois d'ingénieurs informaticiens a augmenté en Europe et en Amérique du Nord ces dernières années, alors qu'il existait aussi en Inde et en Chine des ingénieurs de grande qualité, payés beaucoup moins chers.

Venons en maintenant à la seconde proposition implicite qui se cache derrière les propos de Moshe Vardi et selon laquelle nous serons tous victimes d'une « captation du travail » par les robots, ce qui préluderait, en quelque sorte, à leur prise de pouvoir. La réalité dément tous les jours cette proposition : en effet, de nouveaux métiers naissent et, plus que jamais, on a besoin d'hommes et de femmes capables de se former rapidement aux techniques actuelles, en s'ouvrant aux évolutions les plus récentes. Or, seule une éducation initiale solide donne les moyens et le goût de retourner à tout âge de la vie à l'université pour acquérir de nouvelles

compétences et de nouveaux savoirs. Autrement dit, seule une formation aux disciplines fondamentales complétée par une formation continue, tout au long de la vie, permettra à chacun d'affronter les défis que nous lancent les évolutions technologiques actuelles.

Au reste, l'expérience montre que les hommes et les femmes compétents doivent travailler de plus en plus pour relever les défis du monde moderne. Loin de se réduire, le temps hebdomadaire de travail s'accroît dans la plupart des pays développés, sans que l'on sache précisément le quantifier. Indubitablement, il ne se comptabilise plus en heures passées à l'établi ou à la chaîne ; la lecture et la formation continue doivent y être intégrées ; la négociation, la participation au groupe et l'écoute des autres aussi... En somme, dans la société moderne, les robots ne mettent pas tout le monde au chômage, bien au contraire, car cette société est d'abord une société de la connaissance, qui sollicite tous les talents. Tous y ont donc une chance, car la richesse ne tient pas à l'héritage mais à la maîtrise, à la production et à l'échange de connaissances. Toutefois, ceux qui se contentent d'une activité routinière et répètent inlassablement le même geste dans l'éternité vide de leur temps de travail doivent prendre garde. Il pourra toujours se trouver une machine pour les remplacer !

« Demain, nous serons les esclaves des machines. »

Dans le choix de développement des « artilects », la mise est énorme : la destruction de l'espèce humaine. […] À partir d'un certain seuil, les machines pourraient construire d'autres ordinateurs plus complexes qu'eux qui, à leur tour… et vvoumm ! il y aurait une explosion, une espèce de réaction en chaîne de l'intelligence. Mais plus probablement, compte tenu de la difficulté de créer un cerveau artificiel de même niveau que celui de l'être humain, il faudra beaucoup de temps, au moins cinquante ans. C'est pourquoi j'estime que la question dominera le XXIe siècle.

Hugo de Garis, *Le Monde*, 9 novembre 1999, pages Horizons – entretien 2000, débats pour le siècle à venir

Hugo de Garis, un scientifique britannique, qui travailla un temps dans un laboratoire japonais avant de poursuivre des recherches en Belgique, professe des théories bien singulières : selon lui, les machines deviendront bientôt si rapides et si petites que le rythme de leurs pulsations et leur densité excéderont ceux de notre propre cerveau. Conséquence inéluctable, Hugo de Garis affirme que des intellects artificiels, des sortes d'insectes, désignés sous le terme d'« artilects » naîtront, prolifèreront, prospéreront, prendront le pouvoir sur nous et nous détruiront. Et telles les tribus d'Indiens d'Amérique au moment de la colonisation espagnole, les sociétés humaines se déchireront entre « Terrans », partisans du Vieux Monde où s'affirme encore, pour peu de temps, la suprématie de l'homme, et « Cosmistes » aspirant à voir

s'épanouir l'esprit, grâce aux machines qu'ils auront créées, même si l'humanité devait s'y trouver condamnée.

Comment adhérer, sans sourire, à ce mauvais scénario de science-fiction, auquel le journal *Le Monde*, l'organe le plus officiel de la presse dans notre pays, a prêté ses colonnes à plusieurs reprises ? (Notamment avec une interview le 9 novembre 1999, et un portrait, en septembre 2000. Bien d'autres journaux ont aussi honoré ce chercheur, comme le journal *Libération*.)

Examinons donc les propositions de ce chercheur à l'aune des connaissances scientifiques actuelles, pour en mesurer l'exacte portée.

Commençons par la prémisse. Nul ne sait aujourd'hui jusqu'où iront la miniaturisation des unités de calcul et leur rapidité d'exécution. Contrairement aux affirmations de M. Hugo de Garis, rien ne nous assure de la pérennité de la loi dite de Moore, selon laquelle la vitesse des processeurs et leur capacité de stockage doublent tous les 18 mois. Certes, cette loi empirique émise en 1965 par Gordon Moore, le co-fondateur de la société Intel, se vérifie plus ou moins depuis 1959. Mais tous savent que les principes physiques sur lesquels repose la conception des circuits électroniques actuels, demandent à être revus si l'on désire poursuivre cette progression sur un rythme identique au-delà de l'an 2020. D'ailleurs, on observe déjà, depuis 2016, un tassement du rythme de progression des performances des processeurs, même si aujourd'hui, cette diminution de croissance se trouve partiellement compensé par la parallélisation massive.

Supposons maintenant qu'en dépit des obstacles prévisibles, ces prémisses se réalisent, à savoir que la densité et la rapidité des processeurs excèdent celles de notre cerveau.

Dès lors, comment imaginer que les hangars miniatures de stockage d'informations ainsi fabriqués prendront le pas sur nos facultés intellectuelles et qu'ils nous étoufferont ? Rappelons-le, ni la quantité d'information engrangée, ni la célérité d'accès et de traitement ne produisent spontanément l'esprit. Entre les prouesses des processus matériels engendrés par les automates et leur entendement, réside une savante alchimie algébrique dont nous ne possédons pas encore la pierre philosophale.

Les mécanismes d'auto-organisation, dont Hugo de Garis prétend maîtriser le secret, demeurent encore mystérieux aux yeux de beaucoup. Dans l'état actuel des publications scientifiques, bien peu de choses autorisent à croire qu'une machine saura d'elle-même s'adapter aux conditions environnantes, acquérir des connaissances et parvenir à une conscience. Oh, le rêve est ancien ! En son temps, la cybernétique* en avait déjà caressé l'idée ; certains conçurent des réseaux d'automates interconnectés qui, à l'instar des neurones de notre cerveau, étaient censés se spécialiser dans différentes fonctions pour contribuer collectivement à l'édification d'un esprit artificiel. De nos jours encore, les tenants de l'intelligence artificielle forte et de l'intelligence artificielle générale poursuivent sans relâche le même projet : ils envisagent, eux aussi, de fabriquer une réplique exacte des mécanismes vitaux à l'aide de systèmes de traitement de l'information, c'est-à-dire d'ordinateurs. Or, ceux qui, parmi les scientifiques, prétendent posséder la formule grâce à laquelle une machine matérielle s'animera d'un souffle de vie et d'un esprit, ne donnent pas d'arguments tangibles pour étayer leurs allégations. Il existe une myriade d'instituts qui prétendent examiner ces hypothèses,

par exemple l'Institut sur le futur de l'humanité (Future of Humanity Institute – University of Oxford), l'Institut du futur de la vie (The Future of Life Institute), l'Institut de recherche sur l'intelligence des machines (Machine Intelligence Research Institute – MIRI), le Centre pour l'étude du risque existentiel (Center for the Study of Existential Risk), l'Université de la singularité (Singularity University), l'Institut pour l'éthique et les technologies émergentes (Institute for Ethics and Emerging Technologies), l'Institut des extropiens etc. Mais tant leur multiplicité que l'absence de preuves scientifiques évidentes témoignent du caractère discutable de ces thèses.

