

Faut-il avoir peur de l'intelligence artificielle (IA) ?

Entretien avec Jean-Gabriel Ganascia

Travail, transports, vie quotidienne... L'intelligence artificielle embrasse aujourd'hui de nombreuses facettes de nos vies contemporaines. Un déferlement rapide qui cristallise espoirs et craintes. Celles-ci sont-elles fondées ? Comment lutter contre les idées reçues ? Entretien avec Jean-Gabriel Ganascia, président du comité d'éthique du CNRS et auteur de l'ouvrage "Le mythe de la Singularité — Faut-il craindre l'intelligence artificielle ?" (Seuil).



Siyon Ren via Unsplash CC0

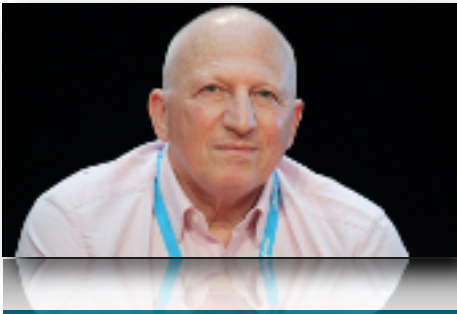
Comment définissez-vous l'intelligence artificielle ?

► **Jean-Gabriel Ganascia** : L'intelligence artificielle est une discipline scientifique dont le nom a été forgé en 1955 par deux jeunes mathématiciens d'à peine 28 ans à l'époque, John

McCarthy et Marvin Lee Minsky. L'objectif initial était de simuler sur des ordinateurs les différentes facultés cognitives — la perception, l'apprentissage, la mémoire, les fonctions exécutives, communicatives ... — dans le but de mieux comprendre ces fonctions mais aussi de rendre des services liés à leur automatisation.

On peut compléter cette définition par une autre, en référence à ce qu'Aristote appelait l'âme au sens rationnel de souffle de vie, ou à ce que l'on entend aujourd'hui par un "agent", et qui correspond à une entité capable de percevoir son environnement, de construire des représentations de cette perception, puis de faire des calculs sur ces représentations, afin de décider et d'agir. En cela, l'intelligence artificielle se définit aussi comme la programmation d'agents artificiels sur des ordinateurs.

Enfin, une dernière acception provient davantage de la science-fiction, voire de l'imaginaire que de la science : il s'agit de construire une entité qui serait douée de capacités intellectuelles équivalentes ou supérieures à celles des hommes. Avec cette dernière définition, l'intelligence artificielle rejoint des mythes anciens que l'on retrouve à travers les âges, dans l'Antiquité, au Moyen-Âge... La légende du Golem en est un exemple, comme celle de Pygmalion : une statue qui prend vie, qui se comporte comme un homme ou une femme pour, éventuellement, conduire à sa perte...



Jean-Gabriel Ganascia via Wikipedia Commons

Vous évoquez en préambule de votre ouvrage *la tribune de Stephen Hawking* prophétisant un développement incontrôlable des machines et la fin de l'humanité telle que nous la connaissons. Vous citez en parallèle de nombreux chercheurs, mathématiciens ou philosophes qui ont réfléchi à la question du développement technologique et qui émettent chacun un avis, plus ou moins pessimiste, sur l'évolution de notre monde. Quelle est votre vision des choses ?

► Lorsqu'on lit et qu'on analyse les arguments sur lesquels elles sont bâties, ces prophéties apocalyptiques ne me semblent pas fondées scientifiquement. Ça ne veut pas dire qu'il n'existe pas de dangers ! Mais, je ne crois pas que ceux-ci puissent venir d'une autonomisation de la technologie qui prendrait d'elle-même le pas sur l'humain pour le détruire. Ces raisonnements s'appuient à la fois sur la notion d'*apprentissage machine* et sur *la loi de Moore*. Or, ni l'apprentissage, ni la loi de Moore ne conduisent inéluctablement à ces évolutions catastrophiques.

Ces risques sont tout à fait illusoire, car la connaissance actuelle que nous avons des technologies de l'intelligence artificielle ne laisse pas présager ce type de scénario. En revanche, le numérique prend une part de plus en plus grande dans notre monde, ce qui comporte un certain nombre de dangers. Mais ceux-ci sont bien moins d'ordre métaphysique que social ou politique !

En effet, à partir du moment où le numérique se généralise, il est de plus en plus difficile de s'en passer. Certes, nous pouvons débrancher nos ordinateurs personnels, mais quid des systèmes informatiques qui régissent les services hospitaliers, les transports ou encore l'état civil ?

