L'IA et le sens commun

Jean-Michel BESNIER

Professeur de philosophie émérite, Sorbonne-Université

Résumé : L'intelligence artificielle (IA) serait incapable de concurrencer le sens commun. Telle est la bonne nouvelle qui garantirait à l'humain son avenir et révélerait la vanité de certaines spéculations transhumanistes. Cette annonce justifie qu'on décrive ici les ambitions portées par l'IA ainsi que les obstacles que celles-ci rencontrent. Mais la désillusion ne saurait être totale : remettre l'IA à sa juste place en pourrait aussi servir la cause.

Mots-clés : transhumanisme, singularité, modèle cognitif, révolution numérique

L'évidence s'impose chaque jour davantage : l'intelligence artificielle (IA) bouleverse notre accès aux connaissances et nous dicte de nouveaux comportements. Elle annoncerait même une mutation anthropologique capitale. Mais il n'est pas d'évidence qui ne puisse être examinée…

Il conviendrait avant tout de s'entendre sur une description minimale de l'IA. Elle est le produit d'une histoire qui a d'abord connu trois étapes, correspondant à trois modèles théoriques que je me contenterai de citer : Elle fut d'abord l'instrument de la résolution de problèmes (ex. les systèmes experts), grâce au traitement computationnel d'informations (modèle cognitiviste ou computationnaliste). Elle devint vite machine à apprendre en recourant à des dispositifs connexionnistes de reconnaissance de formes ou de sons (ex. les jeux vidéos, les machines à traduire) (modèle connexionniste). Elle se développe aujourd'hui sous la forme de « machine ou deep learning » en illustrant la puissance de mécanismes auto-régulés tels que la science des données les exploite à grande échelle (ex. l'analyse des gestes ou des émotions, l'identification de symptômes, l'anticipation d'évolutions à partir de grand nombre…) (modèle énactif).

L'IA a créé une situation épistémologique nouvelle : avec elle, l'algorithme tout puissant consacre le triomphe de la logique contre les mathématiques qui cherchent encore à expliquer les phénomènes grâce aux équations et à des modèles dynamiques, ainsi que l'argumentait Stéphane Mallat au Collège de France. Désormais, on ne connaît plus par les causes mais par les statistiques, on n'explique plus mais on corrèle des

données : ainsi a-t-on bel et bien arraché la connaissance aux préjugés philosophiques qui exigeaient la conscience et l'intentionnalité à la source de tout processus cognitif. L'IA et les algorithmes font triompher une conception pragmatique de la connaissance (au risque de disqualifier la connaissance théorique qui vise la compréhension plus que l'action).

Il est à présent permis de s'engager dans l'évaluation de l'impact de l'IA, héritière de la cybernétique et des sciences de la cognition, sur nos représentations du monde : dans la tradition issue de la Renaissance européenne, la connaissance était associée à une ambition humaniste résumée dans l'antienne cartésienne : se rendre comme maître et possesseur de la nature. Or, inféodée à l'IA, ce projet de connaissance annonce une perte de commande de l'humain sur son destin. De sorte que ce qui devait nous permettre d'être autonomes se soit rendu en quelque sorte autonome par rapport à nous. Ironie de la modernité qui se retourne contre elle-même ! Deux citations mériteraient d'être mises en exergue pour camper cette situation :

> « L'impact de la machine pensante sera un choc d'un ordre certainement comparable à celui de la bombe atomique… Si nous nous mettons à trahir l'homme, il deviendra un homme en colère, et un homme en colère est un homme dangereux ».
> *Norbert Wiener*

> « J'ai peur que l'IA puisse remplacer complètement les humains. Si les gens peuvent concevoir des virus informatiques, quelqu'un pourrait concevoir une IA qui peut s'améliorer et se reproduire. Ce serait une nouvelle forme de vie *capable de surpasser les humains »*
> *Stephen Hawking*