En conclusion, les déclarations d'Hugo de Garis et de multiples personnalités prétendument scientifiques qui annoncent la fin de l'humanité, comme Hans Moravec, Kevin Warwick, Ray Kurzweil et bien d'autres, relèvent plus de la prophétie, c'est-à-dire de la parole inspirée, que du discours rationnel. Il y a donc bien peu de chance que ces prédictions se réalisent et que les ordinateurs électroniques actuels nous réduisent en esclavage.

Pour autant, doit-on se rassurer ? N'y a-t-il aucun risque ? On ne saurait l'affirmer avec certitude. Des dangers existent, mais ils ne viennent pas de là où on les attend. Même si elles ne prennent pas le pouvoir sur nous, les machines ont désormais une part si grande dans la vie quotidienne qu'elles impriment en nous leur marque et qu'elles nous gouvernent de plus en plus. Cette constatation n'est pas neuve, elle a traversé le XX^e^ siècle. Déjà, en 1930, Paul Valéry, ce visionnaire dont la lucidité demeure encore vive, le notait dans ses *Essais quasi politiques*, et en tirait les conséquences : « La machine gouverne. La vie humaine est rigou-

reusement enchaînée par elle, assujettie aux volontés terriblement exactes des mécanismes. Ces créatures des hommes sont exigeantes. Elles réagissent à présent sur leurs créateurs et les façonnent d'après elles. Il leur faut des humains bien dressés ; elles en effacent peu à peu les différences et les rendent propres à leur fonctionnement régulier, à l'uniformité de leurs régimes. Elles se font donc une humanité à leur usage, presque à leur image. »

Depuis que ces lignes ont été écrites, nous assistons à une double évolution. L'une tend à confirmer tous les jours ces propos de Valéry. Au cours de ce siècle, l'empire des machines s'est étendu à la plupart des gestes ordinaires et des pensées quotidiennes. Se déplacer, travailler, se parler, dessiner, écrire, et lire, tout cela passe désormais par le truchement de machines auxquelles nous sommes de plus en plus soumis.

Cependant, tandis que l'homme se retrouve sous l'emprise des machines, un renversement tout à fait singulier, et en quelque sorte inverse, se produit : l'homme a appris, et continue d'apprendre à fabriquer des machines à son image. Les guichets automatiques, les distributeurs de billets, les fers à repasser, les machines à laver, les fours à micro-ondes, tous ces objets qui nous entourent et nous aident nous font de moins en moins peur. Leur maniement est de plus en plus naturel ; ils exigent de moins en moins de nous. Les ordinateurs eux-mêmes se commandent de plus en plus facilement, et semblent de plus en plus familiers. Les machines n'exigent plus de nous que nous nous conformions à leurs exigences ; elles sont de plus en plus faites à notre image, pour nous servir.

Là où un humanisme traditionnel, pénétré d'une culture classique, se contentait et continue de se contenter de

déplorer la présence massive des machines dans le monde et la forme de domination à laquelle elles nous soumettent, l'intelligence artificielle et les sciences cognitives participent d'un nouvel humanisme qui vise à mieux connaître l'homme, et à utiliser cette connaissance de l'homme pour mieux maîtriser son destin.

Ce faisant, ces disciplines nous aident à relever les défis que nous lancent tous les jours ces machines que nous avons fabriquées, et qu'il n'est plus question de ranger au placard des objets encombrants, car elles nous accompagnent quotidiennement. Certes, l'intelligence artificielle ne nous libère pas totalement de ces machines que nous avons créées, mais elle nous aide à les soumettre à nos besoins et à nos capacités.

« Il n'y a pas ou plus de débouchés professionnels en intelligence artificielle. »

« Les entreprises robotisées créent plus d'emplois que les autres », révèle une étude de l'INSEE. Voilà qui rompt avec la vision d'un progrès technique fossoyeur irrémédiable de l'emploi. L'institut explique : « Les nouvelles technologies peuvent être destructrices d'emplois si elles visent à substituer du capital au travail et à accroître la productivité de celui-ci. » Autrement dit : si on a en vue de mettre des machines à la place des hommes et d'économiser encore un peu sur ceux qui restent avec moins de salaire, de qualification, plus de flexibilité et de précarité… on va dans l'impasse sociale et économique. À l'opposé, poursuit l'INSEE, « les changements technologiques peuvent avoir un effet positif sur l'emploi s'ils engendrent un avantage compétitif conduisant à l'expansion des marchés.

Article « Robot et chômage », *L'Humanité*, 9 novembre 1996

Depuis un demi-siècle, les technologies de l'information et de communication ont transformé le monde. En quelques années, le commerce, la finance, les échanges, l'école, le travail, la culture, la politique, se sont totalement modifiés du fait de leur développement. Qu'on se remémore quelques-unes des étapes les plus marquantes de ces évolutions : apparition des mini puis des micro-ordinateurs, nouvelles interfaces avec utilisation de la souris et métaphore du bureau, essor des hypermédias, popularisation du web, nomadisme généralisé, informatique vestimentaire, intelligence d'ambiance, internet des choses… Partout, les évolutions ont été imaginées, conçues, développées, expérimentées dans des

laboratoires de recherche. Partout, l'intelligence artificielle a pris, et continue de prendre, une part déterminante.

Pour mieux le comprendre, prenons quelques exemples.

Le premier article écrit en 1965 par Ted Nelson sur l'hypertexte se réfère explicitement aux techniques d'intelligence artificielle. Il fait appel aux structures de données définies par l'intelligence artificielle pour établir des liens entre les parties d'un texte. Rappelons, à cet égard, pour les monsieur Jourdain de l'Internet qui ne sauraient pas ce qu'est un hypertexte, qu'il s'agit d'un texte enrichi de liens entre ses parties. Cela autorise la navigation, autrement dit, le « butinage » d'un lecteur devenu abeille, qui fait son miel de tout ce qu'il trouve en furetant d'une fleur à une autre du texte. Le web, autrement dit la toile d'araignée mondiale, n'est autre qu'un hypertexte géant à l'échelle de la planète. Le concepteur du web, Tim Berners Lee, l'a ainsi conçu. Et d'ailleurs, le langage le plus utilisé pour rédiger des sites web s'appelle HTML, acronyme d'HyperText Markup Language, ce qui signifie « langage de balises pour hypertexte ». Notons encore, toujours dans le contexte du web, que les moteurs de recherche, et en particulier l'un des plus populaires aujourd'hui, Google, recourent à des techniques d'intelligence artificielle.

