

L'intelligence artificielle est-elle intelligente ? - Faut-il comprendre ?

mardi 21 janvier 2020, par [KRIVINE Hubert](#) (Date de rédaction antérieure : 30 septembre 2019).

Sur la base des progrès foudroyants de l'intelligence artificielle, ses soutiens inconditionnels prétendent que celle-ci va révolutionner les mécanismes de la connaissance et, à la limite, permettre de se passer de théorie. Nous dénonçons la naïveté de ce pragmatisme.

Sommaire

- [Comprendre pour prévoir](#)
- [Prévoir sans comprendre \(...\)](#)
- [A-t-on encore besoin de \(...\)](#)
- [Il n'y a pas d'intelligence](#)
- [Il n'y a pas de données brutes](#)
- [Peut-on numériser l'Univers ?](#)
- [La science ne progresse \(...\)](#)
- [En conclusion ?](#)

Le domaine de l'intelligence artificielle (IA) est un peu comme celui de l'Univers : il est en expansion accélérée et rempli de trous noirs... L'IA est une toute nouvelle discipline, comme l'a été l'imprimerie au milieu du XV^e siècle. Qui aurait pu alors anticiper que son développement irait bien au-delà de son but initial : la propagation des Saintes-Écritures ? Plus récemment encore, qui aurait pu imaginer les conséquences de l'internet, créé à l'origine pour faciliter les échanges entre physiciens de laboratoires éloignés ? Il faut donc être extrêmement prudent sur les potentialités de l'IA, surtout qu'à la différence de ces deux innovations, on ne sait pas pourquoi elle « marche » si bien. Enfin, l'appellation est bluffeuse : l'intelligence artificielle n'est pas à l'intelligence humaine ce qu'est l'insuline artificielle à l'insuline animale, c'est-à-dire, potentiellement au moins, la même chose en mieux.

Nous ne nous intéresserons pas ici aux succès et dangers de l'IA dans tous les domaines, mais essentiellement à ce qui nous semble être ses limites actuelles. Nous ignorons évidemment ce qui se passera dans les siècles à venir...

Comprendre pour prévoir

L'humanité a de tout temps eu besoin de prévoir. Elle l'a fait soit à l'aide de la magie (haruspice, augures et autres pythies), soit par l'observation de corrélations régulières. C'était la prévision du mouvement des astres, des marées, de l'action médicamenteuse de certaines plantes, des propriétés des alliages métalliques, des bienfaits du croisement de plantes et d'animaux, etc. D'où l'importance de la tradition dans les sociétés primitives.

À la Renaissance va se développer l'idée qu'il existe des lois impersonnelles et universelles qui

gouvernent le monde et que la tâche des savants est de les découvrir. Galilée qui affirmait que le « livre de l'Univers est écrit en langue mathématique » en est le précurseur le plus fameux. Dieu n'est certes pas abandonné, mais disons, relégué. Ces lois vont non seulement rendre compte des phénomènes observés, mais en prévoir de nouveaux. La théorie de la gravitation de Newton est emblématique : non seulement elle rend compte avec précision du mouvement elliptique des planètes, mais elle va prévoir le retour de la comète de Halley, la valeur de l'épaississement de la Terre à l'équateur et, un siècle et demi plus tard, la découverte de Neptune grâce aux calculs de Le Verrier. Les ondes radio seront découvertes vingt ans après avoir été prévues par les équations de Maxwell. La théorie générale de la relativité n'est pas fondée sur l'observation que la présence de masses courberait la trajectoire de la lumière, mais au contraire elle a prévu cette (minuscule) déviation qui sera effectivement mesurée quatre ans plus tard par Eddington. On pourrait multiplier les prévisions de phénomènes inouïs - au sens propre du terme - causées par la connaissance de ces lois.

Malheureusement, cette voie royale d'entendement qu'on pourrait résumer en comprendre pour prévoir, va s'obscurcir pour (au moins) deux raisons :

1) Même si on connaît les lois d'un phénomène, elles peuvent être si nombreuses et/ou si compliquées et intriquées que leur mise en œuvre est pratiquement impossible. On doit alors faire appel à des lois statistiques qui prévoiront seulement des moyennes.

