

chercheurs en IA, pensent plutôt qu'il faudra attendre 2040, voire 2100.

Il se pourrait que l'on se trouve dans une situation analogue à celle du jeu d'échecs : on y arrivera, mais pas dans les délais annoncés, car le problème est réellement difficile, ce que l'on découvre année après année. Le plus raisonnable serait sans doute de ne pas tenter aujourd'hui de pronostics !

Liée à cette éventuelle superintelligence artificielle, l'idée d'une « singularité » à venir est évoquée sans qu'on puisse savoir vraiment de quoi il s'agit. Une date où se situerait une sorte d'asymptote verticale dans la courbe mesurant le niveau des intelligences artificielles créées sur Terre ? Un point où les machines écrasent les humains prendraient le pouvoir ? Un instant au-delà duquel il est impossible de comprendre ce qui va se passer à cause des bouleversements entraînés par l'existence d'intelligences artificielles nous dépassant ?

Tout cela n'est pas très sérieux, même en prenant en compte la célèbre « loi de Moore » (vérifiée en gros depuis cinquante ans) qui indique un doublement de capacité des dispositifs informatiques tous les dix-huit mois ou deux ans. D'une part, une fonction exponentielle comme celle envisagée par la loi de Moore n'a pas d'asymptote verticale. D'autre part, dans notre monde physique, toute croissance exponentielle rencontre rapidement des obstacles qui l'arrêtent et personne ne défend sérieusement que la loi de Moore (qui s'essouffle déjà) continuera à être valide indéfiniment. Dernier point : l'augmentation des capacités de calcul et de mémoire ne constitue pas la garantie qu'on saura les utiliser pour créer cette fameuse intelligence générale qui nous résiste.

Les annonces des Gafam (Google, Apple, Facebook, Amazon, Microsoft) sur ces sujets doivent être examinées avec méfiance. Certains spécialistes comme Jean-Gabriel Ganascia, président du comité d'éthique du CNRS et lui-même chercheur en intelligence artificielle, sont plus rigoureux et lucides que les agents publicitaires des firmes de technologie. Jean-Gabriel Ganascia considère que ceux-ci cherchent à promouvoir un projet du dépassement de la politique et des États-nations par des entreprises à l'avidité sans borne. L'idée de la singularité, utilisée pour amuser et effrayer, aurait en réalité surtout pour but d'intimider et de masquer leurs ambitions dominatrices.

Les difficultés à anticiper avec précision n'interdisent pas de s'adonner au jeu de l'anticipation en s'interrogeant sur l'apparition d'une superintelligence artificielle, quitte pour cela à imaginer plusieurs scénarios, sans donner d'échéances et en envisageant qu'éventuellement aucun ne se réalise. Jouons.

Nick Bostrom définit dans son livre une superintelligence comme tout intellect qui

excède radicalement les capacités cognitives humaines et ce dans tous les domaines. Deux hypothèses assez différentes sont envisagées.

(a) Les machines à base de silicium, y compris les ordinateurs quantiques s'ils se révèlent utiles, progressent au point que les dispositifs construits nous égalent et nous dépassent dans toutes les tâches intellectuelles possibles, cela, comme pour AlphaZero, sans que nous soyons associés directement à leur fonctionnement et sans que des éléments biologiques aient été fusionnés ou connectés à ces créatures intelligentes artificielles.

(b) Seule une association entre des éléments (humains ou non) provenant de la biologie et des dispositifs informatiques, donc un mixte de technologies inventées par l'homme et de parties utilisant ce que la vie sur Terre a patiemment élaboré, réussit ce dépassement.

ARTILECTS OU CYBORGS ?

Les êtres intelligents provenant de l'hypothèse (a) sont dénommés *artilects* (pour *artificial intellect*) par Hugo de Garis, un chercheur qui a développé un scénario pour le futur sur lequel nous allons revenir.

La science-fiction a envisagé l'hypothèse (b) depuis longtemps et dénommé cyborgs (pour *cybernetic organisms*) ces êtres hybrides superintelligents grâce au mélange et à l'imbrication de biologie et d'informatique. Imaginons par exemple que, pour suppléer à notre mémoire insuffisante, on nous greffe des puces contenant des téraoctets de données qui nous permettraient directement par la pensée de savoir tout ce qu'il y a dans l'encyclopédie Wikipédia, et que grâce à un module de calcul mathématique lui aussi accessible instantanément par la pensée, nous sachions répondre à n'importe quelle question du type « Combien fait 7^{20} ? » ou « Quelle est la dérivée cinquième de $\sin(\cos(x))$? » Notre intelligence en serait accrue !

