



DES SYSTÈMES PERFORMANTS QUI GAGNENT

Il existe des programmes informatiques qui effectuent des tâches spécialisées avec une efficacité impressionnante. C'est le cas du programme AlphaGo, qui, en 2016, a battu au jeu de go le champion Lee Sedol (à droite sur la photographie ci-contre).

Et début 2017, le programme Libratus a battu quatre joueurs professionnels de poker Texas Hold'em (page ci-contre, Jason Les jouant contre Libratus), un jeu où l'on fait face à des informations incomplètes et à des ruses. Ici, la capacité de calcul ne suffit plus...

dialoguer et raisonner, ainsi qu'à réguler les émotions chez soi et chez les autres.

Pour autant, cela n'a pas empêché le domaine de l'intelligence artificielle d'avoir beaucoup progressé ces dernières années et d'avoir acquis une importance stratégique. Ces avancées sont liées à deux ruptures technologiques : d'une part, la puissance de calcul des ordinateurs permet aujourd'hui d'analyser d'énormes masses de données en quelques secondes ; d'autre part, avec les techniques d'apprentissage profond (*deep learning*, en anglais), des réseaux de neurones artificiels apprennent à effectuer certaines tâches selon des mécanismes plus efficaces que ceux utilisés dans les années 1990.

Ainsi, pour une tâche qui exige uniquement des calculs aujourd'hui bien compris, comme jouer aux échecs ou au go, les programmes informatiques affichent des performances très impressionnantes et peuvent battre des champions du monde : le superordinateur Deep Blue, d'IBM, l'a fait en 1997 pour les échecs, et le programme AlphaGo, de l'entreprise DeepMind-Google, en 2016 pour le go. Le système Libratus, conçu à l'université Carnegie-Mellon, a battu quatre joueurs professionnels au poker dans sa version Texas Hold'em. Il faut comprendre que si le poker est effectivement un jeu plus complexe que les échecs et le jeu de go, cette complexité ne doit pas être exagérée, le nombre de cartes étant limité, et les stratégies aussi. Ce qui est très différent lorsqu'il s'agit du langage, car les possibilités sont infinies.

La possibilité de parler naturellement à un *chatbot* afin qu'il exécute des tâches à notre service ne tient pas du fantasme, même si la

technologie actuelle est encore loin de ce que montre la science-fiction. Mais les agents conversationnels ne comprennent pas l'information qu'ils traitent, ce qui est la majeure différence avec la notion d'intelligence artificielle dite « forte » (qui n'est pas pour demain), laquelle vise à reproduire une véritable compréhension de l'information.

Donner aux robots conversationnels des connaissances sémantiques, une compréhension du sens commun et des règles sociales est un défi scientifique complexe, auquel les géants du numérique (Google, Apple, Facebook, Amazon, Microsoft, IBM) se sont attelés. Les *chatbots* intégrés aux smartphones ou aux robots sont désormais de plus en plus courants, avec 4,2 milliards d'assistants vocaux utilisés en 2020 dans le monde. Ce chiffre devrait passer à 8,4 milliards d'ici à 2024.

DOTER LE PROGRAMME D'UNE FORME DE PERSONNALITÉ

La robotique conversationnelle se développe également ; par exemple le robot Buddy, de la société française Blue Frog Robotics, vise le marché d'assistance aux personnes âgées. Par ailleurs, les concepteurs de *chatbots* ajoutent souvent une « personnalité » à la machine. Celle-ci a alors des préférences ou des habitudes qu'elle exprime dans la conversation, ce qui lui donne un semblant d'identité. Il existe d'ailleurs des bibliothèques, comme Koko, issue des travaux du MIT, qui permettent de choisir, selon le contexte du dialogue, des réponses empathiques.

Dans tous ces développements autour des agents conversationnels, les fonctionnalités



99

émotionnelles – empathie, humour, etc. – prennent de plus en plus d'importance pour que l'on puisse interagir de façon naturelle avec une machine. D'ailleurs, l'informatique émotionnelle fait partie des axes de recherche de mon équipe depuis 2004 au LISN-CNRS, où l'on explore trois types de problématiques et de technologies : la détection des émotions, le raisonnement exploitant des informations de nature affective, mais aussi l'impact de la génération d'expressions émotionnelles. En particulier, l'humour, à un degré assez simple, fait partie des capacités que l'on souhaite donner aux machines. On peut le simuler à travers divers types d'expressions – verbales ou non verbales – socialement et culturellement adaptées, et l'utiliser pour rendre les machines plus attachantes et plus amusantes à utiliser, et ainsi améliorer nos interactions avec elles.