2) Il peut se faire qu'une seule loi simple et bien connue gouverne un phénomène et qu'on soit néanmoins incapable d'anticiper au-delà d'un certain horizon de temps. C'est ce qui s'appelle le chaos déterministe. « Déterministe » parce qu'il y a une loi, « chaos » parce qu'on ne peut néanmoins rien en conclure à terme. Il est dû au fait que de toutes petites variations des conditions initiales ou de l'environnement peuvent générer une divergence exponentielle des solutions. Bien des phénomènes connaissent ce chaos : les trajectoires des molécules d'un gaz, le développement d'espèces en compétition, la météo, le mouvement des planètes du système solaire. Bien sûr cet horizon d'imprédictibilité dépend du système, il peut varier de la millionième de seconde pour le premier exemple aux milliards d'années pour le dernier.

Le chaos déterministe - bel oxymore - ne remet pas en question la causalité, mais interroge, y compris dans les sciences dites « dures », notre capacité à la mettre en évidence.

La conclusion générale est que la compréhension - et même la connaissance de la loi quand on les connaît - ne permet pas nécessairement de prévoir.

Prévoir sans comprendre (le retour ?)

Nos Anciens, sur la base de milliers (millions ?) d'observations, arrivaient à tirer certaines leçons. Mais maintenant nous disposons d'infiniment plus de données (en chiffres, images, sons, vidéos) ; ce sont les "Big Data". Il s'agit de quantités inimaginables : par exemple, chaque jour est généré 2,5 milliards de milliards d'octets [1] ! Ajoutons que l'accumulation d'informations est telle que 90% des données dans le monde ont été créées au cours des deux dernières années seulement. Il est hors de question que ces immenses bases de données soient directement lues par des hommes. Elles doivent être « intelligemment » stockées puis dépouillées par la machine. C'est l'un des objets de la mal nommée « intelligence artificielle ».

Très schématiquement, l'IA moderne se caractérise par l'apprentissage machine, c'est-à-dire que la machine, « instruite » par une base de données, extrapole de l'information sur des données nouvelles. Ces données d'apprentissage peuvent lui être fournies « étiquetées », c'est-à-dire par

exemple, sous forme de milliers de caractères manuscrits préalablement catalogués comme étant a, b, c... z ou de millions d'images d'animaux catalogués comme chats, chiens, tigres, etc. C'est l'apprentissage supervisé. On peut même ne pas étiqueter *a priori* les images, la machine se débrouillera pour faire des regroupements *ad hoc* et créer ainsi de nouvelles catégories ; c'est l'apprentissage non supervisé, plus gourmand en ressources, mais plus facile à mettre en œuvre (pas d'étiquetage qui est un processus long et complexe). Elle peut enfin se procurer elle-même les données d'apprentissage qu'elle testera *in situ* en se proposant des expériences, c'est l'apprentissage par renforcement. Un peu comme un enfant apprend à parler sans connaître la grammaire. C'est avec ce dernier mode d'apprentissage qu'a fonctionné la machine AlphaZero qui a battu le champion du monde de jeu de go, Ke Jie, en mai 2017 [2]. En trois jours elle a joué des millions de parties contre elle-même et a en quelque sorte « compris » comment jouer. Les guillemets sont importants : elle a peut-être compris, mais pas nous ! Personne ne sait expliquer le cheminement qui a permis cette victoire.

A-t-on encore besoin de démonstrations ?

À la différence de l'esprit humain, la machine qui utilise les réseaux de neurones artificiels n'a pas d'outil pour distinguer les corrélations causales des non causales ; et, *a fortiori*, elle ne fournit pas d'explications. Mais est-ce si grave ? Après tout, avec une base de données extrêmement limitée et sans réelle théorie, l'humanité de *l'homo sapiens* s'est très bien développée pendant plus de 150 000 ans. Ne pourrait-on pas penser à plus forte raison que, sur la seule gigantesque base moderne des big data correctement exploitée, elle pourra continuer et même infiniment mieux qu'auparavant ? De toute façon, des corrélations même non causales peuvent être prédictives : ce n'est pas la chute du baromètre qui cause la tempête ! Je cite la position extrémiste et hélas populaire d'un Chris Anderson. Le titre de son célèbre article [3] est éloquent : « La fin de la théorie : le déluge de données rend la méthode scientifique obsolète ». On peut y lire ceci : « Avec suffisamment de données, les nombres parlent d'eux-mêmes ». Et, plus loin : « La corrélation supprime la causalité, et la science peut avancer même sans modèle cohérent, sans théorie unifiée voire sans aucune explication mécaniste du tout. »