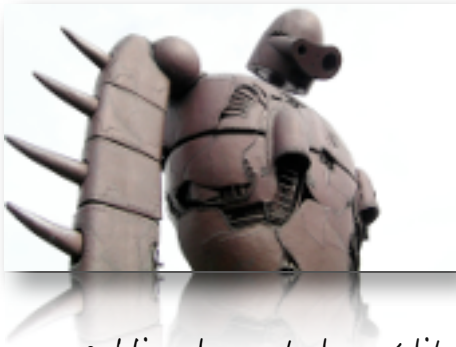
Qu'est-ce que la Singularité technologique, concept que vous placez au centre de votre réflexion ?

► C'est une idée très ancienne qui vient de la science-fiction, dès années 50 et 60, et qui fut inspirée par des mathématiciens comme Stanislaw Ulam, John Von Neumann et Irving John Good. En partant du constat du développement fulgurant des progrès de la technologie des ordinateurs, ces scientifiques dressèrent un parallèle avec le rythme de l'évolution naturelle. Cela les conduisit à se demander qui se passerait si les machines devenaient ultra intelligentes et nous dépassaient.

"Le postulat des singularistes est qu'à un moment donné, si la loi de Moore se poursuit, les performances des machines excéderont les nôtres."

En fait, Von Neumann pressentit ce qu'a ensuite théorisé la loi de Moore : que les technologies se développaient à un rythme exponentiel, c'est-à-dire que la densité des composants doublerait à intervalle régulier — tous les dix-huit mois environ. C'est de cette idée que naît le concept de **Singularité technologique** : leur postulat, c'est qu'à un moment donné, si la loi de Moore se poursuit, les performances des machines excèderont les nôtres. Et pour expliquer cela, Stanislaw Ulam a fait appel à la notion mathématique de singularité, c'est-à-dire de point critique dans une fonction. C'est la même notion qu'a étudié le mathématicien René Thom dans sa **théorie des catastrophes** qui est une branche de la théorie mathématique des singularités.

Une singularité, c'est un point critique, autrement dit un point mal défini — que l'on retrouve également en physique, notamment dans la théorie des trous noirs. L'idée selon laquelle l'évolution des technologies présenterait une singularité a été ensuite reprise par des auteurs de science-fiction dans les années 70, puis par des chercheurs en informatique, tels que **Ray Kurzweil** ou **Hans Moravec**. Ils ont réussi à convaincre des industriels comme les dirigeants de Google, Elon Musk ou encore Bill Gates, qu'en vertu de cette évolution "naturelle" de la technologie, les machines s'autonomiseraient dans un futur proche. Ce qui conduirait à la fin de l'espèce humaine telle que nous la connaissons, comme le prédit Stephan Hawking, ou alors à une certaine forme d'immortalité des hommes qui seraient capables de télécharger leur esprit sur une machine, comme l'évoque Ray Kurzweil.



Nicholas Wang via Flickr CC BY 2.0

Que penser de ce type de scénario ?

► C'est une idée amusante, mais pas très réaliste ! Ce qui m'étonne le plus, c'est que des personnes qui font figure d'autorité dans leur milieu, et qui sont respectées du grand public, donnent du crédit à des théories totalement fantasmatiques ! C'est d'ailleurs la raison pour laquelle j'ai voulu écrire ce livre : il m'a semblé important, en tant que scientifique, d'expliquer pourquoi les arguments employés par ces personnes ne tiennent pas.

Je discute parfois avec des gens qui me disent : "Ils doivent savoir des choses que l'on ne sait pas..." En fait, non : nous avons tous les mêmes connaissances dans ce domaine, et leurs arguments ne laissent pas entendre qu'ils pourraient être au courant de choses que certains de leurs pairs ne sauraient pas, ou qui seraient cachées au commun des mortels.

"On ne peut fonder des affirmations sur des hypothèses aussi hasardeuses que le mur du silicium ou encore l'arrivée des ordinateurs quantiques."

Plus précisément, leurs arguments reposent d'abord sur l'hypothèse selon laquelle la loi de Moore se poursuivrait indéfiniment, alors qu'il ne s'agit que d'une loi d'observation qui atteindra ses limites physiques très bientôt. En effet, nous savons que la miniaturisation ne peut aller au delà d'un certain seuil avec les technologies actuelles de fabrication des microprocesseurs fondées sur l'utilisation du silicium : c'est ce que l'on appelle le "mur du silicium". Certains expliquent que d'autres paradigmes verront le jour, comme par exemple l'ordinateur quantique. Mais, pour l'heure, nous n'avons aucune certitude là-dessus et on ne peut sérieusement fonder des affirmations sur des hypothèses aussi hasardeuses que celles-ci. De manière plus générale, la conjecture selon laquelle la loi de Moore se réaliserait sans le concours des hommes revient à postuler un déterminisme technologique indépendant de nous, alors que l'essence même de la technique provient de nous.

