RENDEZ-VOUS

LE POTENTIEL DANGER DES INTELLIGENCES SURHUMAINES

Entre l'idée de superintelligences menaçantes et celle, plus sympathique, de machines qui ne seraient que notre prolongement, laquelle est la plus vraisemblable?

L'AUTEUR

JEAN-PAUL DELAHAYE
professeur émérite à l'université de Lille et chercheur au Centre de recherche en informatique, signal et automatique de Lille (Cristal)

Un exploit étonnant vient d'être réalisé dans le domaine des recherches en intelligence artificielle (IA). Le programme AlphaZero, conçu et programmé par David Silver et son équipe de la société DeepMind, apprend seul, en jouant contre lui-même et en ne connaissant que les règles, à jouer aux échecs, au go ou au shogi (échecs japonais). Il a dépassé le niveau des meilleurs joueurs humains dans ces trois jeux.

L'exploit est un formidable progrès: l'ordinateur Deep Blue, qui a battu en 1997 le champion du monde d'échecs Gary Kasparov, combinait de bons algorithmes et une connaissance experte des ouvertures de jeu, des fins de partie et de l'évaluation des configurations, ces expertises provenant des meilleurs joueurs humains qui avaient en quelque sorte «instruit le programme».

Avec AlphaZero, plus besoin d'experts humains transmettant un savoir patiemment acquis; pour ce type de jeux, l'IA travaille seule et obtient mieux que ce à quoi aboutissent des siècles de concentration humaine. Les chercheurs qui ont mis au point AlphaZero précisent: «Nous avons appliqué AlphaZero aux jeux d'échecs, de shogi, ainsi qu'au go, en utilisant le même algorithme et la même architecture réseau pour les trois jeux. Nos résultats démontrent qu'un algorithme polyvalent d'apprentissage par renforcement [l'algorithme favorise les choix qu'il fait quand il gagne] peut apprendre, *tabula rasa*, sans connaissances ni données humaines spécifiques.»

À mesure que l'intelligence artificielle progresse, des questions de plus en plus pressantes sont formulées à son sujet. Certains chercheurs et philosophes qualifiés parfois de technoprophètes nourrissent le débat et proposent des visions du futur, variées et contradictoires. Qu'on s'en réjouisse ou qu'on le craigne, nul ne doute que les machines intelligentes, peut-être mises en réseaux ou fusionnées aux humains, façonneront notre avenir.

Parmi ces penseurs qui tentent de déchiffrer le futur, citons Nick Bostrom, Ray Kurzweil, Ben et Ted Goertzel, Hans Moravec, Francis Heylighen, Hugo de Garis, et Clément Vidal. Tous réfléchissent à ce que pourrait être une intelligence artificielle supérieure et à son attitude face à l'humanité.

EXCÈS D'HONNEUR ET INDIGNITÉ

Comme souvent quand on comprend mal un sujet, on procède par excès et, selon son tempérament, on voit exclusivement le verre à moitié vide ou le verre à moitié plein. Pour reprendre les mots que Racine met dans la bouche d'un de ses personnages, l'IA ne mérite cependant ni l'excès d'honneur qu'on lui accorde ni l'indignité liée à la crainte qu'elle suscite.

L'étonnement face aux succès obtenus et la prise de conscience des impacts que cela aura ne doivent pas faire perdre le sens de la mesure. L'IA réussit magnifiquement dans de nombreuses tâches spécialisées, mais reste encore démunie pour tout ce qui demande du sens commun, comme la traduction automatique >

Jean-Paul Delahaye a récemment publié: **Les Mathématiciens se plient au jeu**, une sélection de ses chroniques parues dans *Pour la Science* (Belin, 2017).