Dans un registre différent, les langages de programmation dits « orientés vers les objets » que l'on utilise beaucoup aujourd'hui, viennent en partie des travaux d'intelligence artificielle sur la représentation des connaissances. Et ce sont les besoins de l'intelligence artificielle qui ont initialement conduit des programmeurs à les concevoir.

De même, les logiciels de reconnaissance de la parole comme Siri d'Apple, que l'on emploie dans les serveurs

vocaux ou sur les téléphones portables, de synthèse de la parole, que l'on retrouve dans les agents conversationnels, de reconnaissance des formes visuelles, d'empreintes digitales et de visages, font tous appel à des techniques d'intelligence artificielle et à de l'apprentissage machine qui est une branche de l'intelligence artificielle. Et les assistants personnels, qui remplacent de plus en plus nos antiques carnets de papier, intègrent eux aussi des réseaux de neurones formels grâce auxquels ils reconnaissent nos écritures.

Il existe aussi des petits robots aspirateurs intelligents capables d'explorer une pièce et de la nettoyer automatiquement, en évitant les obstacles, les murs, les escaliers, etc. Les techniques d'intelligence artificielle jouent un rôle clef dans leur fabrication. Et, il en va de même pour les voitures autonomes, pour les robots spatiaux ou pour les drones.

Nous pourrions multiplier les exemples. Dans tous les secteurs d'activités producteurs de richesses, dans le domaine de la santé, de l'électroménager ou, plus exactement, de ce que l'on appelle aujourd'hui la « domotique » ou la « maison intelligente », de l'automobile, de l'aéronautique, des transports ferroviaires, des télécommunications, des médias, etc. l'intelligence artificielle joue un rôle clef. Ajoutons à cela que le secteur très secret du renseignement, à la fois intérieur (DGSI) et extérieur (DGSE), recrute des spécialistes du traitement des masses de données et d'apprentissage machine. Dès lors, comment imaginer sérieusement que la maîtrise des techniques qui contribuent avec succès à la conception des réalisations les plus innovantes dans les secteurs économiques les plus rentables ne procure pas de débouchés professionnels ? On le voit bien, les affirmations selon lesquelles ceux qui seraient spécialisés en

intelligence artificielle n'auraient pas de travail n'ont aucun fondement. À cet égard, à titre personnel, étant enseignant à l'université depuis 35 ans et ayant dirigé pendant 12 ans le diplôme d'études approfondies « Intelligence artificielle, reconnaissance des formes et applications », puis pendant cinq ans le master Erasmus Mundus DMKM (*Data Mining and Knowledge Management* – « Fouille de données et gestion des connaissances »), je peux assurer que les étudiants que j'ai formés en intelligence artificielle, et que je continue de former dans ce secteur, trouvent tous du travail dans le domaine de compétence qui est le leur.

On peut se demander ce qui se produira à l'avenir. Un regard rétrospectif jeté sur les évolutions qui ont eu cours dans les 60 dernières années montre que rien n'est jamais acquis. Aucune évolution n'est totalement prévisible. Rien, dans l'ordre de la technique, n'est prédéterminé. Les étapes évoquées plus haut, ces étapes qui ont scandé les transformations actuelles, par exemple, l'apparition des mini-ordinateurs puis le succès des micro-ordinateurs, les interfaces gestuelles et graphiques, avec la souris et l'écran, l'Internet, le web, les réseaux sociaux et les objets connectés, en ont surpris beaucoup. Et il en a souvent été ainsi dans l'ordre du progrès. Ceux qui n'ont pas vu n'étaient pas nécessairement des esprits obtus, et bien des soi-disant visionnaires se sont trompés, quoiqu'ils ne manquassent pas d'intelligence. Certains arguent de ces incertitudes pour suggérer que les besoins dans le secteur de l'intelligence artificielle ne seraient pas aussi importants dans le futur que ce qu'ils ont été jusqu'ici.

Pourtant, il y a fort à parier que les secteurs de l'économie amenés à se développer dans les pays développés comme le

nôtre toucheront moins à l'agriculture, à l'élevage, à la sidérurgie ou à la production manufacturière qu'aux technologies de l'intelligence. En effet, on constate depuis plus de soixante ans que les connaissances techniques jouent un rôle accru dans la production de richesses. Et il en ira certainement comme cela dans les années à venir. Pour souligner l'importance des mutations en cours, on parle de société de l'information, voire de société de la connaissance, ou même parfois d'âge de la connaissance en les opposant à la société et à l'âge industriels.

Les chefs des 15 gouvernements de l'Union européenne réunis à Lisbonne en l'an 2000 l'avaient bien compris lorsque, prenant acte des défis inhérent à la société contemporaines, ils la caractérisaient comme une « économie de la connaissance » et s'engageaient à préparer la transition de l'Europe vers cette société en modernisant le modèle social, en investissant dans la recherche, l'éducation et les ressources humaines et en luttant contre l'exclusion sociale. Cela devait conduire à créer un espace européen de recherche et d'innovation et à instaurer un climat favorable à la création d'entreprises, et à la mise en place d'une réflexion sur l'éducation et la formation tout au long de la vie.

Indubitablement, le diagnostic était correct et tant les mesures prises à l'époque que les programmes de recherche et d'enseignement financés par la communauté européenne depuis plus de trente ans auraient dû aider l'Europe à relever ces défis et à occuper son rang dans la société de la connaissance qui se met en place. Or, les rapports parlementaires se suivent pour remarquer à la fois le retard pris par rapport aux objectifs de Lisbonne et le décrochage de l'Europe non seulement par rapport aux États-Unis, mais aussi par

rapport à la Chine et aux pays émergents. Citons, à titre d'exemple, Daniel Guarrigue, Rapport d'information sur la politique européenne de recherche et de développement, rapport n° 1095 de l'Assemblée nationale, septembre 2003 ; Hervé Gaymard et Axelle Lemaire, Rapport d'information sur la stratégie numérique de l'union européenne, rapport n° 1409 de l'Assemblée nationale, octobre 2013.

Devons-nous nous y résigner, réduire encore le temps hebdomadaire de travail et considérer à la fois cette régression de l'Europe, l'augmentation du chômage et la crise économique comme des fatalités ? Ou, ne serait-il pas temps de revoir les modalités d'action de la communauté européenne et de la France en matière de financement de la recherche et de l'enseignement, afin de former la jeunesse aux emplois de l'avenir, de former les travailleur à tout âge et de stimuler la création d'entreprises dans le secteur des technologies de l'intelligence et de la connaissance ?