La seconde hypothèse sur laquelle ils se fondent est que selon eux, avec l'apprentissage machine, les machines se perfectionneront d'elles-mêmes jusqu'à nous surpasser et à agir non plus pour nous, mais pour elles. Or si effectivement, avec l'apprentissage profond (deep learning), les machines deviennent extrêmement puissantes au point de nous dépasser parfois dans l'exécution de certaines tâches, comme les jeux ou la reconnaissance des visages, cela ne les conduit pas pour autant à s'autonomiser ! L'apprentissage machine dont nous parlons aujourd'hui n'est en effet qu'un apprentissage supervisé qui demeure incapable d'acquérir de nouveaux concepts par lui-même, et de procéder à l'équivalent de révolutions conceptuelles comme celles de Galilée, d'Einstein ou de Schrödinger en physique.

Dans ce cas, comment répondre aux craintes générées par de tels scénarios, relayées par un certain courant de "techno-prophètes" ?

► Il faut répondre en tant que scientifique. Et ce, non pas parce que le statut de scientifique vous donne de facto l'autorité nécessaire pour le faire, mais parce qu'adopter un discours scientifique permet d'indiquer précisément les limites entre ce que l'on sait avec certitude et ce qui est douteux ou faux.

Je considère que les scientifiques doivent avoir une place dans l'espace public, pour expliquer assez clairement à la fois la puissance des techniques qu'ils contribuent à forger, les enjeux qu'il y a derrière et en même temps les limites de ces techniques, pour balayer tout malentendu.

Dans votre livre, vous évoquez l'horizon d'une hybridation entre homme et machine, aux confins du transhumanisme. Concrètement, à quoi pourrait ressembler ce futur plus ou moins proche ?

► Cette hybridation existe déjà en partie : certains d'entre nous portent des stimulateurs cardiaques, des lunettes ou des montres connectées, des prothèses auditives, des exosquelettes... Ce que l'on peut imaginer, c'est un prolongement de cette hybridation par

le biais de prothèses encore plus perfectionnées et intrusives, comme par exemple des implants cérébraux.

Selon moi, c'est toujours la même question qui va se poser : le remplacement total ou partiel d'un organe ou d'un membre défaillant. Il n'y a donc pas de transformation majeure de l'homme, si ce n'est qu'il va vivre plus longtemps. Attention également à ce que l'on appelle "homme augmenté" : un homme avec une voiture, c'est déjà un homme augmenté ! Enfin, il faut également prendre en compte la possibilité de réversibilité : serons-nous toujours capables d'ôter les dispositifs qui nous équipent ?



Mr Wabu via Flickr CC BY 2.0

Faut-il craindre l'intelligence artificielle, comme le questionne le sous-titre de votre ouvrage ?

► Il y a beaucoup de peurs cristallisées autour de l'I.A., que j'essaie de démonter une à une dans mon ouvrage, car elles ne se fondent sur rien de tangible. Au début, je voulais d'ailleurs appeler ce livre L'Apocalypse, car ces techno-prophètes de l'intelligence artificielle (IA) sont un peu des charlatans, en ce sens qu'ils nous annoncent une apocalypse, au sens étymologique du mot, à savoir une forme de révélation, sans donner de justification tangibles à leurs annonces.

"Les risques technologiques sont maîtrisables, alors que le risque d'un déséquilibre social généré par ces technologies l'est nettement moins."

Mais ce n'est pas parce que ces peurs ne sont pas fondées que le danger n'existe pas. Les transformations du monde consécutives à l'intelligence artificielle sont extrêmement importantes, et que cela ne se fait pas sans risques. Les technologies que nous utilisons au quotidien ne sont pas totalement sûres : il arrive que des voitures autonomes tuent des gens. Elles ne s'imposeront d'ailleurs que si le nombre d'accidents qu'elles génèrent est inférieur à celui des accidents causés par des voitures non-autonomes.

De façon plus générale, ce sont plutôt les conséquences d'ordre politique qui peuvent être catastrophiques. Les risques technologiques, eux, sont maîtrisables, alors que le risque d'un déséquilibre social généré par ces technologies l'est nettement moins. Si la transformation du travail engendre une crise majeure, nous pourrions nous retrouver face à un nouveau phénomène Trump, porté au pouvoir par une partie de la société qui s'est sentie marginalisée.