Nul doute que la préoccupation dont l'IA fait aujourd'hui l'objet, dans les cursus universitaires ou dans les médias, ne soit le signe d'une prise de conscience et d'une inquiétude. On découvre que la technique ne désigne plus seulement les outils que nous fabriquons et que nous pouvons utiliser ou non, mais qu'elle est devenue une culture qui nous immerge et conditionne nos manières de penser et d'agir. La technologie précipite une mutation qui bouleverse notre conception du monde et qui transforme nos attentes et nos idéaux. Le ressort de cette mutation est presque anodin : la révolution numérique à l'origine du triomphe actuel de l'IA procède d'un dispositif technique unique, d'une déconcertante simplicité (la conversion de données qualitatives en suites de 0 et de 1), et capable de s'appliquer à toute réalité (l'écrit, le son, l'image, le corps, les émotions…). Le projet de *mathesis universalis* rêvé par Descartes (la science universelle de l'ordre et de la mesure) ou par Leibniz (la caractéristique universelle) s'est réalisé concrètement. Le pouvoir nous paraît en effet acquis d'unifier l'ensemble de ce qui existe et d'agir sur lui, grâce à la maîtrise des paramètres permettant la conversion du réel en langage machine, grâce à la modélisation et la simulation en réalité virtuelle. L'ambition de réaliser un cerveau artificiel n'est qu'un avatar de la confiance placée dans ce pouvoir. L'IA est par là même devenue le nom donné à cette maîtrise du réel, grâce à l'application de procédures de conversion numérique qui en permettent la description, l'exploitation et la transformation.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

Devant l'OPECST (l'Office parlementaire d'évaluation des choix scientifiques et technologiques), le 7 décembre 2019, le mathématicien et député Cédric Villani présentait l'IA comme un ensemble de dispositifs permettant à des algorithmes de donner des réponses précises, selon trois types de méthode caractéristiques des trois étapes que j'ai décrites :

1. en mimant la décision humaine intelligente (*cf.* les systèmes experts avec des arbres de décision)
2. en mobilisant un apprentissage fonctionnant sur la base de la reconnaissance par corrélation (*cf.* les systèmes connexionnistes avec une logique d'essais-erreurs)
3. en appliquant des solutions après exploration et calcul sur un grand nombre de situations-problèmes (*cf.* Alphago et les systèmes convolutifs)

Cédric Villani résumait l'impact de l'IA en introduisant une réserve sur laquelle je voudrais m'arrêter : elle peut s'appliquer à presque tout, dit-il, par exemple reconnaître des situations, analyser des données, optimiser des paramètres, jouer…, mais elle n'est pas bonne pour prendre des décisions relevant du sens commun. Toute la question est à présent de déterminer ce que peuvent être ces décisions relevant du sens commun… Le plus simple serait de se dire que ce sont celles qui exigent de l'intuition plus que du calcul. De l'intuition, c'est-à-dire de la réaction spontanée, non formalisable. L'IA trouverait donc sa limite dans son incapacité à gérer le caractère informel des comportements humains qui sont ajustés à des situations sans le vouloir expressément, sans le préméditer, sans même l'anticiper. Or, c'est pourtant de ce genre de comportements dont l'IA veut nous débarrasser : pour conjurer l'arbitraire et la subjectivité, pour interdire le caprice et le hasard, pour déjouer l'ambiguïté et l'opacité. Le rêve de l'IA n'est-il pas, en effet, de transformer toute activité en une succession d'automatismes que pourront réaliser des machines dépourvues d'intention et de motivation ?

De fait, l'histoire de l'IA depuis la naissance de la cybernétique, dans les années 1940-1950, est associée à l'ambition de réaliser un organisme artificiel qui imiterait l'intelligence humaine, en réagissant de manière adaptative à des stimuli (ou des inputs) de plus en plus élaborés. Grâce à des boucles de rétroaction qui imitent les phénomènes d'homéostasie de l'organisme vivant (l'effet rétroagit sur la cause qui modifiera l'effet, pour préserver l'équilibre ou le couplage avec l'environnement). Derrière cette approche de l'intelligence comme la description d'un mécanisme capable de recevoir des informations et d'émettre en retour des réactions susceptibles de modifier ces informations en un sens favorable à un ajustement toujours meilleur (c'est ici la description de l'objet prétendument intelligent : il capte une information et y répond par un comportement adéquat), il y a un parti-pris assumé par le positivisme et le behaviorisme : l'intelligence est de l'ordre de l'observation et de l'expérimentation, mais pas d'une supposée vie intérieure ou d'une subjectivité capable de délibérer avant de prendre les bonnes décisions. L'intelligence n'a pas besoin de l'hypothèse d'une intentionnalité ou même de celle d'une conscience. C'est pourquoi elle peut habiter le végétal, l'animal ou la machine. Le test de Turing, on le sait, traduit ce parti-pris :