« L'intelligence artificielle constitue un danger existentiel majeur et inéluctable pour l'humanité. »

Tandis que l'impact à court terme de l'IA dépend de qui la contrôle, ses impacts à long terme dépendent uniquement de savoir si elle peut être contrôlée.

Stephen Hawking, Stuart Russell, Max Tegmark, Frank Wilczek, *The Independent*, 1er mai 2014

Le 1er mai 2014, une tribune alarmiste parue dans le journal *The Independent* et signée par quatre éminents scientifiques, Stephen Hawking astrophysicien fameux, Stuart Russell professeur à l'université de Berkeley et auteur d'un manuel sur l'intelligence artificielle qui fait autorité, Max Tegmark physicien et professeur au MIT et Frank Wilczek, physicien, professeur au MIT et prix Nobel de physique, nous alertait des dangers que l'intelligence artificielle fait courir à l'humanité. Selon ces quatre personnalités, nous atteindrons très bientôt un point de non-retour au-delà duquel nous irons inéluctablement à notre perte sans jamais pouvoir revenir en arrière. Aujourd'hui, il serait encore temps ; demain, plus rien ne sera possible !

Cet appel à la vigilance fût suivi de beaucoup d'autres lancés par les mêmes, par exemple par Stephen Hawking à la BBC, ou par Stuart Russel qui a parrainé en 2015 deux lettres ouvertes publiées sur le site de l'Institut du futur de la vie, l'une sur les dangers de l'intelligence artificielle,

l'autre sur les méfaits potentiels des armes autonomes, ou par d'autres, qu'il s'agisse de philosophes, comme Nick Bostrom, ou d'hommes d'affaires très en vue comme Elon Musk et Bill Gates. Partout dans le monde, ces personnalités font des émules. En France, c'est le cas avec un médecin qui est en même temps un homme d'affaires très médiatique, Laurent Alexandre, et un philosophe qui s'est spécialisé dans le transhumanisme, Jean-Michel Besnier. En l'occurrence, ce dernier se demandait dans une émission de radio si l'homme était encore maître de l'intelligence artificielle (France Info, décembre 2014). À titre d'illustration, il mentionnait l'assistance au pilotage des avions et les « robots traders » à la bourse qui l'un et l'autre dessaisissent l'homme du pouvoir d'intervenir.

Ces déclarations publiques annoncent toutes un événement majeur et inquiétant consécutif à l'utilisation massive des technologies de l'information. Elles pointent sur les conséquences dramatiques de cet événement pour l'humanité, sur son inéluctabilité et sur son imminence. Ces trois points méritent d'être examinés l'un après l'autre.

Commençons par les supposées conséquences néfastes pour l'humanité dans son ensemble. Aux dires de Stephen Hawking, le déploiement des technologies d'intelligence artificielle sur des ordinateurs hyperpuissants constituerait « notre plus grande menace existentielle », car les humains ne pourront plus rivaliser avec des machines devenues plus intelligentes qu'eux. Cela sous-entend que ces machines ultra-intelligentes, du fait de leur intelligence, entreraient en rébellion contre nous, prendraient le pouvoir et nous réduiraient en esclavage, ce qui signifie que les machines que nous fabriquerons auront des désirs, des aspirations, des

besoins distincts des nôtres et de ceux que nous leur avons insufflés, autrement dit qu'elles se constitueront en sujets autonomes agissant pour eux-mêmes et donc doués d'une conscience et d'une volonté propres. Or, pour l'instant, les scientifiques ne savent pas comment procéder pour concevoir de telles machines et l'on est loin de comprendre les mécanismes à l'origine de la conscience et de la volonté. D'après leurs auteurs, ces prédictions reposent sur le degré de complexité des machines, calculé en nombre de composants, qui avec le temps deviendrait équivalent, puis supérieur à celui du cerveau. Or, même si l'accroissement des capacités de calcul des machines a permis, ces dernières années, à l'intelligence artificielle de réaliser des prouesses sur des tâches spécifiques, comme les jeux (jeux d'échec, jeu de go ou poker), la reconnaissance faciale, la reconnaissance de la parole ou la conception de voitures autonomes, la complexité, la quantité de stockage d'information et la rapidité de calcul ne produisent pas à elles seules de l'intelligence, loin s'en faut ! De nombreuses facultés cognitives – par exemple, le rire et le rêve – demeurent encore très difficiles, voire impossible à simuler sur des ordinateurs, quand bien même il n'existe pas d'argument tangible permettant d'affirmer qu'elles sont à jamais hors de portée de l'intelligence artificielle.

Ajoutons que ces annonces supposent aussi que les machines formeront une coalition hostile aux hommes, ce qui, là encore, ne repose sur aucun fondement et, en conséquence, paraît difficile à admettre et, plus encore, à prouver…

Le deuxième point porte sur l'inéluctabilité de cet événement catastrophique, ce qui suppose, implicitement, que la technologie se déploie de façon autonome, indépendam-

ment de nous. Cette inéluctabilité se conjugue avec l'imminence, à savoir avec le troisième point, qui découlerait, selon certains auteurs, d'un calcul mathématique issu de l'extrapolation de la loi de Moore. Rappelons que cette loi, émise en 1964, par Gordon Moore, le fondateur de la société Intel, est une loi d'observation qui constate que, depuis 1959, les performances des processeurs doublent tous les 18 mois. Si l'on prolongeait indéfiniment cette loi, nous obtiendrions à terme des ordinateurs infiniment puissants, ce qui paraît tout à la fois contraire à l'intuition et contraire aux anticipations scientifiques des physiciens. C'est pourtant sur l'extrapolation de cette loi d'observation que certains ingénieurs se fondent pour affirmer que les ordinateurs nous dépasseront bientôt. C'est ce qui justifie l'affirmation selon laquelle cette catastrophe que l'on appelle la « Singularité technologique », avec un « S » majuscule pour signifier son unicité et son exceptionnalité, serait à la fois inéluctable et imminente.

En dépit de l'autorité des scientifiques qui, comme Stephen Hawking, se prononcent là, et de la célébrité de personnalités qui, comme Elon Musk ou Bill Gates, leur emboitent le pas, on doit considérer avec circonspection ces prédictions en se demandant ce qui les justifie au plan scientifique. Dans le grand public, beaucoup de personnes impressionnées par la renommée supposent que ces grands personnages disposent d'informations confidentielles qui motivent leurs inquiétudes. Or, à supposer que les informations dont ils ont connaissance soient secrètes, au point qu'ils préfèreraient les tenir scellées, même si la survie de l'humanité était en jeu, cela voudrait dire qu'ils en seraient d'une façon ou d'une autre les bénéficiaires et que, dans

cette éventualité, leur attitude devrait être suspectée puisqu'ils nous dissimuleraient des informations essentielles pour notre devenir collectif. En revanche, si ce n'était pas le cas et s'il n'y avait rien là de caché, ils devraient être en mesure d'expliquer clairement ce qui justifie leurs craintes, ce qu'ils ne font pas de façon convaincante dans leurs déclarations publiques qui demeurent toutes bien elliptiques. Bref, nous ne devons pas renoncer à notre sagacité critique en nous laissant abuser par des annonces « apocalyptiques », au sens étymologique, c'est-à-dire qui prétendent révéler une catastrophe imminente, qu'aucun argument rationnel ne justifie.