Une troisième hypothèse est parfois évoquée, celle d'un cerveau planétaire naissant des réseaux de machines qui s'échangent des informations et constituent une sorte d'être décentralisé et pensant ; nous y reviendrons.

LE PESSIMISME DE HUGO DE GARIS

Pour Hugo de Garis, « la société se divisera en trois groupes philosophiques majeurs, agressivement opposés. Le premier groupe sera constitué par les *cosmistes* (terme dérivé du mot *cosmos*), favorables à la construction d'artilects. Le deuxième groupe sera constitué des *terrans* (dérivé du mot *Terra*, la Terre) opposés à la construction d'artilects. Le troisième groupe sera celui des *cyborgistes*, qui souhaitent devenir eux-mêmes des artilects en ajoutant des composants à leur propre cerveau humain. »

>



3

LE TRANSHUMANISME

Nick Bostrom, philosophe suédois et cofondateur de la World Transhumanist Association, pense que l'homme doit tenter de se transformer physiquement et ne voit pas de limite à cette ambition : « Nous ne sommes pas contraints d'accepter ce que la nature a fixé, nous ne sommes par exemple pas obligés d'accepter de mourir après quelques décennies sur Terre. Nous ne sommes pas plus obligés d'accepter les limites dans nos capacités mentales. La vie meilleure envisagée par les transhumanistes consiste en plus du contrôle sur notre propre vie, et en la possibilité de rester totalement maître de nos idéaux. Évidemment, ces idéaux, dans un monde posthumaniste, varieraient d'un individu à l'autre. Les transhumanistes veulent que chacun puisse faire ses propres choix. Le sens à donner à l'idée d'une vie meilleure sera une affaire personnelle. Dans le monde idéal du transhumanisme, la mort serait volontaire. Si vous pensez que vous avez assez vécu, que vous avez apporté au monde ce que vous aviez à y apporter, que vous avez atteint vos objectifs et qu'il n'y a plus de raison de vivre, vous arrêterez de prendre des suppléments de vie ou d'utiliser toute autre technique pour perdurer. En revanche, si vous croyez qu'il serait peut-être bon de vivre encore et de voir ce que la vie peut vous apporter, vous en auriez encore la possibilité. Dans un monde développé technologiquement, nous serons toujours confrontés à des défis variés supérieurs à ceux d'aujourd'hui. Prenons le cas des mathématiques. Vous pouvez faire facilement certains calculs qui étaient

autrefois inconcevables : vous n'avez qu'à pianoter sur votre calculatrice. Cette nouvelle situation n'a pas aboli les défis mathématiques, elle a, au contraire, ouvert la voie à des défis plus intéressants et plus nombreux. »

Le transhumanisme est souvent critiqué, voire considéré comme scandaleux. Une déclaration du Vatican intitulée « Communion et service : la personne humaine créée à l'image de Dieu » affirme ainsi : « Changer l'identité génétique de l'homme en tant que personne humaine par la production d'un être infrahumain est radicalement immoral. Le recours à la modification génétique pour produire un surhomme ou un être doté de facultés spirituelles essentiellement nouvelles est impensable, puisque le principe de la vie spirituelle de l'homme (le principe qui informe la matière pour en faire le corps d'une personne humaine) n'est pas produit par des mains humaines et n'est pas sujet à la manipulation génétique. L'unicité de chaque personne humaine, en partie constituée par ses caractéristiques biogénétiques et développée par l'alimentation et la croissance, lui appartient intrinsèquement et ne peut pas être instrumentalisée dans le but d'améliorer certaines de ces caractéristiques. [...] L'homme est créé à l'image de Dieu mais il n'est pas Dieu. »

Les idées des transhumanistes s'opposent aussi à celles de Hugo de Garis qui ne croit pas que l'homme augmenté des transhumanistes puisse rester compétitif face aux machines superintelligentes qui extermineront les êtres humains et les cyborgs.