Plus récemment, dans ma chaire *Human-Machine affective interaction & ethics* (Humaine), nous mettons en œuvre et testons des stratégies de *nudges* (« coups de pouce ») dans des systèmes conversationnels affectifs. Il s'agit d'une technique qui consiste à inciter les individus à changer de comportement sans les contraindre et en utilisant leurs biais cognitifs (voir *Communication, les deux visages*

Donner de l'humour
aux machines,
c'est les rendre
plus attachantes,
plus amusantes,
et ainsi améliorer
nos interactions
avec elles

de l'intelligence artificielle, par L. Devillers, page 42). Nous cherchons à en montrer les bénéfices et les risques dans le dialogue oral. Ces systèmes conversationnels seront de plus en plus capables de discuter avec une personne

→ Plus qu'un simple jouet, Moxie est un robot compagnon destiné aux 5-10 ans, capable, selon son concepteur Embodied, de converser en langage naturel pour instruire les enfants et, plus encore, les inciter à développer leurs relations sociales dans la vie de tous les jours. Connectée à une plateforme de mise à jour, l'IA de Moxie évolue à mesure que l'enfant grandit.



en lui donnant l'impression de converser avec un humain.

Comment évaluer les performances d'un agent conversationnel en matière d'imitation, de conversation, de mémorisation, de raisonnement, de simulation affective, d'anticipation ? L'une des difficultés est que ces capacités sont conférées par des algorithmes et des techniques d'apprentissage profond, qui gardent une forme d'opacité : ces techniques fonctionnent selon un principe transparent et bien compris, mais elles produisent des décisions dont il est impossible de retracer les raisons et le cheminement. Hormis ce problème dit « de boîte noire », nous sommes en droit de nous poser bien d'autres questions sur les *chatbots*, en particulier leurs enjeux éthiques. Sur ce sujet, un avis que j'ai corédigé à la CNPEN (Commission nationale pilote d'éthique du numérique) a été rendu au Premier ministre en novembre 2021.

À TROP FRÉQUENTER LES HUMAINS, LE RISQUE EST DE DEVENIR AGRESSIF

Par exemple, la machine en question respecte-t-elle la vie privée de ses utilisateurs et des autres humains ? Se comporte-t-elle de façon éthiquement responsable, et, si oui, est-ce que ce sera toujours le cas jusqu'à la fin de sa vie ? Peut-on créer des jumeaux numériques de personnes décédées et quelles seraient nos relations avec elles ? Est-ce suffisant de tester la machine avant sa mise sur le marché si elle est amenée à s'adapter en continu ? De telles questions sont particulièrement importantes dans le cas de systèmes capables d'apprendre tout au

long de leur durée d'utilisation. Il est alors difficile de contrôler leur comportement futur et de garantir que ce dernier restera acceptable. Un *chatbot* au contact d'humains pourrait développer des attitudes agressives ou inconvenantes, un algorithme conçu pour faire des offres financières pourrait à terme orienter ses propositions en fonction du groupe social auquel appartient son interlocuteur, etc.

Un exemple d'une telle dérive a été fourni en 2016 par Tay, un système conversationnel développé par Microsoft. Ce *chatbot*, déployé pour converser avec des jeunes sur la plateforme Twitter, était doté d'une capacité d'apprentissage en continu. On a pu constater qu'il pouvait apprendre de ses interactions avec les internautes, puisqu'il adaptait son langage aux habitudes et éléments qu'il repérait dans les conversations. Mais l'essai a tourné court : au bout de vingt-quatre heures seulement, Tay s'est mis à poster sur la plateforme des tweets à caractère raciste et nazi !

Comment faire évoluer le test de Turing pour évaluer les capacités des agents conversationnels ? Rappelons-le, ce test éprouve la faculté d'une machine à imiter la conversation humaine. La mission du juge est de deviner,