« Grâce à l'intelligence artificielle, nous téléchargerons nos consciences et deviendrons immortels ! »

Une analyse de l'histoire de la technologie montre que le changement technologique est exponentiel, contrairement à la perception « intuitivement linéaire » du sens commun. [...] Dans quelques décades, l'intelligence des machines surpassera l'intelligence humaine, conduisant à la Singularité – un changement technologique si rapide et profond qu'il représente une rupture dans le tissu de l'histoire humaine. Les apports comprennent la fusion de l'intelligence biologique et non-biologique, des hommes logiciels immortels et des niveaux d'intelligence ultra-élevés qui se propagent dans l'univers à la vitesse de la lumière.

Ray Kurzweil, *The Law of Accelerating Returns*, 7 mars 2001

Face aux inquiétudes de ceux qui voient dans l'intelligence artificielle un péril existentiel pour l'humanité, s'oppose, comme le revers d'une même médaille, l'enthousiasme de ceux qui aspire au salut de l'humanité par les machines. Partout, la même croyance dans le déploiement autonome de la technologie qui s'animerait d'une vie propre indépendante de l'homme. Mais, contrairement aux premiers qui manifestent leur peur devant l'inconnu, les seconds se réjouissent d'une prolongation de la vie au-delà de la mort biologique. Pour eux, c'est certain, la technologie permettra très bientôt de faire sortir la conscience de son substrat naturel, le cerveau, et de l'hybrider à des ordinateurs hyperpuissants dans lesquels elle poursuivra son cours. Cela ouvre, si ce n'est sur

la vie éternelle, du moins sur une prolongation indéfinie de la vie de l'esprit.

Cette espérance s'inscrit dans une perspective plus large selon laquelle l'évolution des machines prendrait le relais de l'évolution naturelle. Plus précisément, la grande épopée du progrès qui aurait commencé avec l'organisation de la matière physique pour aller ensuite d'abord à la vie, depuis ses formes élémentaires, jusqu'à ses formes les plus achevées, en particulier jusqu'à l'espèce humaine, puis aux cultures humaines, en particulier aux langues, aux spiritualités variées, aux arts et aux connaissances scientifiques, avant de parvenir aux réalisations matérielles actuelles, se poursuivrait avec la technologie. Cependant, tandis que, toujours dans cette perspective, l'humanité aurait été depuis son origine, ou presque, et jusqu'à peu, le moteur de ces transformations, elle n'en deviendrait bientôt plus que l'objet, et perdrait sa liberté, alors que la technologie prendrait le relais et contribuerait à l'essor de l'esprit. En cause, les ordinateurs hyperrapides qui, doués de facultés d'apprentissage automatique sur de très grandes masses de données, seront très bientôt en mesure de se perfectionner d'eux-mêmes et de se reproduire, sans le secours des hommes. L'accroissement des performances des machines, qui seul permet cet apprentissage automatique, repose sur la généralisation de la loi de Moore selon laquelle la vitesse et la capacité de stockage des processeurs s'accrurent et continueront de s'accroître sur un rythme exponentiel jusqu'à dépasser soudainement les capacités humaines. D'après les calculs savants de scientifiques reconnus comme Ray Kurzweil, directeur scientifique chez Google, cela se produira en l'an 2045. Cependant, là où, utilisant les mêmes arguments, d'autres craignent ce

moment critique et les risques de basculement dans l'au-delà de l'humain qu'il provoquera, ceux-ci espèrent qu'une symbiose de l'espèce humaine et des machines adviendra et conduira à l'épanouissement de l'esprit.

D'après eux, cette grande course de l'évolution universelle s'achèvera grâce à la technologie dans une sorte de plérôme, entendu au sens d'une plénitude de l'être déployant, enfin, l'ensemble de ses potentialités.

L'un des plus fervents zélateurs de cette hypothèse, Ray Kurzweil, décrit, dans ses nombreux ouvrages, les six phases de cette grande histoire. Selon lui, la première aurait commencé avec le Big Bang, et se serait poursuivie avec la naissance du premier électron, des protons et des atomes, puis avec l'élaboration progressive de la matière organisée. Serait venue ensuite l'ère de la vie, avec l'ADN, les cellules, les tissus biologiques et les premiers organismes. Dans un troisième temps, seraient advenus des mammifères dotés de cerveaux de plus en plus perfectionnés, jusqu'à l'homme. Au cours d'une quatrième phase, plus brève, les technologies conçues par l'Homme se seraient perfectionnées à grande vitesse. Aujourd'hui, nous parviendrions à une cinquième phase où les technologies initialement conçues par l'homme pour le servir prendraient leur autonomie, se perfectionneraient d'elles-mêmes et se grefferaient sur la matière organique pour donner naissance à des cyber-organismes et à une humanité augmentée. Enfin, dans une sixième phase, apothéose de l'esprit, l'univers se réveillerait et s'emplirait d'une intelligence d'ordre essentiellement technologique dont le règne succèdera, à n'en pas douter, à celui du vivant.

Ces thèses, pour étonnantes qu'elles apparaissent aux néophytes, font florès dans les entreprises de haute techno-

logie. Beaucoup de personnalités averties et influentes du monde contemporain les discutent au sein d'une kyrielle de groupes de réflexion assez fermés dont l'Institut du futur de la vie, financé assez largement par des industriels du secteur des technologies de l'information comme Elon Musk, et l'Université de la singularité qui compte parmi ses fondateurs des grands industriels comme Google, Cisco, Nokia, Genentech, Autodesk, etc. Pour donner une idée de l'écho que recueillent ces conceptions, évoquons les bonnes œuvres du milliardaire russe Dmitry Itskov : ce dernier craint qu'une fois nos consciences téléchargées sur des machines hyper-intelligentes, nous ne soyons désincarnés, au sens propre, car réduits à un état purement informationnel. Pour nous sauver, sa philanthropie l'a conduit à monter une entreprise qui vend des corps robotisés que nous pourrons mouvoir à distance avec notre conscience téléchargée afin de nous réincarner et de poursuivre ainsi une existence mondaine. Comme il reste moins de 30 ans avant l'instant ultime, tous peuvent dès à présent commencer à concevoir ce corps qui accueillera leur esprit pour l'éternité, ou presque, et s'exercer à sa manipulation en ayant recours aux technologies actuelles des interfaces cerveau-ordinateur. Il suffit de se rendre sur le site « 2045.com » de cette société et d'appuyer sur le « bouton d'immortalité » *(Immortality Button)* pour initier la conception de son futur avatar…

Doit-on croire à ces thèses et se soucier tout de suite de son salut par l'intelligence artificielle en achetant un avatar robotisé ? C'est bien évidemment une question très personnelle à laquelle nous n'oserons pas apporter de réponse ici, d'autant plus que la science actuelle n'est pas en mesure de démontrer rigoureusement l'impossibilité d'un tel scénario,

même si, pour de multiples raisons, la probabilité qu'il advienne paraît aujourd'hui extrêmement faible, si faible qu'on doit pouvoir la considérer comme négligeable. Parmi ces raisons, citons en deux. D'une part, la loi de Moore n'est pas éternelle. Dès aujourd'hui, beaucoup entrevoient les limites des technologies du silicium ce qui conduisit déjà, en 2016, à un tassement du rythme de progression des processeurs, autrement dit à un signe avant-coureur de la fin de la loi de Moore. D'autre part, quand bien même la loi de Moore se poursuivrait indéfiniment, l'intelligence ne se réduit pas à une fréquence de calcul. Des phénomènes comme l'émergence d'une conscience ou d'une volonté demeurent encore si mystérieux que l'on n'a aucune idée de la façon dont une machine pourrait être construite pour les reproduire.

