

> faut prendre des décisions fondées sur des informations incomplètes tout en affrontant les bluffs et autres ruses.

Quant aux performances des agents conversationnels, la possibilité de parler naturellement à ces systèmes afin qu'ils exécutent 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. Les assistants conversationnels ne gèrent pas encore bien un dialogue, n'ont pas de sens commun et ne comprennent pas les règles sociales; ils répondent à des questions sans se préoccuper de l'historique ni du contexte de la conversation, d'où un dialogue qui paraît souvent simpliste ou incohérent.

DES ROBOTS QUI BAVARDENT

Dans leur état actuel, ces systèmes repèrent des mots (voire des expressions entières) sémantiquement importants dans le discours de l'interlocuteur, afin de retrouver des réponses dont le schéma est programmé et qui permettent de mener le dialogue d'une manière plus ou moins intelligente, sans qu'une réelle compréhension du sujet et des phrases prononcées soit nécessaire.

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. Bien que l'objectif soit encore loin d'être atteint, ces entreprises ont développé des agents conversationnels qui répondent à des questions en langage naturel et dont certains rendent déjà service aux particuliers.

Ainsi, en 2011, le programme Watson d'IBM a gagné au jeu télévisé Jeopardy, une sorte de « Questions pour un champion », où il fallait répondre en langage naturel aux questions. Siri, installé sur les iPhone d'Apple, est capable de répondre à nos questions prononcées oralement, et une version plus perfectionnée sera bientôt disponible. Autre exemple: l'assistant vocal Alexa, intégré dans la petite enceinte cylindrique Echo d'Amazon, peut répondre à vos demandes, anticiper vos besoins, contrôler les autres objets connectés du domicile... Commercialisé depuis deux ans, ce système équipe déjà 20 % des foyers aux États-Unis!

La robotique conversationnelle se développe également. Les robots Nao et Pepper, de la société japonaise Softbank Robotics, le robot



L'humour fait partie des capacités que l'on souhaite donner aux machines

Buddy de la société française Blue Frog Robotics, pour ne citer qu'eux, sont capables de quelques dialogues préséncarés. Jibo, de la société américaine du même nom, est un robot de petite taille, sans bras ni jambes, dont l'écran rond simule une tête. Il est censé aider pour des tâches simples, mais il a aussi été créé pour faire office de compagnon amusant, qui connaît par exemple quelques plaisanteries.

Par ailleurs, les concepteurs d'agents conversationnels ajoutent de plus en plus 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. La machine peut ainsi prononcer des phrases telles que « J'aime



jouer» ou des expressions empathiques comme «Je comprends votre détresse, je vais vous aider». Il existe d'ailleurs des bibliothèques de réponses empathiques, comme Koko (<https://itskoko.com/>), issue des travaux du MIT et qui permet de choisir, suivant le contexte du dialogue, certaines réponses exprimant de l'empathie.

Dans tous ces développements autour de l'intelligence artificielle et des agents conversationnels, les fonctionnalités 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, la détection des émotions et l'humour font partie depuis 2001 des axes de recherche de mon équipe au Limsi, où l'on explore trois types de problématiques et de technologies: la détection des émotions, le raisonnement exploitant des informations de nature affective et la génération d'expressions d'ordre affectif.

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.

L'humour est en effet omniprésent dans les relations sociales humaines et constitue l'un des moyens les plus répandus pour produire un affect positif chez d'autres personnes. Plusieurs études ont montré que l'humour innocent augmente l'appréciation, renforce l'amitié, atténue le stress, encourage la créativité et améliore le travail d'équipe. Il peut aussi aider à détendre une situation, à surmonter des échecs ou encore à mettre une situation en perspective.

MACHINES EMPATHIQUES

Par exemple, quand le robot perçoit que vous êtes malheureux, il peut plaisanter pour vous faire rire et vous reconforter; s'il détecte chez vous de la colère, il peut vous rappeler avec humour que la situation n'est pas si dramatique. L'humour d'autodérision est également très utile lorsque le robot commet une erreur de reconnaissance ou d'interprétation, à condition bien sûr qu'il ait détecté son erreur: si le robot est capable de vous émouvoir et de vous divertir, il vous sera plus facile de vous attacher à lui et de lui pardonner ses imperfections.

Parler, répondre avec pertinence, apprendre, raisonner, détecter les émotions de l'interlocuteur, faire de l'humour sont ainsi des capacités que l'on attend d'une machine intelligente. Mais comment évaluer les capacités réelles d'une machine donnée? 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 aux robots ou aux agents conversationnels au travers d'algorithmes et de techniques d'apprentissage profond, procédures 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 d'expliquer 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 robots ou les agents conversationnels. 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 sa fin de vie? Suffit-il de tester la machine avant sa mise sur le marché si elle s'adapte en continu?

TESTER LES MACHINES TOUT AU LONG DE LEUR VIE

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 robot 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 était capable d'apprendre de ses interactions avec les internautes, puisqu'il adaptait son langage aux habitudes et autres 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!

Ces considérations et l'exemple malheureux de Tay montrent que tester les capacités des systèmes à intelligence artificielle est indispensable pour que l'on puisse interagir avec eux en toute confiance. D'où la question: comment faire évoluer le test de Turing pour évaluer les capacités des machines et leur code moral?

Comme on l'a vu, ce test éprouve la faculté d'une machine à imiter la conversation humaine. La mission du juge est de deviner, >