Pour lui, l'idée est que tous les modèles sont faux et souvent pollués d'idées préconçues alors que les bases de données, à condition d'être suffisamment gigantesques, ne sauraient mentir. Peut-on considérer que, dès 1960, Deng Xiaoping avait bien anticipé cette philosophie lorsqu'il affirmait : « Peu importe qu'un chat soit blanc ou noir, s'il attrape la souris, c'est un bon chat » ? Il s'agissait d'introduire plus de pragmatisme (au sens de plus de marché) dans l'économie, sans s'embarrasser des objections théoriques qu'on pouvait lui opposer. Il y a moins « extrémiste » dans l'abandon de la théorie [4] : « La science vit ainsi une révolution épistémologique avec la mise en œuvre depuis une dizaine d'années seulement d'un « quatrième paradigme » de la découverte scientifique, à partir de l'analyse et de l'exploitation intensive des données, sans nécessité a priori d'un modèle décrivant le réel. Cette révolution touche tous les secteurs scientifiques, tout particulièrement les domaines de la biologie-santé et des sciences humaines et sociales. »

Il n'y a pas d'intelligence, il n'y a que des preuves d'intelligence

Comment définir l'intelligence de la machine sans l'avoir fait déjà pour celle des humains [5] ? Turing [6] escamote habilement cette question en proposant seulement de les comparer par un test. Un expérimentateur dialogue via un clavier (ou même aujourd'hui par la voix) avec un interlocuteur caché. Si l'homme est le plus souvent incapable de savoir s'il a eu affaire à une machine ou pas, on dira que la machine a passé le test de Turing. Bien sûr, la durée du test est importante et

aujourd'hui aucune machine ne l'a emporté dans des temps raisonnables. Malgré (ou peut-être à cause de ?) sa très grande simplicité, le test de Turing est jugé peu intéressant par certains spécialistes de l'IA. On peut d'ailleurs penser que les machines réussiraient le test de Turing s'il ne comportait que des épreuves conventionnelles du type de celle utilisées pour déterminer le « Quotient intellectuel » des individus.

Essentiellement, l'intelligence de l'IA procède par induction. Ce qui veut dire que la machine ne peut prévoir que sur la base (gigantesque, certes) du déjà connu ou advenu. En caricaturant un peu, pour la machine, ce qui arrivera est déjà arrivé ou proche de l'être, mais sans les big data, on (les humains) ne l'aurait jamais décelé. Sauf que dans des situations politiques, financières et économiques inédites, les prévisions des big data échouent. Voir par exemple les « prévisions » de la crise des subprimes partie des États-Unis en 2007. Nate Silver [7] montre comment la singularité absolue du tableau économique des USA à cette époque rendait inopérante toute extrapolation.

Au fond, le « raisonnement » par induction suppose que quand un événement s'est répété n fois, il se répétera une $(n + 1)$ -ième fois, et ce d'autant plus sûrement que n est grand ; mais c'est sans considérer les conditions - éventuellement changeantes - qui ont permis cette continuité ; il faut une hypothèse implicite d'uniformité. Avec cette hypothèse, ni vous, ni vos parents ne mourront jamais ! (puisque vous constatez qu'ils ont vécu sans cesse tous les jours depuis 25 000 jours, ils vivront encore le jour suivant). Le raisonnement par induction, courant dans la vie quotidienne, peut donc suggérer une hypothèse ; il ne la démontre en aucun cas.

Il n'y a pas de données brutes

Il n'y a pas de données innocentes ; la notion de données brutes est un oxymore, a écrit très justement l'historienne des médias nord-américaine Lisa Gitelman. Les données sont des productions humaines qui peuvent être socialement ou techniquement biaisées - et pas nécessairement de façon volontaire.

On prend et on mélange les données là où on les trouve, comme un ivrogne cherche sa clé perdue la nuit seulement sous les lampadaires. Sont nombreux les exemples du danger d'analyse de données « brutes », sans réflexion sur leur production et dont l'augmentation en taille n'en diminuerait pas le faux sens.

En fait, l'IA ne fait que multiplier les dangers de biais inhérents à toutes les analyses « classiques ».

Peut-on numériser l'Univers ?

L'humain interagit avec toute la Nature (et ça fait du monde !), pas la machine qui n'en connaît qu'une petite partie, et de plus, digitalisée, c'est-à-dire, en dernière analyse, uniquement représentée par une succession - gigantesque, certes -, mais finie, de 0 et de 1. Mais la carte (numérique) n'est pas le terrain. Croire que la Nature suffisamment digitalisée est la Nature nous semble être une illusion totale, et ce quel que soit le degré de digitalisation. C'est pourtant le credo de quelques ayatollahs des big data. Extrapolant les succès spectaculaires de l'IA, ils imaginent qu'on pourra demain faire de la physique sans physicien, voire de la médecine sans médecin et pourquoi pas, des condamnations sans juge.