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

si un robot est capable de tromper un humain, il peut être dit intelligent au même titre que cet humain lui-même. Ce test de Turing est naturellement susceptible de s'appliquer aussi aux comportements émotionnels : si un robot est capable de simuler des émotions, rien n'empêche de le considérer comme capable d'éprouver ces émotions. Au même titre que n'importe quel humain dont on ne peut dire s'il éprouve ou non ce qu'il donne à voir (les signaux traduisant la colère, la peine ou la joie… *Cf.* Le paradoxe du comédien de Diderot). L'informatique émotionnelle est un rejeton de l'IA : elle se demande ce qu'est une émotion et si on peut la modéliser pour permettre qu'elle soit simulée par un robot. Dans la mesure où on la réduit à la production et la réception de signaux, on ne voit pas pourquoi elle ne relèverait pas du calcul et de l'algorithmicité auxquels on réduit toute information dans le contexte des procédures de numérisation.

Cette conclusion très pragmatique (puisqu'elle permet d'opérer sur ce qui apparaît complexe du point de vue de la psychologie introspective qui postule une vie intérieure s'exprimant par des conduites et des attitudes) est une objection aux SHS (sciences humaines et sociales) et elle se trouve par exemple au cœur de l'opposition entre la psychanalyse et les neurosciences associées à la psychologie cognitive (*cf.* le psychanalyste Gérard Pommier vs le cogniticien Stanislas Dehaene). Enjeu du débat : les comportements cognitifs ou émotifs sont ou ne sont pas le simple résultat d'une activation ou inhibition neuronale, qui établit des circuits synaptiques susceptibles d'évoluer au gré de la modification de l'environnement dans lequel ils s'expriment… La modélisation des neurones qui fonctionnent de manière binaire, avec des modulations et des poids synaptiques variables, se prête naturellement à la traduction numérique, comme l'avait vu John von Neuman (dans son livre paru en 1957, *Le cerveau et l'ordinateur*, qui distinguait quand même le langage analytique binaire de l'ordinateur, du langage statistique et modulable du cerveau). On voit en tous cas que les prétentions de l'IA à imiter l'humain se construisent sur la base d'une réduction de la complexité que supposaient la psychologie introspective (comment faire de soi-même un objet de connaissance ?) et par la suite la psychologie des profondeurs (autrement dit : la psychanalyse). C'est sans doute ce que voulait pointer Cédric Villani, en invoquant les décisions du sens commun comme le résultat de démarches qui se laissent mal formaliser de manière algorithmique.

La référence au sens commun nous conduit à observer ceci : le recours au qualificatif d'intelligent pour désigner une voiture, une ville, un capteur biologique, une spéculation boursière… relève d'une déflation sémantique, un peu à la manière dont procède la novlangue selon Orwell : on ne définit plus l'intelligence que comme la réponse adaptative d'un dispositif capable d'enregistrer des stimuli, et à partir de là, on en infère que n'importe quel objet est susceptible d'en relever ! L'étrange est que l'on s'habitue très bien à faire cet usage minimal d'un concept qui était réputé compliqué. La même chose est dite de la conscience chez les roboticiens ou les cognitivistes qui la limitent à l'aptitude pour un dispositif à faire la distinction entre soi et non-soi. À partir de cette approche sommaire, n'importe quel robot peut être désigné comme « conscient » et l'IA pourra espérer produire des objets conscients de soi – ce que l'on désigne par l'IA forte, celle qui nous rendra superflus, comme le prédisent les théoriciens de la Singularité !

