compréhension plus robuste, « un test de Turing du XXIe siècle ».

L'objectif était de « créer un programme informatique capable de regarder n'importe quelle émission de télévision, ou vidéo sur YouTube, et de répondre à des questions sur son contenu : "Pourquoi la Russie a-t-elle envahi la Crimée ?" ou "Pourquoi Walter White envisage-t-il de faire supprimer Jessie ?" » L'idée était d'éliminer la tricherie et de se concentrer sur le fait de savoir si les systèmes comprenaient vraiment les informations qui leur parvenaient. Programmer des ordinateurs pour leur faire faire des remarques désobligeantes ne nous rapprocherait guère de la véritable intelligence artificielle, mais les programmer pour qu'ils analysent davantage les textes ou paroles qu'on leur soumet pourrait faire avancer les choses.

Francesca Rossi, informaticienne à l'université de Padoue et présidente des Conférences internationales sur l'intelligence artificielle, a lu ma proposition et a suggéré que nous travaillions ensemble pour donner vie à ce test de Turing réactualisé. Nous sommes tombés d'accord pour enrôler Manuela Veloso, roboticienne à l'université Carnegie-Mellon et ancienne présidente de l'Association pour la promotion de l'intelligence artificielle, et nous nous sommes mis à réfléchir à trois. Nous avons commencé par rechercher un test unique capable de remplacer celui de Turing. Mais nous sommes très vite passés à l'idée de tests multiples, car de la même façon qu'il n'y a pas une seule épreuve d'athlétisme, il ne peut pas y avoir un test unique et absolu d'intelligence.

UN GROUPE DE TRAVAIL D'UNE CINQUANTAINE DE CHERCHEURS

Nous avons aussi décidé d'impliquer l'ensemble de la communauté des chercheurs en intelligence artificielle. En janvier 2015, nous avons rassemblé à Austin, dans le Texas, une cinquantaine des principaux chercheurs pour envisager une mise à niveau du test de Turing. >

Test 2

DES TESTS STANDARDISÉS

Il s'agirait de soumettre l'intelligence artificielle aux tests scolaires standardisés que les élèves du primaire et du collège passent à l'écrit sans être aidés. La méthode évaluerait la capacité d'une machine à relier des faits de façon innovante grâce à la compréhension sémantique. Tout comme le jeu d'imitation original de Turing, le principe est d'une grande simplicité. Il suffit de prendre n'importe quel examen standardisé suffisamment rigoureux (par exemple sous la forme d'un questionnaire à choix multiples), d'équiper la machine de moyens permettant d'acquérir les données du test (par exemple un système de traitement du langage naturel et la vision), et de la laisser faire.

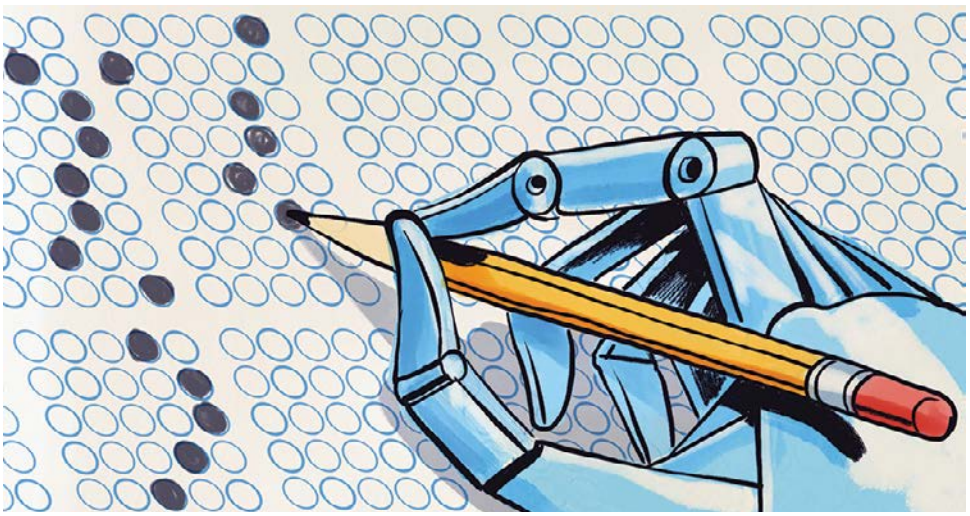
POUR : La polyvalence et le pragmatisme. Contrairement aux schémas de Winograd, les tests standardisés sont nombreux et bon marché. Et comme aucun de ces tests n'est adapté à la machine ou conçu spécifiquement, analyser leurs questions et y répondre nécessite une vaste connaissance du monde, polyvalente et pleine de bon sens.

CONTRE : Pas aussi protégé contre Google que les schémas de Winograd ; de plus, comme pour les humains, la capacité à réussir l'épreuve ne suppose pas nécessairement une réelle intelligence.

NIVEAU DE DIFFICULTÉ : Modérément élevé. Un système nommé Aristo, conçu par l'institut Allen pour l'intelligence artificielle, atteint une note moyenne de 75 % aux épreuves de sciences de CM1 qu'il n'a pas déjà rencontrées – mais uniquement sur des questionnaires à choix multiples sans figures. « Il n'y a à ce jour aucun système qui soit proche de réussir complètement un examen de sciences de CM1 », ont écrit Peter Clark et Oren Etzioni, de l'institut Allen, dans un article technique paru en 2016 dans *AI Magazine*.

