

IA : Peut-on domestiquer l'infini ?

#ia

#philosophie

Image d'illustration © Shutterstock



Guillaume Von der Weid

- 17 mai 2024

CONTRIBUTION LECTEUR // L'IA dispose certes d'une puissance de calcul largement supérieure à celle que le cerveau humain est capable, mais elle reste dépendante de lui pour se voir indiquer les valeurs lui permettant de préserver nos relations humaines et de respecter nos principes moraux. C'est en tout cas la conviction de Guillaume Von der Weid, philosophe et fidèle lecteur d'*Usbek & Rica*.

L'IA avance à pas de géant. Dans le commerce, le transport, la finance, la justice, l'éducation, la santé, le divertissement... elle semble faire aussi bien, sinon mieux, que nous. Après avoir externalisé la force par l'outil, l'action par la machine, nous avons sorti l'intelligence de notre crâne, à cette différence près qu'elle n'est plus seulement un instrument à notre service, mais la capacité à choisir les instruments, plus seulement un gouvernail par lequel nous nous dirigeons, mais un gouvernement nous indiquant où aller. Elle contient en effet tous les éléments d'une décision éclairée : l'information, les principes de choix et le but à atteindre.

L'IA représente ainsi une réserve infinie de puissance, non plus simplement comme une assistance asymptotique à nos manières de faire, mais comme une créativité sans limite. Elle semble à la fois connaître le monde et savoir quoi en faire. Elle dispose du navire des méga-données, du gouvernail des algorithmes et de la destination de l'optimisation. Doit-on se soumettre à l'infini numérique ?

1/3 de l'intelligence humaine

Aussi intelligente soit-elle, l'IA n'est qu'une technique qui dépend de l'usage qu'on en fait. Rappelons d'abord que son « intelligence » renvoie au sens anglo-saxon du mot, signifiant *recueil des données* (d'où la « Central Intelligence Agency »). Or cette intelligence d'identification, de tri et de calcul, qu'on pourrait appeler *rationnelle*, ne représente qu'un tiers de l'intelligence humaine, qui est aussi capable de poser des buts (intelligence finaliste) et de choisir des valeurs (intelligence morale).

Au contraire du calcul qui donne une impression de maîtrise, les relations sont risquées



Ce tiers ne saurait être tout-puissant sans manquer son but. Risque qu'Ivan Illitch a nommé la « contre-productivité » : toute technique qui vise à satisfaire un besoin se retourne contre son but initial si elle dépasse un certain seuil : trop de voitures immobilisent (embouteillage), trop de médecine tue (maladies iatrogènes, résistance aux antibiotiques, modes de vie malsains), trop d'école abêtit (passivité, ennui, conformisme) et... trop de réseaux sociaux désocialisent. De même, la puissance même de l'IA ne doit pas entraver son but réel.

Quel est ce but ? Ce n'est pas la richesse, qui n'est qu'un moyen (Midas, qui est mort de faim après avoir vu son vœu de transformer en or tout ce qu'il touchait exaucé, l'avait oublié). Ce n'est pas non plus la satisfaction des besoins, qui se passe très bien de la richesse, mais la reconnaissance qu'elle confère, c'est-à-dire, en somme, *l'amour des autres*. Le but, ce n'est pas le rapport optimisateur, mais les relations entre des gens *tels qu'ils sont*. Seulement, au contraire du calcul qui donne une impression de maîtrise, les relations sont risquées. Leur construction prend du temps et peuvent être décevantes, voire nuisibles. C'est pourquoi le rationnel tend à empiéter sur le relationnel.

Kasparov décrit comment, à la suite de sa [défaite face au logiciel Deep Blue](#), l'IA a changé la nature du jeu d'échec. Alors qu'une partie était une histoire faite de stratégies, de bluffs, de faiblesses à combler, de postures agressives ou de défenses étouffantes, les joueurs ont poussé les joueurs à penser comme elles et à apprendre par cœur des enchaînements de positions, indépendamment du sens de la partie. Le calculable tend à assécher l'historique. Ce qu'on gagne en efficacité, on le perd en intérêt. Mais encore une fois, l'IA est devenue si forte, écrasant à tout coup les meilleurs joueurs, qu'on s'en est désintéressé, pour recommencer à admirer des champions comme Magnus Carlsen, qui est pourtant moins fort que la première appli tournant sur votre téléphone. C'est qu'on n'admire pas une performance, mais une *valeur*.

Mesurer la valeur immuable

Ce qui fait signe vers une troisième dimension. Pas seulement celles d'un calcul qui permet de voir ou de relations qui permettent de vivre, mais celle *des valeurs* dont les relations dépendent. Car l'intérêt qui nous réunit peut conduire à l'abaissement instrumental du cynisme, qui utilise les autres pour générer une valeur strictement économique, comme à l'élévation des valeurs morales, qui considèrent autrui comme une finalité indépassable. Or l'IA calculera toujours comment réduire les coûts ou maximiser les gains, y compris en faisant semblant de se détacher du calcul, notamment par sa capacité à mimer l'empathie. Aussi apprenante et ajustée soit-elle, elle ne pourra « ressentir » ou « intuitionner » des principes qui peuvent conduire à sacrifier des vies au nom de la justice, à choisir un état sous-optimal au nom d'un idéal politique, à trancher des dilemmes moraux dont toutes les issues sont mauvaises. Ainsi, l'IA doit être encadrée par des principes qui ne peuvent lui venir que *de l'extérieur*, puisqu'ils réduisent son périmètre d'évaluation par conversion de ses valeurs profitables en valeurs immuables.

L'IA ne doit jamais être aux commandes. Certes, sa puissance de calcul la rend infiniment meilleure que nous dans potentiellement tous les domaines. Mais cette polyvalence reste unidimensionnelle. Pour s'intégrer à des activités qui visent non pas seulement l'amélioration quantitative, mais aussi le sens et les valeurs, elle réclame d'être triplement domestiquée : pour rester dans sa zone d'efficacité, préserver nos relations humaines, et respecter les principes moraux qui nous rendent *dignes* d'intérêt.



Guillaume Von der Weid

- 17 mai 2024