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

Mis en perspective, le bouleversement provoqué par la révolution numérique et l'IA, pose à l'humain la question de sa destination. Le paléoanthropologue (et le philosophe platonicien) était convaincu que l'hominisation avait tenu à deux facteurs en coévolution au fil du temps, à savoir : l'outil et le langage, la technique et l'art politique… La question s'est posée au cours de l'histoire : la technique et le langage peuvent-ils demeurer en équilibre ? Longtemps, on s'est inquiétés du trop grand privilège accordé au langage par rapport à la technique (*cf.* L'Encyclopédie de Diderot et d'Alembert ou le thème « des deux cultures » formulé en 1959 par Charles Percy Snow). Aujourd'hui, cette inquiétude s'est inversée. Elle conduit à se demander si l'outil n'a pas pris le pas sur le langage. Certains transhumanistes le disent clairement, le cybernéticien Kevin Warwick, par exemple : la technologie numérique, celle qui permet la réalisation de dispositifs d'émission-réception de signaux et la création d'interfaces cerveau-machine, devrait nous permettre d'en finir avec le langage qui est entaché de tous les vices : ambiguité, polysémie, dissimulation, mensonge… Le langage serait la cause de tous les malheurs de l'humanité, mais nous allons bientôt pouvoir nous en passer et installer une communication télépathique qui permettra la transparence et donc, une sécurité universelle. Pour un philosophe ou un spécialiste de sciences humaines, cette idée est non seulement absurde mais également ruineuse : l'évacuation du signe et du dialogue au profit du signal et de son traitement automatique, revient à prononcer la mort de l'humain. Du point de vue anthropologique, le recours hégémonique au numérique et à l'IA signale la défaite de l'espèce humaine et sa relève programmée par les machines (deep learning, systèmes autonomes auto-régulés, robots androïdes ou non…).

Si l'on résiste au caractère apocalyptique du propos, on peut s'en remettre au sens commun qui n'a pas besoin de la leçon des paléoanthropologues pour étayer ses réserves :

• Par exemple, en rejetant la croyance selon laquelle l'humain serait devenu « le maillon faible » dans les sociétés technologisées. Une vieille idée (émise par Galton au début du vingtième siècle) qui prendrait une réelle consistance avec l'annonce de la disparition d'emplois même qualifiés (voir un rapport de l'OCDE, en 2018 : seuls 13% des emplois résistent encore à l'autonomisation). Mais le rôle joué par les micro-tâches de l'intelligence humaine pour faire fonctionner les plateformes numériques est de plus en plus mis à jour (*cf.* le livre d'Antonio A. Casilli, « En attendant les robots », Le Seuil 2019).

• Autre exemple, le sentiment qu'on ne fait plus droit au dialogue dès lors qu'il y a des machines pour produire les résultats escomptés, comme si le langage ne servait qu'à produire des effets utiles – ce sentiment suscite une résistance salutaire, notamment dans le domaine médical : on doute que la santé puisse s'établir sur fond d'un escamotage de la relation clinique (*ie.* dialoguée) entre le médecin et son patient. On répugne à une approche vétérinaire de l'humain, celle d'une médecine s'imposant de n'assurer que le fonctionnement des métabolismes et de ne plus considérer l'organisme autrement que sous l'angle algorithmique.

• Dernier exemple : l'ensemble des activités de services est désormais numérisé, digitalisé, livré aux algorithmes, à la prospection et à l'exploitation des données. « Le

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

syndrome de la touche étoile », les agents conversationnels sont les symboles de cette déshumanisation soft qui bénéficie souvent d'une démission de notre part, d'une véritable servitude volontaire (voir mon livre L'homme simplifié, éd. Fayard 2012). Le sens commun réclame qu'on en revienne aux échanges informels, qu'on s'oppose aux injonctions dominantes : il faut aller vite, être réactifs, affronter la concurrence… Le sens commun revendique pour les humains le privilège de n'être pas réduits à la condition de simples supports de données, destinés à accroître la nourriture de machines qui deviendront de plus en plus performantes, au fur et à mesure qu'elles engrangeront plus de données…).