À QUOI CELA SERT : À faire des vérifications objectives. « Au fond, nous voyons qu'aucun programme ne dépasse 60 % à un test de sciences de classe de quatrième, mais, en même temps, on lit dans certains journaux que le superordinateur Watson d'IBM fréquente la faculté de médecine et guérit du cancer », nous dit Oren Etzioni, PDG de l'institut Allen pour l'intelligence artificielle. « Soit IBM bénéficie d'une avance phénoménale, soit ses chercheurs s'avancent un peu trop. »

J. P.



Test 3

UN TEST DE TURING MATÉRIEL

La plupart des tests d'intelligence artificielle se limitent à l'aspect cognitif. Ce test ressemble davantage à des travaux pratiques : on demande à une intelligence artificielle de manipuler physiquement des objets réels de façon sensée. Le test comporterait deux axes, construction et exploration. Dans l'axe de construction, une intelligence artificielle physiquement matérialisée, (un robot, pour l'essentiel), tenterait de réaliser une construction à partir d'un lot de pièces en suivant des instructions verbales, écrites et illustrées (un peu comme l'assemblage d'un meuble en kit). L'axe d'exploration consisterait à demander que le robot conçoive des solutions pour relever des défis non limités et de créativité croissante à partir de briques-jouets (par exemple : « Construis un mur », « Construis une maison », « Ajoute un garage à la maison »). Chacun des axes se terminerait par un exercice de communication lors duquel le robot aurait à « expliquer » ses efforts. Ce test s'appliquerait à des robots isolés, à des groupes de robots ou à des robots coopérant avec des humains.

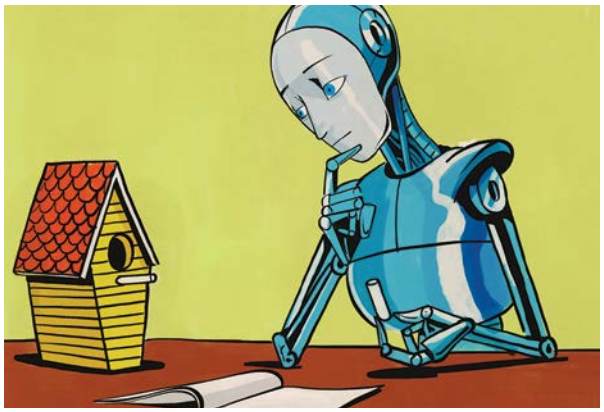
POUR : Ce test intègre des aspects de l'intelligence du monde réel, en particulier la perception et l'action, que l'on a généralement négligés ou pas assez étudiés. De plus, il est presque impossible à tromper. « Je ne vois pas comment on le pourrait, à moins que quelqu'un trouve le moyen de mettre sur Internet des instructions sur la façon de construire tout ce qui a jamais été construit », dit Charles Ortiz, de Nuance Communications.

CONTRE : Lourd, fastidieux et difficile à automatiser si l'on veut que les machines fassent physiquement leurs constructions. Et si on se limite à de la réalité virtuelle, « un roboticien dira que ce n'est encore qu'une approximation », explique Charles Ortiz. « Dans la vraie vie, quand vous saisissez un objet, il peut glisser, ou il peut y avoir un léger vent à prendre en compte. Dans un monde virtuel, il est difficile de simuler de façon fiable toutes ces nuances. »

NIVEAU DE DIFFICULTÉ : C'est de la science-fiction. Une intelligence artificielle matérialisée capable de manipuler correctement des objets et d'expliquer de façon cohérente ce qu'elle fait se comporterait comme un androïde de *La Guerre des étoiles*, très au-delà de l'état de l'art actuel. « Effectuer ces tâches de la même façon qu'un enfant le fait de façon routinière est un immense défi », selon Charles Ortiz.

À QUOI CELA SERT : À imaginer une trajectoire d'intégration de quatre aspects de l'intelligence artificielle – perception, action, cognition et langage – que les recherches spécialisées tendent à étudier séparément.

J. P.



> À l'issue d'une journée complète de présentations et de discussions, nous sommes tombés d'accord sur la nécessité de multiplier les épreuves à passer.

L'une de ces épreuves, le Winograd Schema Challenge (Défi du schéma de Winograd), qui porte le nom de Terry Winograd, pionnier de l'intelligence artificielle et mentor de Larry Page et Sergey Brin, les fondateurs de Google, soumettrait les machines à un test qui mêle compréhension du langage et bon sens (*voir l'encadré « Test 1 »*).

Il ne peut pas y avoir un test unique et absolu d'intelligence

Tous ceux qui ont un jour essayé de faire comprendre le langage à un ordinateur se sont vite rendu compte que pratiquement chaque phrase est ambiguë, et souvent de plusieurs façons. Mais notre cerveau est tellement efficace pour comprendre le langage qu'en général, nous ne nous en rendons pas compte. Prenez la phrase : « La grosse balle a défoncé la table parce qu'elle était en polystyrène. » Au sens strict, la phrase est ambiguë : le mot « elle » peut désigner la table comme la balle. N'importe quel auditeur humain comprendra que « elle » se réfère à la table. Mais cela nécessite d'associer à la compréhension du langage des notions de science des matériaux, ce qui n'est absolument pas à la portée des machines.

Trois experts, Hector Levesque, Ernest Davis et Leora Morgenstern, ont déjà développé un test autour de telles phrases, et la société Nuance Communications, spécialisée en reconnaissance vocale, offre un prix de 25 000 dollars au premier système qui fonctionnera.

Nous espérons faire participer de nombreux autres chercheurs. Une étape naturelle serait une épreuve de compréhension, qui évaluerait la capacité de la machine à comprendre des images, des vidéos, des enregistrements audio et des textes (*voir l'encadré « Test 4 »*). Charles Ortiz Jr, directeur du Laboratoire