Au reste une question majeure demeure : pourquoi des sociétés aussi prestigieuses que Google et des hommes d'affaires aussi connus qu'Elon Musk ou Bill Gates investissent-ils autant dans les institutions précédemment mentionnées dont les thèses se révèlent aussi peu étayées au plan scientifique ? Y croient-ils vraiment ? Ou ne s'intéressent-ils qu'à des thèmes populaires qui défraient les chroniques journalistiques, pour faire parler d'eux ?...

« La machine est l'avenir de l'homme. »

Les êtres humains ont créé un million d'explications sur la signification de la vie dans les arts, en poésie, dans les formules mathématiques. Certainement, les êtres humains doivent être la clef de la signification de l'existence, mais les êtres humains n'existent plus.

Steven Spielberg, *AI – Artificial Intelligence*, 2001
(début de la troisième partie du film, après que David a passé plus de 2000 ans sous la glace – paroles prononcées en voix off et attribuées à un robot)

Quel est l'avenir de l'homme ? Peut-on croire Aragon lorsqu'il affirme que c'est la femme ? D'un point de vue individuel, la chose se défend, encore que les avis demeurent partagés, même chez les poètes et les chansonniers. Citons par exemple Jacques Brel qui dit, dans *La ville s'endormait*, n'être « pas bien sûr/Comme chante un certain/Qu'elles [les femmes] soient l'avenir de l'homme ». Mais la question s'entend autrement ici : c'est de la phylogénie et non de l'ontogénie dont nous souhaiterions traiter, de l'homme en tant qu'espèce et non de l'individu, de nos successeurs, c'est-à-dire de ceux qui recueilleront l'héritage de l'humanité, et non de nous-mêmes.

Pour dire les choses différemment, selon les biologistes, nous descendons du singe ou de ce que l'on appelle, en jargon scientifique, des primates supérieurs. Nous sommes donc redevables à ces ancêtres, dont nous constituons, en quelque sorte, l'avenir rétrospectif. Et sans aucun doute, la dénomination de primates se réfère-t-elle à ce caractère pre-

mier du singe par rapport à nous. Cette dette à l'égard de nos grands aînés explique les raisons de ceux de nos contemporains qui proposent de leur octroyer le bénéfice de droits humains. Rappelons que pour les défenseurs du projet « grand singe » (« The Great Ape Project »), nous devons respect à toutes les espèces vivantes en raison de leurs capacités cognitives. Or les primates supérieurs disposent de facultés intellectuelles supérieures à celles de certains d'entre nous, par exemple à celles des enfants en bas âge, des personnes âgées, de certains accidentés et des humains atteints de déficience mentale. On devrait donc étendre les droits de l'Homme, de sorte que les primates supérieurs puissent en tirer un avantage à la mesure de la considération que nous leur devons.

Notons, bien sûr, qu'à titre d'ingénieur, la tentation serait grande de reprendre l'argument pour défendre des droits des machines analogues aux droits de l'Homme. C'est d'ailleurs ce qui fait l'objet d'un rapport du parlement européen. Cependant, notre question ici est autre : nous souhaiterions savoir ce qui, dans une perspective darwinienne, succédera à l'espèce humaine, lorsque son règne s'achèvera. À cet égard, on conçoit, à la fin d'un ouvrage consacré aux idées reçues sur l'intelligence artificielle, que les questions d'avenir n'y traitent pas rétroactivement du passé et de nos ancêtres, mais au contraire, à titre prospectif, du futur, ou plus exactement de notre postérité, lorsque l'espèce humaine aura disparu.

Pour les écologistes, l'avenir appartient aux espèces les plus résistantes. Selon eux, les rats seraient bien placés ; les goélands et les scorpions aussi ; peut-être les serpents auraient-ils une chance… Or, dans l'hypothèse d'une disparition de l'humanité, ces espèces se présenteront au mieux

comme des parasites, au pire comme des charognards, qui engloutiront nos dépouilles sans pouvoir prétendre ni tirer parti de notre héritage, ni *a fortiori* constituer notre avenir. Et il en irait de même si une espèce naturelle plus intelligente et plus résistante venait à nous envahir et à nous éliminer de la surface de la Terre. Comment imaginer que de tels prédateurs supportent des vestiges des civilisations qui les ont précédés et qu'ils ont à jamais anéanties ?

N'existe-t-il pas d'autres perspectives, moins sombres pour l'humanité après son extinction ? Comment pourrions nous subsister au-delà de nos individualités, de nos familles, de nos nations, de nos races et de notre espèce ? À supposer que le monde ne régresse pas vers un état minéral inerte, deux perspectives s'ouvrent à nous pour échapper à une éradication totale des restes de nos civilisations.

Selon la première, la greffe de machines sur et dans nos corps nous permettra de vivre, en dépit des conditions matérielles les plus pénibles. Que l'on songe aux exo-squelettes, aux scaphandres sous-marins, aux pompes à insuline, aux stimulateurs cardiaques et autres prouesses de la micro-robotique et des nanotechnologies, et nous aurons une idée de ce que l'avenir réserve à nos descendants. Sans aucun doute, ce seront des hybrides de machines et d'organismes animaux, autrement dit des « cyborgs ». Ainsi, greffés sur des machines, nos héritiers en seront les parasites et les commensaux.

Dans la seconde, après la disparition de son autonomie matérielle, l'espèce humaine s'éteindra totalement. Et, avec elle, disparaîtront à jamais les souvenirs de nos aventures, de nos chefs-d'œuvre, de nos guerres, de nos héros, des idéaux qui nous ont animés, des spiritualités auxquelles nous avons

crû, etc. À moins que des êtres bienveillants forgés par nous, pour nous survivre, ne prennent le relais de nos mémoires et ne transmettent notre héritage. Ces robots témoigneront, pour les siècles des siècles, de notre grandeur. Ils assumeront à jamais notre mémoire.