> Dans son livre *The Artilect War*, ouvrage intermédiaire entre science-fiction et essai sérieux de futurologie, Hugo de Garis considère inévitable une guerre entre ces groupes. Pour lui, elle sera plus sanglante encore que toutes celles ayant eu lieu sur Terre et provoquera au cours du *xxi^e* siècle des hécatombes de plusieurs milliards de morts. Cette guerre a peu de chance de s'achever en faveur des terrans qui, s'ils veulent gagner, doivent agir rapidement avant que les deux autres groupes, capables de s'associer, comprennent qu'ils sont menacés par les terrans. Dans un entretien diffusé sur Discovery Channel, Hugo de Garis a exprimé son inquiétude : « Je suis content d'être en vie maintenant. Je mourrai sans doute paisiblement dans mon lit. Cependant, je crains vraiment pour mes petits-enfants. Ils seront mêlés à la guerre d'Artilect et seront probablement détruits par elle. »

Assez curieusement, les transhumanistes, peu appréciés en France, où la très grande majorité des articles et livres qui les mentionnent le font pour les condamner ou s'en moquer, défendent des idées bien optimistes comparées à cette guerre annoncée par Hugo de Garis. Les transhumanistes soutiennent l'idée qu'il faut aller, sans retenue, vers l'augmentation et le perfectionnement technique des humains et non seulement porter des lunettes, des prothèses auditives ou des *pace-makers*, mais utiliser tout ce qu'on pourra pour allonger nos vies, renforcer nos capacités intellectuelles et physiques. En clair, ils sont favorables à l'idée des cyborgs, et seraient donc dans le camp cyborgiste tentant de s'associer aux cosmistes pour éviter d'être préventivement éliminés par les terrans.

Concernant ce qui pourrait se passer entre cyborgistes et cosmistes, les avis des transhumanistes et de Hugo de Garis divergent. La plupart des transhumanistes envisagent la possibilité d'une coexistence entre les cyborgs en partie humains et les machines devenues superintelligentes sans composants biologiques. Les transhumanistes croient que le fait d'associer des éléments vivants (eux en particuliers) et la technologie électronique permettrait de maintenir dans la compétition les humains augmentés qui, peut-être devenus immortels, resteraient la clé du futur de l'humanité.

Hugo de Garis défend le contraire : à côté des capacités que pourraient acquérir les artilects, le projet transhumaniste ne fait pas le poids. Pour lui, la voie du mélange du vivant avec le futur des technologies informatiques est une illusion : les artilects n'auront que faire des hybrides monstrueux que seraient devenus les humains modifiés et s'en débarrasseront.

Nick Bostrom, cofondateur de la World Transhumanist Association, n'est pas persuadé

que les artefacts adopteront une attitude pacifique envers leurs créateurs humains éventuellement devenus cyborgs. Il nous incite à surveiller nos créations intelligentes futures qu'il voit comme des menaces. L'ouvrage de Nick Bostrom a été lu par Elon Musk, Bill Gates et Robin Li (du géant chinois Baidu) et semble les avoir ébranlés, ce qui les a amenés à répandre un message de crainte et des mises en garde à propos de nos futures IA.

Pour Hugo de Garis, il n'y aura pas d'intelligence artificielle amicale tolérante aux humains ou même aux cyborgs: «L'idée d'une queue qui remue le chien est évidemment ridicule. Un chien est beaucoup plus gros que sa queue, donc la queue ne peut pas remuer le chien. C'est pourtant ce que proposent les transhumanistes: les artefacts (esprits artificiels, machines massivement intelligentes) resteraient amicaux avec les humains. [...] Je trouve cette conception arrogante et naïve. Elle suppose que les êtres humains sont suffisamment intelligents pour anticiper les motivations d'une créature disposant de capacités des milliards de milliards de fois supérieure à celles humaines.»

AUTRES SCÉNARIOS, AUTRES ARGUMENTS

La position de Hugo de Garis semble aveugle à un mouvement qui enlèverait tout sens à l'idée de guerre exterminatrice: la construction d'un superorganisme mondial associant et liant tous les êtres qui, à la surface du globe, ne seraient pas en concurrence, mais associés.

Cette idée d'une unité à l'échelle globale est en général considérée comme religieuse. Elle est présente dans la philosophie bouddhiste et chez le philosophe jésuite Pierre Teilhard de Chardin (1881-1955), qui a évoqué une convergence vers la «noosphère», une «pellicule de pensée enveloppant la Terre, formée des communications humaines».