Illusion totale, vraiment ? Mais l'homme n'a pas non plus directement accès à toute la Nature. C'est seulement par ses sens qu'il interagit avec elle et il ne peut donc pas voir - ni sentir, ni toucher - tout le terrain. Ce qu'il en voit, par exemple, est pixélisé sur les 120 millions de cellules

photosensibles (cônes et bâtonnets) qui peuplent sa rétine. Or, aujourd'hui les photos digitalisées peuvent atteindre, voire dépasser cette finesse. Les capteurs artificiels n'ont donc rien à envier à nos capteurs naturels.

Mais le « terrain » ne se réduit pas seulement à l'image qu'en perçoit notre rétine. Il faut considérer tout ce qui le constitue avec sa géologie, son histoire, ses millions d'espèces vivantes et mortes, ses odeurs, son prix au mètre carré, sa beauté, la poésie qui lui est attachée, etc. D'une façon ou d'une autre le cerveau humain y est sensible même si on ne sait ni comment ni à quel degré ; voilà qui fait pratiquement une infinité d'éléments (en interaction). Croire que le « terrain » au plein sens du terme est pixélisable, c'est-à-dire représentable par une suite finie (même très grande) de 0 et de 1, semble aussi fou que croire que les êtres humains (et le reste) étant composés de molécules en interaction, on arrivera à l'explication de la prise de la Bastille en 1789 par l'étude (très !) approfondie des forces entre atomes. Ce serait ce qu'on peut appeler un réductionnisme échevelé.

On n'évoquera jamais trop ce titre d'article aussi lapidaire que profond de Philip Warren Anderson : « *More is different* ». Il faut changer de théorie quand on change d'échelle en temps, en taille ou en complexité. Que le tout soit autre chose que la somme de ses parties est bien connu. C'est ce qui caractérise le phénomène d'émergence.

La science ne progresse pas par accumulation de données

La science ne progresse pas par accumulation de données. Si la découverte du boson de Higgs ou des ondes gravitationnelles n'a pu avoir lieu qu'en manipulant les millions, voire les milliards de données de big data, ces découvertes sont dans leur genèse « classiques » : on savait ce qu'on cherchait. Des milliers de scientifiques et techniciens, à l'aide de centaines de millions de dollars, etc., ont dû imaginer des dispositifs diaboliquement astucieux pour mettre en évidence des effets extraordinairement faibles, qu'on aurait manqué si on ne les cherchait pas là où la théorie les prévoyait. C'est par exemple, pour les ondes gravitationnelles, une variation de longueur de moins d'un milliardième de milliardième de mètre d'un bras d'interféromètre de 3 km. Aucun big data n'aurait pu les trouver.

Ce n'est pas nouveau ; il en a été de même pour le neutrino. Prévu en 1930 par Wolfgang Pauli, c'est une particule neutre qui n'interagit pratiquement pas avec la matière et est donc très difficile à mettre en évidence (il faudra attendre 1956 !). On voit mal comment les big data, aussi big soient-elles, sans guide théorique de recherche, auraient permis ces découvertes.

Plus généralement, les avancées qui ont révolutionné la physique (et même aussi la philosophie), à savoir la théorie atomique, la mécanique quantique et la relativité, n'ont rien à voir avec une accumulation forcée de data. Einstein a élaboré la relativité restreinte sur la base des contradictions logiques internes aux équations de Maxwell (celles qui régissent les courants électriques et expliquent les ondes radio) et la relativité générale à cause des contradictions théoriques apparues au sein de la relativité restreinte. Newton n'avait pas vu plus de pommes tomber que ses prédécesseurs pour élaborer sa théorie de la gravitation.

Les données - bien sûr indispensables à la vérification de la théorie - ne viendront qu'ensuite. Ce sont ces théories permettant une nouvelle auscultation du ciel qui fourniront le point de départ à la fabrication de bien des big data et pas l'inverse. Par exemple, la théorie de la relativité (énoncée par Einstein en 1915) prévoit une courbure des rayons lumineux passant près d'une étoile massive (qui sera vérifiée en 1919 par Eddington). C'est le phénomène de lentille gravitationnelle, source nouvelle d'information sur la distribution des masses de l'univers. Il en va de même pour les ondes gravitationnelles qui vont encore enrichir nos big data.