Un cinéaste de science-fiction, Steven Spielberg, fit sienne cette hypothèse : la fin du film *AI, Artificial Intelligence*, qu'il a fait paraître sur les écrans en 2001, évoque un futur peuplé de machines nostalgiques qui, toutes semblables à des girafes langoureuses et douées d'une exquise sensibilité, se souviennent de notre présence. Il ne manque d'ailleurs pas de défenseurs de la thèse selon laquelle les machines constitueraient, en quelque sorte, l'avenir de l'homme.

En somme, dans le meilleur des mondes, c'est-à-dire dans celui où une trace de notre présence subsistera, l'une des deux éventualités que nous venons d'envisager se réalisera. Or, dans ces deux éventualités, les machines prennent une place centrale. Dans la première, nos descendants se transformeront en cyborgs ; les machines les assisteront dans toutes les tâches les plus élémentaires, au point qu'ils ne sauront plus s'en passer. Dans la seconde, on confiera à des robots bienveillants le soin d'assumer notre héritage une fois que notre espèce aura disparu. Dans tous les cas, la machine est l'avenir de l'homme !

Et comme, selon les philosophes, l'impératif éthique commande tout à la fois de se perfectionner et d'œuvrer au bonheur des autres, quoi de plus éthique que de fabriquer des machines intelligentes ! En effet, à défaut de contribuer au bonheur de nos congénères, les machines intelligentes attesteront de nos qualités et de notre perfection tout en se souvenant, pour les siècles des siècles, de nos présences et de nos réalisations.

conclusion

Désormais, la maîtrise des automates qui peuplent notre univers matériel et psychique ne passe plus par une anticipation mentale. Pris isolément, chaque maillon des enchaînements logiques exécutés par les machines se déchiffre aisément ; nous en appréhendons le sens et la teneur ; pris dans la masse, nous ne les percevons plus, ils deviennent insignifiants et abstraits. Le nombre et la fréquence des opérations nous submergent. La complexité de l'intelligence artificielle dépasse notre entendement immédiat. L'expérience montre que dans le temps de l'action, nous restons démunis. D'ailleurs, l'analyse des accidents les plus récents dans les centrales nucléaires ou les avions, là où les technologies modernes se trouvent les plus répandues, prouve que les techniques contemporaines se révèlent remarquablement fiables et que les dysfonctionnements proviennent presque tous d'un défaut de communication entre hommes et machines. Trop rapides, elles nous fourvoient ; trop lentes, elles nous ennuient ; notre rythme n'est pas le leur. Nous interprétons mal les informations qu'elles nous fournissent ; et mal informés, nous les commandons mal.

Dans la vie de tous les jours, pour pallier toutes ces difficultés, nous usons d'une ruse. Au lieu d'appréhender rigoureusement les mouvements des machines, nous leur attribuons une personnalité, et cette stratégie enfantine porte ses fruits. Pour surprenante qu'elle paraisse, cette

constatation se vérifie quotidiennement : qui sait, à part les ingénieurs qui les ont conçus, comment fonctionnent les téléphones portables ? Pourtant, tous, même les plus jeunes et les plus ignorants, savent les utiliser. Les ingénieurs tirent grandement parti de cette ruse par laquelle notre esprit fait semblant de croire que les machines possèdent des intentions, des connaissances, des émotions, afin de mieux les maîtriser. Ils cherchent à susciter un anthropomorphisme partiel, au moyen de signes extérieurs facilement repérables. L'objet, par sa forme, nous parle et nous invite à le saisir de telle ou telle manière. C'est ce que Donald Norman, l'homme qui conçut les interfaces graphiques du Macintosh, appelle d'un terme difficilement traduisible, l'*affordance* des objets. Une poignée de porte nous dit implicitement : « Pousse ! », s'il lui manque une prise pour tirer. De même un programme informatique nous en raconte tout autant, sinon plus, par les icônes qu'il nous donne à voir et par les bruits de clochettes, les couinements ou les claquements qu'il nous fait entendre, que par les messages écrits qu'il nous envoie. *A fortiori*, un robot communique mieux par sa physionomie que par ses paroles ; les automates qui nous font part de leurs états d'esprit par quelques mimiques (sourires, pleurs, bouche ouverte…) sont souvent bien plus faciles à utiliser que ceux qui exigent de décrypter leur mode d'emploi.

Si cette ruse de l'intelligence, qui prête un esprit aux machines, doit être prise au sérieux, c'est qu'à côté de la rationalité scientifique du chercheur et de l'ingénieur coexiste une autre rationalité que nous qualifierons d'animiste, en ce qu'elle procède par attribution d'une « âme » aux objets. Entendons-nous bien, il ne s'agit pas d'un animisme métaphysique ayant partie liée à un chamanisme

ancien. Nous ne présumons pas ici l'existence de génies tutélaires ou d'esprits vagabonds. Comme nous venons de le voir, l'homme d'action prête, très temporairement, des buts, des sentiments, une personnalité, c'est-à-dire une « âme » aux machines modernes, et ce prêt est payé de retour s'il lui permet d'anticiper leurs comportements et, par là, de les soumettre à ses besoins.

La sortie en 2001 du film de Spielberg *IA – Artificial Intelligence* fait manifestement état d'un tel retour à l'animisme au point que l'on pourrait, en en paraphrasant le titre, résumer ce retournement par une formule lapidaire : IA – Informatique Animiste. Ce renouveau de l'animisme à l'aube du XXIe siècle témoigne-t-il d'une régression vers une forme de pensée archaïque ? À moins que l'animisme en question ne relève d'un mode de rationalité toujours à l'œuvre, dans toutes les sociétés ou bien encore qu'il manifeste l'essor d'une pensée radicalement nouvelle qu'il convient d'élucider pour saisir la singularité de notre époque ? Telles sont les questions sur lesquelles ouvrent aujourd'hui les développements actuels de l'intelligence artificielle.

ANNEXES

glossaire

Algorithmes génétiques : ils miment, avec des techniques algorithmiques, autrement dit avec des ordinateurs, les phénomènes d'adaptation des espèces aux conditions environnementales. Ils simulent l'évolution de populations en se plaçant dans une perspective darwinienne, où les individus les plus adaptés se reproduisent et donnent à leurs enfants une partie de leur patrimoine génétique, tandis que les moins adaptés disparaissent sans descendance. Introduits en 1967 par J. Bagley, ils se sont ensuite fortement développés sous l'influence de plusieurs chercheurs dont J. Holland, D. Goldberg, etc.

Connexionnisme – néo-connexionnisme : à partir du milieu des années 1940, on a essayé de simuler l'intelligence en construisant des réseaux d'automates reliés entre eux par des connexions analogues aux synapses, qui s'établissent spontanément dans notre cerveau, entre les neurones. Le connexionnisme recouvre ces premières tentatives à l'aide de réseaux d'automates, qui s'organisent spontanément, comme on suppose que le font les cellules de notre cerveau. Après avoir rencontré des échecs dans les années 1950, ces tentatives ont de nouveau motivé les chercheurs dans les années 1980 au cours desquelles elles ont beaucoup progressé. Ce renouveau des approches connexionnistes se range sous le vocable de néo-connexionnisme.

