réponses et de synthèse de la parole. Connectés sur Internet ou embarqués sur un objet ou un robot, ces programmes, invisibles aux yeux des utilisateurs, sont ce qu'on nomme des agents conversationnels, ou encore robots bavards (*chatbots* en anglais). Le premier d'entre eux était Eliza, construit au MIT (l'institut de technologie du Massachusetts) vers 1964-1966 par Joseph Weizenbaum; ce système simulait un psychothérapeute rogérien, dont la stratégie consiste, pour l'essentiel, à répéter les propos du patient.

Au-delà des avancées scientifiques et techniques mises en œuvre dans les agents conversationnels, l'interaction croissante que nous avons avec ces derniers soulève des questions plus fondamentales, voire stratégiques et éthiques, sur les capacités des machines et sur la relation intersubjective qu'instaurent avec celles-ci leurs utilisateurs humains. En particulier, si l'on veut vivre harmonieusement avec des machines qui nous aident et avec lesquelles nous dialoguons, il apparaît de plus en plus important d'évaluer de façon pertinente, par des batteries de tests reproductibles, ces capacités ainsi que leur impact sur la relation de l'humain avec la machine.

LE TEST DE TURING, INSUFFISANT POUR PLUSIEURS RAISONS

Que faut-il tester plus précisément, et comment? On s'est longtemps focalisé sur l'«intelligence» des machines. D'ailleurs, le domaine dit de l'intelligence artificielle, expression introduite dans les années 1950 par le chercheur américain John McCarthy, a pour objectif de «doter des machines de systèmes informatiques ayant des capacités intellectuelles comparables à celles des hommes».

Or cela suppose notamment que l'on sache comment comparer ces capacités. En 1950, le mathématicien britannique Alan Turing proposait le «jeu de l'imitation» pour déterminer si une machine est intelligente ou non. Il préfigurait l'orientation des futures technologies de l'information et de la communication. Le «test de Turing», comme on le nomme aujourd'hui, consiste à confronter par le biais d'une conversation textuelle deux interlocuteurs – un humain et une machine – à un juge, lequel doit déterminer qui est la machine et qui est l'humain. Si la machine réussit à faire croire à l'examinateur qu'il a affaire à un interlocuteur humain, c'est qu'elle a des capacités qui ressemblent à ce que nous appelons de l'intelligence.

Le test de Turing est-il la meilleure façon d'évaluer une machine conversationnelle? À ce jour, aucun programme informatique n'a réussi à passer le test de Turing de façon indiscutable. Cependant, le jeu de l'imitation proposé par Turing apparaît de plus en plus insuffisant, pour plusieurs raisons. L'une d'elles est qu'il ne détermine pas directement les capacités, cognitives ou autres, du programme: il teste seulement si ce programme peut simuler un comportement humain et tromper l'examinateur (*voir l'article de Gary Marcus, pages 26 à 31*).

Un autre argument remettant en cause la puissance du test de Turing a été avancé en 1980 par John Searle. Selon ce philosophe américain, un ordinateur (ou un logiciel) ne fait que manipuler de la syntaxe, et réussir le test de Turing ne prouverait donc rien. Pour l'illustrer, John Searle a proposé une expérience de pensée dite de la chambre chinoise. Supposez que vous vous trouviez à l'intérieur d'une pièce muni d'un ensemble de caractères chinois et d'un manuel d'instructions de type: «Si la question posée est constituée de telle suite de caractères, donnez la réponse formée avec telle autre suite de caractères.» Vous pouvez alors répondre aux questions d'un locuteur chinois situé à l'extérieur de la pièce et fournir des réponses adéquates grâce au manuel. Ce faisant, vous donnez l'impression à votre interlocuteur de savoir parler sa langue, sans qu'il soit nécessaire que vous la compreniez.

Les systèmes actuels d'intelligence artificielle font de même: ils imitent (ou cherchent à imiter) le comportement et le langage des humains, mais ils ne comprennent pas le sens de l'information qu'ils traitent.

Pour en revenir au test de Turing, il est insuffisant pour une autre raison encore: il ne prend pas en compte les diverses et nombreuses facettes de l'intelligence humaine. On définit souvent l'intelligence comme la capacité à mener un raisonnement abstrait, mais

BIOGRAPHIE

LAURENCE DEVILLERS **dirige au Limsi (Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur, unité propre du CNRS) l'équipe Dimensions affectives et sociales dans les interactions parlées et est membre de la Cerna, la commission de réflexion sur l'éthique d'Allistene (Alliance des sciences et technologies du numérique). Elle vient de publier *Des robots et des hommes : mythes, fantasmes et réalité* (Plon, 2017).**

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 dans la photographie ci-dessous)*. Et début 2017, le programme Libratus a battu quatre joueurs professionnels de poker Texas Hold'em *(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...

© Geordie Wood *(à gauche)* - Nate Smallwood *(à droite)*

l'intelligence émotionnelle ou sociale est un autre aspect primordial du comportement humain, qui requiert d'autres compétences. De façon générale, l'intelligence regroupe un vaste ensemble de fonctions mentales : apprentissage, compréhension et organisation du réel en concepts, interprétation des règles sociales et culturelles, capacité à utiliser le raisonnement causal, imagination, prospection et flexibilité, mais aussi habileté à percevoir et à exprimer les émotions, à les intégrer pour comprendre, dialoguer et raisonner, ainsi qu'à réguler les émotions chez soi et chez les autres.

Le test de Turing ne prend pas en compte de nombreuses fonctions mentales

L'intelligence est un concept tellement polysémique que nous ne parvenons pas encore à la caractériser de façon satisfaisante. 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 ont 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 l'a fait en 2016 pour le go. Certains programmes sont également capables de bluffer : ainsi, il y a quelques mois, le système Libratus, conçu à l'université Carnegie-Mellon, aux États-Unis, a battu quatre joueurs professionnels au poker dans sa version Texas Hold'em. Or le poker est un jeu complexe où il >