En résumé, la science ne procède pas par accumulation et systématisation de données - même si c'est une étape qui peut être importante, mais par la résolution de problèmes [8]. Problèmes qui peuvent être internes à la théorie existante ou résulter des contradictions entre théorie et expériences (ou observations). C'est tout le problème ouvert de la créativité.

Il y a enfin une autre différence de taille : la machine est dédiée. Elle doit résoudre, tout chose égale par ailleurs, une tâche qu'on lui fixe. Elle vit dans un petit monde. Même si le nombre de parties de jeu de go est des milliards de fois supérieur au nombre total d'atomes de l'univers, la machine n'interagit que faiblement avec tout l'univers (elle répond seulement aux coups de l'adversaire). En ces sens, c'est un jeu « simple ». Un robot rat se débrouillera bien mieux qu'un vrai dans un labyrinthe, mais qu'arrive une odeur de brûlé, le vrai rat cherchera à fuir, pas le robot.

Le vrai rat possède en quelque sorte une culture, fruit d'un processus d'évolution darwinienne d'interactions avec le reste du monde qui aura duré des milliards d'années. Ce processus est foisonnant, c'est-à-dire sans but. Il est beaucoup plus lent qu'un processus piloté par un but, mais sur le long terme, il est beaucoup plus efficace. C'est lui qui est à l'origine du « sens commun », chose la plus difficile à acquérir - si tant est que ce soit possible un jour - par une machine [9].

En conclusion ?

Voici un extrait de l'interview d'Antoinette Rouvroy parue sous le titre « Mais pourquoi faudrait-il s'en inquiéter si l'on gagne en efficacité ? », dans *Le Monde* du 30 décembre 2017 : « Nous allons vers un changement épistémologique majeur. S'en remettre à ce type de calcul traduit un renoncement aux ambitions de la raison moderne qui liait les phénomènes à leurs causes. Ces ambitions de la raison permettaient d'envisager la prévention, d'agir sur les causes pour changer les effets. Au lieu de cela, on se dirige vers un système de pures corrélations.

On ne cherche plus à comprendre l'environnement, on cherche à le prédire. Notre rapport au savoir change, mais aussi notre rapport au monde : on se focalise davantage qu'auparavant sur les risques. Voir et comprendre sont supplantés par détecter et prévenir. On passe d'une civilisation du signe, qui était porteur de sens, à une civilisation du signal, qui est une donnée qui ne signifie rien en soi. »

Autrement dit : avec la science on tente d'agir sur le monde, avec l'IA fétichisée, c'est le monde qui agit sur nous, ce qui est bien différent ! Voilà qui redonne une bonne actualité à un Marx qui affirmait en substance qu'il ne fallait pas tant interpréter le monde, mais qu'il s'agissait de le transformer.

Hubert Krivine

P.-S.

• Créé le Lundi 30 septembre 2019, mise à jour Lundi 30 septembre 2019, 09:08 :
<https://npa2009.org/idees/culture/lintelligence-artificielle-est-elle-intelligente>

Notes

[1] Un octet, ce sont 8 chiffres binaires (bits), c'est-à-dire une suite de 8 zéros ou 1. Il permet de coder $2^8=256$ caractères, soit largement plus que tout un alphabet avec majuscules, accents et ponctuation.

[2] Ce qui semblait totalement hors de portée cinq ans auparavant...

[3] Consultable en ligne sur <https://www.wired.com/2008/06/pb-theory/>

[4] Jedan-Philippe Bourgoïn, Voyage au cœur du big data. Clefs CEA, 64, juin 2017.

[5] On connaît les nombreuses polémiques autour de la signification du QI (le quotient intellectuel).

[6] Alan Turing (1912-1954), génial mathématicien anglais qui a su casser le code secret allemand pendant la Seconde Guerre mondiale. Persécuté pour son homosexualité, il se suicida en croquant une pomme empoisonnée.

[7] The signal and the noise : The art of science and prediction, Penguin, 2012.

[8] Prophétie attribuée à Einstein : les machines un jour pourront résoudre tous les problèmes, mais jamais aucune d'entre elles ne pourra en poser un !

[9] Par exemple la modification minuscule de quelques pixels sur une photo de mouton broutant dans un pré peut le faire identifier par la machine comme une table. Ce que le sens commun interdirait évidemment !