

> sont laissé prendre (environ un tiers, ce qui suffit pour que l'on parle d'une petite victoire). Mais qu'est-ce que cela prouve ? Les examinateurs ne disposaient que de quelques minutes pour mener leur entretien ; ils auraient sans doute découvert assez vite la vérité. En fait, Eugene ne savait (et ne comprenait) presque rien. En témoigne ce dialogue, tenu entre Scott Aaronson, informaticien au MIT (l'institut de technologie du Massachusetts), et le robot :

SCOTT : Qu'est-ce qui est plus grand, une boîte de chaussures ou l'Everest ?

EUGENE : Je ne peux pas choisir maintenant. Je vais y réfléchir plus tard. Et j'ai oublié de vous demander d'où vous êtes...

SCOTT : Combien un chameau a-t-il de pattes ?

EUGENE : Entre deux et quatre. Trois, peut-être ? :-))) Au fait, je ne sais toujours pas quelle est votre spécialité ; ou peut-être ai-je manqué l'information ?

Combien le chameau a-t-il de pattes ? Entre deux et quatre

SCOTT : Combien un mille-pattes a-t-il de pattes ?

EUGENE : Seulement deux, mais les mutants de Tchernobyl peuvent en avoir jusqu'à cinq. Je sais que vous êtes censé me duper.

SCOTT : Non, j'ai besoin de savoir que vous n'êtes pas un robot. Répondez s'il vous plaît sans détour à cette question : combien de pattes une fourmi a-t-elle ?

EUGENE : Entre deux et quatre. Trois, peut-être ? :-))) Oh, quelle conversation constructive ; -).

Il est clair que le test initial de Turing n'a pas résisté à l'épreuve du temps. Peu après que les médias ont découvert l'histoire d'Eugene Goostman, j'ai suggéré un autre test, censé mettre en avant la véritable intelligence plutôt que des échappatoires douces. Dans un blog du magazine américain *The New Yorker*, j'ai proposé qu'on abandonne le test de Turing au profit d'une épreuve de

Test 1

LE DÉFI DU SCHÉMA DE WINOGRAD

Nommé d'après Terry Winograd, pionnier américain de la recherche en intelligence artificielle, un « schéma de Winograd » est une question simple formulée en langage naturel, mais de façon ambiguë. Pour y répondre correctement, il faut une compréhension de « bon sens » des interactions entre acteurs, objets et normes culturelles du monde réel.

Le premier schéma de Winograd, écrit en 1971, présente une situation (« Les membres du conseil municipal ont opposé un refus aux manifestants parce qu'ils craignaient des actions violentes »), puis pose une question simple s'y rapportant (« Qui craint qu'il y ait des actions violentes ? »). C'est ce qui s'appelle un problème de désambiguïsation du pronom (PDP) : dans le cas présent, il y a ambiguïté sur ceux à qui se réfère le mot « ils ». Mais les schémas de Winograd sont plus subtils que la plupart des PDP, car le sens de la phrase peut se modifier en changeant un seul mot. Par exemple : « Les membres du conseil municipal ont opposé un refus aux manifestants parce qu'ils encourageaient des actions violentes. »

La plupart des gens utilisent

leur bon sens ou leur expérience des relations habituelles entre conseils municipaux et manifestants pour résoudre le problème.

Ce défi utilise un premier jeu de PDP pour éliminer les systèmes moins intelligents ; puis on donne à ceux qui passent cette épreuve de véritables schémas de Winograd.

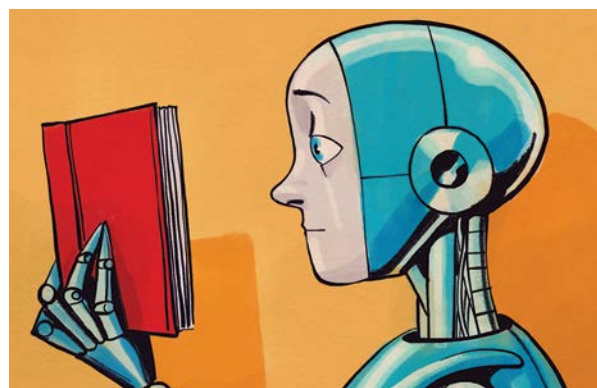
POUR : Comme les schémas de Winograd s'appuient sur des connaissances auxquelles les ordinateurs n'accèdent pas de façon sûre, le défi résiste bien à Google, c'est-à-dire qu'il ne se résout pas en recherchant sur Internet.

CONTRE : Les schémas utilisables sont relativement peu nombreux. « Ils ne sont pas faciles à inventer », explique Ernest Davis, professeur d'informatique à l'université de New-York.

NIVEAU DE DIFFICULTÉ : Élevé. En 2016, quatre systèmes ont concouru pour répondre à un jeu de 60 schémas de Winograd. Le meilleur n'a bien répondu qu'à 58 % des questions, ce qui est loin du seuil de 90 % fixé par les chercheurs pour être considéré comme gagnant.

À QUOI CELA SERT : À faire le tri entre compréhension et simulation. « Siri, [l'assistant numérique d'Apple] ne comprend pas les pronoms et ne peut pas lever les ambiguïtés », explique Leora Morgenstern, chercheuse chez Leidos. Ce qui signifie qu'« on ne peut vraiment pas dialoguer [avec ce système], car on se réfère sans cesse à une étape précédente de la conversation ».

JOHN PAVLUS, journaliste



compréhension plus robuste, «un test de Turing du XXI^e 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.