De même, le fait que les États aient moins de prise sur ce qui est développé au niveau de l'intelligence artificielle peut engendrer des déséquilibres politiques majeurs. Et c'est dans les années qui vont suivre qu'il va falloir se montrer vigilants, de façon collective, pour ne pas se laisser dépasser.

Les Etats ont donc un rôle à jouer ?

► Il faut véritablement qu'ils prennent la mesure des enjeux liés au développement de l'intelligence artificielle, car ce sont des questions d'ordre politique qui se posent. Avec le numérique, la notion d'État telle qu'elle a pu avoir cours dans le passé, et en particulier celle de souveraineté sur laquelle l'État de droit moderne se fonde, devient obsolète. Car l'État ne domine plus seul sur un territoire... Si les États ne s'emparent pas de ces sujets, de grands groupes vont avoir une emprise de plus en plus importante sur nos sociétés et risquent de se substituer, au moins partiellement, à eux.

Les enjeux de la souveraineté sont donc totalement à repenser. L'intelligence artificielle nous amène à redéfinir ces concepts, qui relèvent de la philosophie politique. C'est d'ailleurs ce travail de redéfinition qui m'a motivé au départ !

"L'intelligence artificielle nous amène à redéfinir des concepts qui relèvent de la philosophie politique."

Ce renversement est à rapprocher de celui induit par la généralisation de la **sousveillance**, notion introduite par le technologue canadien Steve Mann. Ce concept part du principe que, grâce aux smartphones, tout le monde peut accéder et diffuser de l'information sur Internet : dès lors, celui qui veille n'est pas nécessairement au-dessus, comme dans la surveillance, mais il peut-être en dessous et diffuser l'information mettant en cause ceux qui sont au-dessus de lui...

La société de sousveillance telle que nous la connaissons actuellement va à l'encontre de la société de surveillance, qui s'est mise en place à la fin du XVIII^e siècle, avec l'État de droit et dans laquelle la maîtrise de l'information, tout en étant réservée aux personnes qui disposaient du pouvoir, maintenait et même renforçait leur pouvoir.

Faut-il, alors, "réguler" l'intelligence artificielle ?

► Tout dépend de ce que l'on entend par réguler : si l'on considère que les machines vont s'autonomiser elles-mêmes, alors oui, il faut réguler. Mais comme je vous le disais, je n'y crois pas une seconde. En revanche, qu'il faille s'intéresser aux questions de sécurité qui sont liées au développement de l'intelligence artificielle, absolument. Si les agents doivent

adopter des comportements qui n'étaient pas prévus au départ, il faudrait prévoir une manière de contraindre leurs actions.

Il faut, selon moi, réfléchir aux valeurs humaines que l'on peut intégrer aux machines, afin que leurs actions ne soient pas uniquement décidées par des systèmes reprogrammables par apprentissage et, en conséquence, un peu imprévisibles. A cet égard, il existe un projet en cours, **le projet Ethicaa**, financé par l'Agence nationale de la recherche, qui cherche à insuffler une certaine éthique dans les agents autonomes. La recherche peut justement jouer un rôle fondamental dans l'arbitrage de cette question, et c'est d'autant plus intéressant lorsque l'on fonde ce respect des valeurs humaines sur un point de vue scientifique.

Ces questions d'éthiques surgissent lorsqu'il y a conflit entre des exigences contraires, voire contradictoires, qu'il faut surmonter, et cela passe par la création d'outils de supervision dédiés. À cet égard, je suis en profond désaccord avec de nombreux chercheurs américains qui considèrent que les valeurs morales découlent de l'apprentissage. Je considère pour ma part qu'il y a l'apprentissage d'un côté et les prescriptions de l'autre. Et que, clairement, la norme morale ne peut s'acquérir par l'apprentissage à partir d'une simple observation du réel. Il faut introduire les deux : des capacités d'adaptation au monde et des valeurs morales.

* Jean-Gabriel Ganascia, *Le mythe de la Singularité – Faut-il craindre l'intelligence artificielle ?*, **Seuil**, 2017

* En savoir plus sur le cycle de conférences **Aux Sources du Numérique**, organisé par Renaissance Numérique et Le Tank, dans lequel intervenait Jean-Gabriel Ganascia vendredi 21 avril 2017.

➤ à lire aussi : **Quelle gouvernance pour les robots et algorithmes ?**

➤ à lire aussi : **FIC 2017 : vers un (véritable) espace numérique européen ?**