

DANIEL ANDLER

Professeur émérite de Sorbonne Université, membre de l'Académie des sciences morales et politiques, philosophe

Patrick Nicolet, directeur général de Linebreak SA, ancien directeur des technologies de Capgemini

Sans plus attendre, Daniel, merci de planter le décor.

Daniel Andler, professeur émérite de Sorbonne Université, membre de l'Académie des sciences morales et politiques, philosophe

Je viens de publier un livre de 400 pages sur l'IA, et les gens me demandent parfois comment j'ai pu l'écrire si vite, compte tenu du fait que l'IA existe depuis peu – un an ou peut-être un peu plus. En fait, l'IA n'est pas née avec ChatGPT. L'idée de base existe sous une forme rudimentaire depuis au moins deux siècles si l'on remonte à Charles Babbage et Lady Lovelace, et même plus si l'on remonte à Jacquard et Pascal, Hobbes et Leibniz. Sous sa forme moderne, l'IA a été lancée par Alan Turing en 1950 et baptisée en 1956 lors d'une réunion à Dartmouth, aux États-Unis.

Comment un peu d'histoire peut aider à comprendre la situation actuelle ? Premièrement, cela dissipe l'idée selon laquelle les systèmes d'IA actuels sont sortis de nulle part, le résultat d'une révélation qui a changé du jour au lendemain le destin de l'humanité. C'est plutôt le résultat d'un processus long et sinueux, qui s'est heurté à des difficultés qui l'ont contraint à abandonner ses hypothèses initiales et à repenser de façon radicale : au lieu de considérer les processus mentaux comme un type de logique, on a commencé à les considérer comme un type de perception. Au lieu d'essayer d'imiter le type de pensée que nous concevons consciemment, l'IA a visé le type d'informations que les neurones peuvent traiter, informations auxquelles nous n'avons pas d'accès direct. Au lieu d'essayer de transformer directement une architecture de von Neumann en une machine à penser, elle a choisi d'éduquer les réseaux neuronaux.

Une autre raison de se souvenir de la naissance de l'IA est le nom qu'elle s'est donné, qui masquait une ambiguïté : visait-elle l'intelligence ou autre chose ? Dès le premier jour, il y a eu deux projets derrière le projet. L'un consistait à créer un système informatique qui penserait comme les humains et serait intelligent dans le sens où les humains sont intelligents. L'autre était de trouver des moyens d'automatiser les solutions à autant de types de problèmes que possible, depuis le jeu d'échecs à la traduction, depuis la reconnaissance de formes à la navigation des robots... À première vue, ce sont des objectifs distincts. Pourtant, l'idée fondamentale était que la réflexion n'est en réalité rien d'autre que la capacité à résoudre des problèmes. Un système entièrement intelligent serait capable de résoudre toutes sortes de problèmes et, inversement, plus un système pourrait résoudre de problèmes,

plus il se rapprocherait de l'intelligence totale. Ainsi, l'IA a entrepris d'automatiser un problème après l'autre. Cela s'est avéré plus difficile que prévu. Les systèmes d'IA ne pouvaient pas comprendre les choses à partir de rien : ils avaient besoin d'informations riches, trop riches pour être transmises à la petite cuillère aux ordinateurs de von Neumann par des programmeurs humains. On s'est donc tourné vers des réseaux neuronaux capables d'apprendre par eux-mêmes à partir d'exemples. Après un démarrage lent, les réseaux neuronaux ont rencontré un succès retentissant.

Mais voici le problème : les systèmes construits par l'IA, qu'il s'agisse de raisonneurs à l'ancienne ou de percepteurs de la nouvelle vague, étaient des solutions à des problèmes spécifiques, une population d'algorithmes spécialisés qui ne ressemblaient en rien à l'intelligence humaine. Il semblait que l'un des deux objectifs que l'IA s'était fixés au départ avait été abandonné. L'opinion majoritaire chez les professionnels considérait comme une réalité qu'il y avait suffisamment de problèmes ou de tâches qui attendaient d'être automatisés, ou d'être automatisés plus efficacement, pour occuper les ingénieurs en IA ; et elle continue de le penser.

Mais le rêve d'une machine qui serait véritablement intelligente, une véritable machine pensante, qui posséderait une « intelligence générale artificielle » ou AGI, ou encore une « intelligence de niveau humain », est de retour. L'avènement des LLM, de l'IA générative, a fait pencher la balance : la capacité à composer à la demande des textes et des images cohérents et souvent pertinents de toute nature et sur n'importe quel sujet n'est pas seulement, comme chacun s'est vite rendu compte, un véritable changement de donne en termes d'applications dans d'innombrables domaines ; cela rend également plus plausible que l'AGI soit à portée de main dans quelques années seulement.