Cybernétique : forgé en 1947 par Norbert Wiener à partir du grec *kubernêtikê* où il signifiait le pilote d'un navire, littéralement, l'homme de barre, ce terme désigne l'étude des systèmes complexes, de leur évolution et de leur régulation. Dans son principe, la cybernétique peut être appliquée à l'observation des phénomènes physiques, biologiques ou sociaux.

Fonction booléenne : George Boole (1815-1864) mathématisa les lois de la pensée en traduisant la logique des propositions dans un formalisme algébrique. Il assimila le faux au nombre 0 et le vrai au nombre 1, puis il introduisit des fonctions numériques correspondant aux opérations logiques élémentaires comme la conjonction (le « et »), la négation ou la disjonction, autrement dit le « ou ». Depuis, en hommage à George Boole, on a pris l'habitude d'appeler « fonctions booléennes » toutes les fonctions mathématiques analogues qui portent sur l'ensemble comprenant les deux entiers 0 et 1.

Informatique évolutionniste : elle simule, sur des ordinateurs, des phénomènes d'évolution. Les algorithmes génétiques relèvent de l'informatique évolutionniste ; la simulation de colonies de fourmis, ou de phénomènes physiques dits de recuits simulés, aussi.

Ingénierie des connaissances : les connaissances techniques prennent une part de plus en plus grande dans les sociétés contemporaines, au point qu'on qualifie parfois ces dernières de « sociétés de la connaissance ». L'intelligence artificielle joue alors un rôle clef, car elle permet de forma-

liser les connaissances expertes pour en faciliter le stockage, l'échange et la mise en œuvre par des programmes informatiques. Cependant, la formalisation des connaissances ne relève pas uniquement de la compétence d'informaticiens. Elle requiert aussi la maîtrise de principes de psychologie cognitive et de sociologie des organisations. À cette fin, une nouvelle discipline est née aux marges de l'ingénierie des systèmes informatiques ; elle porte le nom d'ingénierie des connaissances.

Intelligence collective : selon Aristote, l'homme, comme les abeilles, les termites ou les fourmis, est un animal politique. Son intelligence ne tient pas seulement à ses capacités individuelles, mais aussi à la société dans laquelle il s'insère. Autrement dit, l'intelligence ne provient pas uniquement des capacités de notre cerveau ou de notre psychisme, mais aussi de la collectivité dans laquelle nous nous insérons. L'éthologie a essayé de caractériser les principes élémentaires qui donnent naissance à ces phénomènes dits d'intelligence collective ou d'intelligence en essaim ; l'intelligence artificielle tente de les simuler à l'aide de réseaux d'automates.

Mathématiques discrètes : les mathématiques discrètes constituent une branche des mathématiques ; elles traitent des objets discontinus, comme par exemple les nombres entiers ou les graphes.

Système expert : un système informatique construit à partir du savoir d'hommes de métier et apte à résoudre les problèmes relevant de leur champ de compétence est appelé un « système expert ».

pour aller plus loin

Une encyclopédie générale en français qui donne des informations précises sur toutes les techniques informatiques : Akoka J. & Comyn-Wattiau I. (dir.), *Encyclopédie de l'informatique et des systèmes d'information*, Vuibert, 2007.

Deux ouvrages classiques en français sur l'intelligence artificielle : Laurière J.-L., *Intelligence artificielle : résolution de problèmes par l'Homme et la machine*, Eyrolles, 1987 et Nilsson N., *Principes d'intelligence artificielle*, Cépadues, 1988.

Un livre classique sur les fondements épistémologiques de l'intelligence artificielle par l'un des pionniers de ce domaine : Herbert Simon, *Sciences des systèmes, sciences de l'artificiel*, Bordas, 1991.

Un ouvrage de référence en anglais et actuel sur l'intelligence artificielle : Russel S. & Norvig P., *Artificial Intelligence a Modern Approach*, Prentice Hall Series in Artificial Intelligence, 1995.

Un livre passionnant sur la vie d'Alan Turing : Andrew Hodges, *Alan Turing ou l'énigme de l'intelligence*, Bibliothèque scientifique Payot, 2004.

Un recueil de textes classiques sur la cybernétique et les sciences cognitives. On y trouvera en particulier le fameux article de Turing traduit en français : *Sciences cognitives, textes fondateurs (1943-1950)*, textes rassemblés et traduits par A. Pelissier, présentés et annotés par A. Tête, PUF, collection « Psychologie et sciences de la pensée », 1994.

Les deux articles originaux d'Alan Turing sur l'intelligence des machines : Alan Turing (1948) "Intelligent Machinery", *National Physical Laboratory Report*, B. Meltzer and D. Michie (eds), Machine Intelligence 5, Edinburgh University Press, 1969 ; Alan Turing (1950) "Computing Machinery and Intelligence", *Mind* 59 (236), pp. 433-60.

Un livre écrit par un philosophe, John Searle, où il décrit la très célèbre expérience dite de la « Chambre chinoise » et où il introduit la distinction entre l'intelligence artificielle forte et l'intelligence artificielle faible : John Searle, *Du cerveau au savoir*, Hermann, éditeurs des sciences et des arts, 1985.

Un ouvrage d'imagination et de fantaisie sur l'intelligence artificielle et la créativité : Douglas Hofstadter, *Gödel, Escher, Bach, les brins d'une guirlande éternelle*, InterÉdition, 1985.

Deux ouvrages de fonds sur les interfaces homme-machine : Byron Reeves, Clifford Nass, *The Media Equation*, Cambridge University Press, 1996, et Donald Norman, *The Design of Everyday Things*, Basic Books, 2002.

Deux livres sur le cerveau et les neurosciences : Antonio Damasio, *L'Erreur de Descartes : la raison des émotions*, Odile Jacob, 1995, et Jean-Noël Missa, *L'Esprit-cerveau : la*

philosophie de l'esprit à la lumière des neurosciences, collection « Pour demain », Vrin, 1993.

Un article sur le projet européen *Human Brain Project* de simulation du cerveau à l'aide de techniques d'intelligence artificielle : Lee Gomes, « Facebook AI Director Yann LeCun on His Quest to Unleash Deep Learning and Make Machines Smarter », spectrum.ieee.org.

Un livre sur la modélisation des joueurs d'échecs avec des techniques d'intelligence artificielle : Fernand Gobet, *Les Mémoires d'un joueur d'échecs*, Éditions Universitaires (Fribourg), 1993.

Un livre sur la construction du programme Deep Blue qui l'a emporté sur le champion du monde en titre aux échecs, Garry Kasparov, en 1997 : Hsu, Feng-Hsiung, *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion*, Princeton University Press, 2002.

Enfin, il existe aussi une jolie bande dessinée sur l'histoire de l'intelligence artificielle : Jean-Noël Lafargue, Marion Montaigne, *L'Intelligence artificielle. Fantasmes et réalités*, La Petite Bédéthèque des Savoirs, 2016.