Le philosophe belge Francis Heylighen défend l'idée que la superintelligence sera distribuée à travers le réseau Internet, qui fonctionnera comme un supercerveau: «Ce cerveau mondial relèvera tous les défis auxquels est confronté le superorganisme mondial. Ses capacités iront bien au-delà de nos capacités actuelles. Il possédera les attributs divins: omniscience (il saura tout ce qu'il faut pour résoudre nos problèmes), omniprésence (il sera présent partout, à tout moment), omnipotence (il pourra produire n'importe quelle chose ou service de la manière la plus efficace) et omnibienveillant (il visera le plus grand bonheur pour le plus grand nombre).» L'exact opposé du pessimisme de Hugo de Garis!

Clément Vidal, philosophe, spécialiste de cosmologie à la Vrije Universiteit de Bruxelles,

rappelle la phrase du poète William Stafford que «toute guerre fait deux perdants» et que la violence est le plus souvent irrationnelle. Steven Pinker détaille cette pacification du monde qui, même si on la trouve bien trop lente, se mesure et s'explique. D'une part, l'interdépendance de tous vis-à-vis de tous nous lie, qu'on le veuille ou non. Des arguments de théorie des jeux et des simulations informatiques sur la base du modèle du dilemme des prisonniers s'ajoutent à cette analyse des faits. Ils montrent que la coopération est l'issue rationnelle de nombreuses configurations sociales et politiques.

CONVERGENCE DES BUTS

L'éthique de la complexité (*voir cette rubrique dans Pour la Science de septembre 2018*) est un dernier argument montrant qu'une convergence des motivations aboutissant à une unification des intérêts de tous sur Terre n'est pas absurde. Si tous les êtres intelligents choisissent de servir un ensemble de valeurs universelles communes comme cette éthique l'envisage, alors ils agiront avec des objectifs partagés et les raisons de mener des guerres cesseront d'opérer.

Bien sûr, quoique convergents, ces divers d'arguments ne constituent pas une preuve définitive que lorsque des superintelligences existeront sur Terre (si cela se produit), alors elles formeront avec nous les humains, devenus plus ou moins cyborgs, un tout pacifié sans qu'aucune partie ne tente d'en éliminer d'autres. Il me semble cependant que l'on peut défendre l'idée très simple que l'humanité n'est pas simplement l'addition des corps biologiques présents sur Terre, mais l'ensemble de tout ce qu'elle a produit et même de tout ce qui lui permet d'y vivre. Ce corps collectif qu'est la Terre évoluant vers plus d'intelligence et des liens de plus en plus resserrés et unis, en particulier par les multiples réseaux récemment déployés à la surface du globe, ne déclenchera probablement jamais cette guerre interne imaginée par Hugo de Garis, qui serait une forme d'automutilation, voire de suicide.

Nous sommes là maintenant au-delà de la science-fiction; cependant, prenons garde de ne pas condamner toute réflexion sur ces questions en les considérant comme stupides et prématurées. AlphaZero nous montre que ce que nous arrivons à tirer de nos machines est incroyable: l'intelligence artificielle cesse progressivement d'être un rêve. Par ailleurs, le développement des réseaux de toutes sortes tisse un cocon dense de liens entre tous les acteurs présents dans le monde. Sans le moindre doute, il se construit une structure nouvelle où tout dépend de tout et dont nous ne sommes, comme nos machines intelligentes, que des composants. ■

BIBLIOGRAPHIE

Wikipédia, **La singularité technologique**, consulté en décembre 2018.

D. Silver et al., **A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play**, *Science* vol. 362, pp. 1140-1144, 2018.

J.-G. Ganascia, **Le Mythe de la singularité**, Seuil, 2017.

N. Bostrom, **Superintelligence**, Dunod, 2017.

S. Pinker, **La Part d'ange en nous**, Les Arènes, 2017.

B. et T. Goertzel (dir.), **The End of the Beginning: Life, Society and Economy on the Brink of the Singularity**, Humanity+ Press, 2015.

H. de Garis, **The Artefact War - Cosmists vs. Terrans**, ETC Publications, 2005.



Mycelium chair, du studio Klarenbeek & Dros, est une chaise en mycélium de champignon qui a poussé autour de déchets agricoles.