Est-il nécessaire d'être complétement critique devant l'algorithmicité croissante de la vie quotidienne ? Non, on peut certes se féliciter de bien des progrès, en particulier dans le domaine de la médecine (le cheval de Troie du transhumanisme qui rêve de tuer la mort) ou bien dans celui de l'accès aux informations destinées à former des connaissances (Internet sert la cause de la recherche scientifique). Mais il faut éviter de se payer d'illusions et tâcher d'évaluer les transformations que nous laissons se développer dans nos organismes et dans nos têtes (*cf.* le livre de Nicolas Carr : *Internet nous rend-il stupides ?* Voir aussi la vogue du concept de « stupidité fonctionnelle » dans les entreprises : être réactifs, plutôt que réflexifs…). Parmi les illusions à dissiper, celle consistant à s'imaginer que la révolution numérique devrait être acceptée car elle serait l'équivalent de la révolution de l'écrit qu'a consacrée l'imprimerie de Gutenberg, c'est-à-dire un facteur d'émancipation. Michel Serres qui alimentait cette comparaison dans La petite poucette, n'était peut-être pas aussi clairvoyant qu'on le dit. La présente révolution qui l'enthousiasmait est volontiers argumentée avec la thèse selon laquelle nous serions en train de promouvoir un « humanisme numérique ». Mais le malentendu doit être levé sur ces « humanités numériques » qui prétendent relayer l'humanisme issu de la Renaissance. Le philosophe Sloterdjick posait la question en 1999 : quel humanisme est encore possible, après les critiques qui lui furent adressées au XX^{e} siècles (sommairement résumées, ces critiques consistaient à dénoncer un prométhéisme débridé ayant conduit au fantasme d'un homme nouveau et à une société prétendument intégralement rationalisée, ayant justifié aussi une vision de l'humain qui le réduit au biologique…). Pour Sloterdjick, l'humanisme issu du monde gréco-romain était d'emblée fondé sur l'écrit (l'échange de lettres puis de livres entre amis). Or, constatait-il, l'écrit n'est plus le bon support. C'est sans doute sur le numérique qu'il va falloir miser pour réactiver les vertus de communication universelle que portait l'humanisme. La perspective était tentante et conduisit Sloterdjick à invoquer un posthumanisme. Mais la voie m'a semblé incertaine : quelles limites révèlent le point de vue de Sloterdjick (et celui, également, de Serres) ? Celles de ne pas vouloir distinguer entre les signes et les signaux. Ni Sloterdjick ni Serres n'accordent d'importance au dialogue de soi avec soi par lequel on définissait jadis la vie intérieure ; tous deux paraissent s'accommoder de l'euphorie techniciste qui s'emploie à nous réduire à l'état de boites noires dotées du pouvoir d'interagir avec des objets dits intelligents. Ce qu'ils écartent est justement ce qui nous distingue des dieux, des animaux et des machines : la fonction symbolique – celle qui nous dispose à recourir aux vertus des mots pour mettre à distance le réel, celle qui nous invite à recourir à l'intelligence

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928

pour surmonter les automatismes de l'instinct, celle qui nous arrache aux impératifs utilitaires auxquels on voudrait nous soumettre.

À l'heure où nous réalisons que le monde est loin d'être homogène, que les références ne sont pas universellement partagées, que des fractures se multiplient dans les sociétés technologisées (les mouvements de contestation tous azimuts en témoignent), il est important de s'attacher au sens commun : la surenchère technologique, l'application sans mesure du digital (par ex. celle qui annonce les grandes surfaces commerciales sans personnel) ne sont peut-être pas sans dommages. Tout ce qui est techniquement réalisable ne doit pas être forcément réalisé. Ainsi l'éducation, sur laquelle chacun est tenté de miser, ne serait confiée à l'IA qu'au prix de l'élimination de la dimension socialisante des apprentissages scolaires, et donc des éléments de structuration psychique qui ont besoin du tiers (et non pas du binaire) et qui profitent du décentrement par les savoirs.

Published in : DELTORN, Jean-Marc ; PICHENOT, Evelyne (dir.), *Algorithmes et Société*, Editions des archives contemporaines, ISBN : 9782813003928, doi : 10.17184/eac.9782813003928