Mais cela repose sur deux hypothèses peu plausibles. La première est que la tendance actuelle à la réussite est vouée à se poursuivre jusqu'à ce que l'ensemble du répertoire des types de problèmes que l'esprit humain peut résoudre ait été conquis par l'IA. La deuxième hypothèse est qu'une fois que cela se produira, l'intelligence de niveau humain aura été atteinte.

Quant à la première hypothèse, la moins improbable, il y a deux raisons d'être prudent : premièrement, les systèmes spectaculaires actuels sont loin d'être parfaits et loin d'être entièrement compris. Ils constituent une base trop fragile pour prédire le succès à venir. Le deuxième problème est que même si les succès actuels laissent présager de nouveaux progrès, ils ne confortent pas l'idée selon laquelle des problèmes de toutes sortes sont à notre portée ; en fait, il est clair que ceux qui le sont obéissent à de sérieuses contraintes. Quant à la promesse selon laquelle l'intelligence de niveau humain est à notre portée, j'affirme qu'elle est complètement vaine. Je ne peux proposer aujourd'hui que deux arguments. Le premier, c'est que les leaders scientifiques les plus éminents de l'IA aujourd'hui s'accordent tous sur la nécessité de nouvelles connaissances, sans lesquelles l'IA stagnera. Aujourd'hui, l'IA est peut-être à la veille d'un tournant similaire à la révolution des réseaux neuronaux, mais elle ne sait pas encore quelle direction prendre. La seconde est l'observation selon laquelle l'intelligence humaine ne se limite pas à résoudre des problèmes, et je ne vois pas comment l'IA telle qu'elle est conçue actuellement peut faire autre chose que résoudre des problèmes.

Ces hypothèses ne sont pas seulement invraisemblables ; elles sont également potentiellement nocives. Elles envoient les professionnels dans une quête futile, celle de penseurs artificiels et totalement autonomes, au lieu de s'en tenir à ce que je considère comme la vocation majeure de l'IA, qui est de fournir à l'humanité des auxiliaires puissants et fiables qui peuvent nous aider à surmonter une partie des problèmes actuels aussi bien techniques, scientifiques, sociaux que politiques, ainsi que faciliter les tâches quotidiennes qui nécessitent réellement une aide. Et elles encouragent une falsification majeure : celle de faire passer des systèmes mécaniques pour de véritables êtres humains en silicone. L'ironie est que certaines personnes s'inquiètent du soi-disant « risque existentiel » posé par l'intelligence humaine et, en un rien de temps, par l'intelligence surhumaine. À mon avis, cette inquiétude est déplacée. Ce qui m'inquiète cependant, c'est la combinaison de la croyance infondée selon laquelle l'AGI est imminente avec une priorité déplacée accordée à l'objectif d'implanter l'IA dans autant de contextes que possible, dans le seul but d'utiliser un outil aussi merveilleux, peu importe les conséquences au sens le plus large.

Le principal défi aujourd'hui est de faire de l'IA une discipline d'ingénierie à part entière, capable de produire, de manière bien comprise, des objets fiables dotés de garde-fous intégrés contre une utilisation inappropriée.

Patrick Nicolet

Merci, Daniel. Je vais résumer et d'après ce que j'ai compris, et nous sommes d'accord au sein du panel, ce que nous voyons aujourd'hui est une IA spécifique et même si l'IA générale reste une possibilité, nous ne voyons pas comment y parvenir aujourd'hui. Nous parlons beaucoup d'intelligence artificielle aujourd'hui parce que, comme vous l'avez décrit, il y a une avancée majeure. Auparavant, on pouvait saisir des données dans l'intelligence artificielle ; elles étaient simples puis sont devenues plus complexes, mais le résultat restait simple. Avec ChatGPT, on peut désormais produire des résultats complexes. Avec une saisie et des résultats complexes, on peut utiliser du texte, des images, des vidéos, du son et produire la même chose, ce que nous ne pouvions pas faire auparavant. Là est la percée. Cela a un impact non seulement lorsque vous jouez avec vos enfants, mais aussi dans l'entreprise, et comme nous l'avons entendu hier lors de la discussion avec Virginie Robert, cela peut interférer dans les processus démocratiques.