

En bref

> Nous interagissons de plus en plus avec des machines, ou des robots, qui nous parlent.	> Ces «agents conversationnels» simuleront des capacités qui auront l'air d'être élaborées sur le plan cognitif, mais aussi sur le plan émotionnel, alors qu'ils ne comprennent rien.	> Il est essentiel d'évaluer les capacités de tels systèmes, voire d'en réaliser un suivi : étant donné la faculté d'apprentissage de ces machines, leur comportement pourrait changer avec le temps.	> Il faut aussi étudier le comportement des humains vis-à-vis des « robots bavards », afin d'éviter certains abus et effets pervers, l'isolement social par exemple.
--	---	---	--

96

«Si vous souhaitez joindre le service commercial, tapez 1. Le service technique, tapez 2.» Cela vous dit quelque chose? Eh bien, l'époque où une demande de renseignements exigeait de taper sur les touches d'un ordinateur ou d'un téléphone avant d'obtenir une information ou de joindre le bon interlocuteur touche à sa fin. De plus en plus, nous utilisons des programmes informatiques auxquels on s'adresse en parlant naturellement, et qui nous renseignent aussi en langage naturel, que ce soit à l'écrit ou à l'oral. Connectés sur internet ou à une base de données, embarqués sur un ordinateur, un téléphone ou encore 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, élaboré à l'institut de technologie du Massachusetts (MIT) vers 1964-1966 par Joseph Weizenbaum : ce système répondait à l'écrit en simulant un psychothérapeute dont la stratégie était de répéter les propos du patient.

Si les *chatbots* ont longtemps été développés de manière à répondre à un sujet de discussion précis ou une tâche particulière, les *chatbots* dits à domaine «ouvert» sont conçus pour parler de tout. Ce qui sous-entend certaines capacités élaborées de la machine, ou plus exactement des logiciels dont elle est munie. Les *chatbots* étaient jusqu'à présent composés de plusieurs modules : de reconnaissance de la parole, de compréhension, de dialogue, de production et de synthèse de la parole. Les nouvelles générations de *chatbots* sont de plus en plus performantes grâce à l'évolution des techniques d'apprentissage des

machines, à la puissance des processeurs et à la taille des bases de données. Les modèles de traitement automatique de la langue les plus récents, appelés des «transformeurs», sont des réseaux de neurones qui, à partir de vastes corpus linguistiques, apprennent les régularités les plus saillantes, sans être influencés, par l'ordre des mots. Des *chatbots* sont maintenant capables d'aborder n'importe quel sujet sans contrainte, par exemple Blenderbot, le modèle conversationnel de Facebook. Ou encore Meena, système fondé sur l'apprentissage profond de bout en bout entraîné sur des dialogues entre humains, qui montre des performances dépassant tous les *chatbots* existants.

LE TEST DE TURING, INSUFFISANT POUR PLUSIEURS RAISONS

Au-delà des avancées scientifiques et techniques mises en œuvre dans les *chatbots*, 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 assistent et avec lesquelles nous dialoguons, il apparaît de plus en plus important d'évaluer de façon pertinente et continue, par des batteries de tests reproductibles, ces capacités ainsi que leur impact sur la relation entre humain et machine.

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» a pour objectif, selon son pionnier John McCarthy, de «doter des machines de systèmes informatiques ayant des capacités intellectuelles comparables à celles des hommes».

En 1950, pour établir si une machine est intelligente ou non, le mathématicien britannique Alan Turing proposait le «jeu de l'imitation». 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 décider qui est qui. Si la machine réussit à faire croire à l'examineur 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 (voir *Interro écrite pour l'intelligence artificielle*, par M. Mitchell, page 54) ? Le jeu de l'imitation proposé par Turing apparaît de plus en plus insuffisant, et ce pour plusieurs raisons. L'une d'elles est qu'il ne détermine pas directement les capacités, cognitives ou autres, du programme : il évalue seulement si ce programme peut simuler un comportement humain et tromper l'examineur. 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 *chatbot* ne fait que manipuler la syntaxe, et réussir ce jeu d'imitation ne prouverait donc rien. Les *chatbots* imitent le langage des humains, mais ils ne comprennent pas le sens de l'information qu'ils traitent.

Le test de Turing
a ses limites :
il évalue seulement
si le programme
peut simuler un
comportement humain et
tromper l'examineur

Le test de Turing est insuffisant pour une autre raison : il ne prend pas en compte les diverses et nombreuses facettes de l'intelligence humaine. On la réduit souvent à la capacité de mener un raisonnement abstrait, mais l'intelligence émotionnelle ou sociale est un autre aspect primordial de l'humain, qui requiert d'autres compétences. L'intelligence est un concept polysémique qui 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,



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