

Il y a eu une sorte de conjonction de planètes avec l'arrivée, presque au même moment, de beaucoup plus de puissance de calcul, de plus en plus de corpus de données pour nourrir l'apprentissage, et du développement de nouveaux algorithmes dits « d'apprentissage profond » (*deep learning*, en anglais). D'un coup, des problèmes qui nous résistaient depuis des années se sont mis à tomber. Cela avait un côté génial. Mais, encore une fois, cette approche n'a pas remplacé ce qui existait avant. Même quand AlphaGo, de DeepMind, a battu le joueur de go Lee Sedol, il n'utilisait pas uniquement des algorithmes d'apprentissage profond.

Le revers de ces méthodes n'est-il pas leur opacité ?

En effet, quand on fait tourner un algorithme d'apprentissage profond, on ne sait pas expliquer pourquoi on arrive à un résultat particulier. Les longs calculs réalisés ne font pas à proprement parler un raisonnement, en tout cas un raisonnement qu'un humain serait capable de comprendre. On peut penser que, tant pis, seule l'efficacité prime, mais ce n'est pas si simple. Prenons deux exemples en médecine : un algorithme d'apprentissage qui aide à retrouver des tumeurs cancéreuses (*voir La fée IA au chevet des malades, par N. Ayache, page 18*) examine des milliers d'images annotées par des médecins, et à partir de toute cette connaissance se prononce sur les images qu'on lui soumettra. Impossible pour un humain de se former en étudiant toutes ces images : il y en a trop. Et de toute façon, il y aura un médecin, voire une

équipe, pour discuter l'avis de la machine qui sera un avis comme un autre, pris comme tel.

En revanche, en matière de diagnostic médical, si vous rentrez dans un programme un grand nombre d'informations sur un patient, et qu'à la fin ce logiciel décide « c'est une hépatite », ça ne peut pas suffire au médecin qui a besoin d'explications. Il ou elle a besoin d'entendre que, en fonction des observations du malade, statistiquement ce peut être telle maladie avec 95 % de chances, mais aussi telle autre avec 5 % de chances, et qu'il faudrait poser telle question au malade pour écarter telle possibilité, ou demander tel examen complémentaire, etc. Dans les deux cas, il y a un travail collaboratif entre machine et humains. Dans le premier, nul besoin d'explications (les techniques d'apprentissage automatique un peu brutales sont efficaces). Dans le second, des explications sont indispensables.

Où finit l'informatique « ordinaire », où commence l'intelligence artificielle ?

Cette distinction n'a pas vraiment de sens. On a, à l'intérieur de l'informatique, un vrai continuum.

Quelle est alors votre définition de l'intelligence artificielle ?

Pour Alan Turing, une activité d'une machine sera qualifiée d'« intelligence artificielle » si elle est considérée comme intelligente quand un humain s'y livre. À ses yeux, ce ne peut être qu'une imitation de l'intelligence humaine, une