

L'IA : maître ou esclave ?

En masquant la dimension exclusivement calculatoire de l'IA, la projection anthropomorphique conduit à lui déléguer des responsabilités qui nous reviennent.

Publié le 22-01-2025 à 16h18

Enregistrer



Sommet de l'ingéniosité humaine, l'intelligence artificielle se diffuse partout. ©Copyright (c) 2024

Une opinion de Guillaume von der Weid, philosophe en éthique de la santé

Sommet de l'ingéniosité humaine, l'intelligence artificielle se diffuse partout. Son omnipotence est telle qu'elle semble devoir dépasser non seulement nos compétences, mais notre compréhension. D'où l'inquiétude qu'elle nous échappe et se retourne contre nous. N'est-elle pas un outil, certes éminemment perfectionné, mais dont le résultat dépend de l'usage qu'on en fait ? Ou bien une puissance autonome qui finira par nous asservir ? Deux caricatures symétriques qui renvoient, pour la première, à une erreur quant à sa nature, pour la seconde, à une faute quant à nos responsabilités.

Le périmètre de l'intelligence

Erreur quant à sa nature d'abord. Ultime externalisation de la puissance, l'IA a sorti l'intelligence de notre cerveau, comme l'organe a sorti l'action du végétal pour devenir une sorte de "sur-plante" (la main fait mieux que la branche), comme l'outil est devenu un "sur-organe" (le marteau fait mieux que la main) et la machine un "sur-corps" (la pelleteuse fait mieux que l'ouvrier). L'IA serait-elle la dernière étape du "sur-homme" ?

Ce serait prendre la partie pour le tout. Car l'IA ne fait qu'autonomiser la partie calculante de l'intelligence humaine. Importée de l'anglais, l'expression même d'"IA" nous pousse à mésinterpréter le terme d'*intelligence* où elle signifie "recueil de données" plus que réflexion humaine, comme en témoigne l'expression "*Central Intelligence Agency*" (C. I. A.). Or le calcul rationnel ne représente qu'un tiers de l'intelligence, celle qui dispose des meilleurs moyens en vue d'une fin, mais est incapable de les évaluer moralement et, encore moins, d'en déterminer le but.

➔ **À lire aussi**
Face à l'IA : calculer moins et penser plus

Cette erreur de périmètre vient de la surpuissance même de l'IA. Matrice algorithmique qui infinitise notre puissance rationnelle, elle structure des données massives pour détecter des régularités et en tirer des règles, puis des règles de règles dans une architecture logique tantôt renforçante, tantôt correctrice, qui finit par accomplir des analyses pénétrantes et des tâches complexes. Or cette puissance fascine. Quand, en 1997, *Deep Blue* battit Gary Kasparov, champion du monde d'échecs, celui-ci se sentit dépouillé d'une compétence spécifiquement humaine. Lee Sedol, qui perdit au jeu de go en 2016, fit au contraire remarquer qu'*AlphaGo* l'avait moins "battu" qu'"écrasé". L'IA ne joue pas – édifiant un plan de bataille avec une attaque et une défense, de coups de bluff et des positions de repli –, elle choisit des embranchements et, à force d'arborescences, finit par pulvériser une pensée humaine qui, elle, s'appuie sur l'indétermination du possible pour exprimer une stratégie. L'idée que les champions auraient "perdu" revient à projeter sur l'IA une pensée qui ne s'y trouve pas. On commet la même erreur en attribuant aujourd'hui à ChatGPT les idées des textes qu'il produit par calcul de probabilités, alors qu'elles appartiennent aux humains qui ont écrit les myriades de textes qu'elle synthétise sans comprendre.

Calcul et valeurs

Or, en masquant la dimension exclusivement calculatoire de l'IA, cette projection anthropomorphique conduit à lui déléguer des responsabilités qui nous reviennent. D'abord, l'auto-développement de l'"apprentissage profond" pourrait devenir le meilleur exemple de la théorie de la contre-productivité chère à Ivan Illitch, qui affirme que toute technologie tend à franchir un seuil au-delà duquel elle accomplit le contraire de ce qu'elle était initialement destinée à faire: trop de voitures immobilisent (embouteillages), trop d'école abrutit (conformisme), trop de médecine rend malade (modes de vie malsains, résistance aux antibiotiques, maladies nosocomiales, etc.). Trop d'algorithmes diminuent l'efficacité globale de l'IA, comme on le voit avec des réseaux sociaux qui, par valorisation automatique des publications les plus engageantes, finissent par promouvoir "fake news" et discours haineux *antisociaux*. Pour être efficace et précisément parce qu'elle se développe toute seule, la puissance de l'IA doit donc être limitée.



L'intelligence humaine ne détermine pas seulement les moyens d'agir et les principes d'action, mais encore les buts de la vie, ces idéaux lointains qui donnent sens à notre existence.

Mais pour se limiter, il lui faut des principes extérieurs à elles, fondés sur des valeurs. Comment la voiture autonome, par exemple, pourrait-elle décider, en cas d'accident, qui sacrifier entre un groupe de personnes se tenant sur la route, ou son conducteur en fonçant dans un arbre pour les éviter ? De même, selon quel critère de choix attribuer un organe : l'espérance de vie du patient après la greffe, l'urgence du besoin ou l'égalité dignité des candidats ? L'IA calculera les moyens les plus efficaces pour accomplir les principes que nous lui aurons fixés, mais sera incapable de trancher ces dilemmes. Les merveilles de l'IA risquent d'écraser le jeu de la délibération en prétendant calculer le meilleur choix, alors même que *le recours au calcul* est en réalité le choix principal.

Donner sens à l'impuissance

Enfin, l'intelligence humaine ne détermine pas seulement les moyens d'agir et les principes d'action, mais encore les buts de la vie, ces idéaux lointains qui donnent sens à notre existence. C'est ce qu'on voit bien en médecine. Quand on est malade, qu'on souffre, qu'on sent qu'on va mourir, ce sont eux qui nous aident à tenir. Or combien de malades, malgré tous les efforts de la médecine moderne, même optimisée par l'IA, souffrent, vieillissent, et meurent ? La réponse est simple : 100 %. C'est pourquoi, comme nous le rappelle le serment d'Hippocrate, la médecine ne tient pas qu'à la guérison, mais doit *prendre soin*. Non seulement traiter une maladie mais soigner un malade dans sa singularité, ses attentes et ses peurs. C'est le rôle de la compassion et de l'accompagnement. C'est le rôle ultime des soins palliatifs, de "tout ce qui reste à faire quand il n'y a plus rien à faire". C'est dire s'il ne s'agit plus d'efficacité. L'IA est impuissante à donner sens à notre impuissance.

➔ **À lire aussi**
Apprendre à penser face à ChatGPT

Loin d'être un surhomme, l'IA fait ainsi plutôt penser au Puntila de Bertold Brecht qui, esclave de l'alcool et bipolaire, infect à jeûn mais fraternisant dans l'ivresse, affirme à son valet Matti, brandissant sa fourchette dans une effusion humaniste : "*Tu vois, parfois, quand j'ai des accès de sobriété, au lieu de voir deux fourchettes, je n'en vois qu'une... je ne vois que la moitié du monde !*" Évitions qu'une surdose d'algorithmes nous fasse oublier la potion révélatrice des deux autres tiers du monde – qu'on appelle l'humanité.