

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.