

Test 4

L'I-ATHLON

Dans une batterie de tests partiellement ou complètement automatisés, on demande à une intelligence artificielle de résumer le contenu d'un fichier audio, de restituer le scénario d'une vidéo, de traduire du langage naturel à la volée et d'exécuter d'autres tâches encore. L'objectif est d'établir un score objectif d'intelligence.

La caractéristique de cette méthode est l'automatisation de la conduite du test et de sa notation, sans contrôle humain. Retirer les humains du processus d'évaluation de l'intelligence d'une machine peut faire sourire, mais Murray Campbell, chercheur en intelligence artificielle chez IBM et membre de l'équipe qui a développé Deep Blue, affirme que c'est nécessaire pour en garantir l'efficacité et la reproductibilité.

Une notation de l'intelligence artificielle qui serait déterminée par un algorithme permettrait aussi aux chercheurs de ne plus utiliser comme étalon l'intelligence humaine « avec tous ses biais cognitifs », remarque Murray Campbell.

POUR : L'objectivité, en théorie du moins. Une fois que le jury de l'I-athlon aurait déterminé comment noter chaque test et pondérer les résultats, ce seraient les ordinateurs qui noteraient et pondéreraient. L'appréciation des résultats serait aussi tranchée que l'examen d'une photo d'arrivée d'épreuve olympique. La diversité des tests aiderait aussi à identifier ce que les chercheurs d'IBM nomment des « systèmes à intelligence large ».

CONTRE : Un risque de résultats incompréhensibles. Les algorithmes des I-athlon pourraient décerner des notes élevées à des systèmes d'intelligence artificielle fonctionnant d'une façon que les chercheurs ne comprendraient pas complètement. « Il est possible que certaines décisions de systèmes évolués d'intelligence

artificielle seront très difficiles à expliquer [aux humains] de façon concise et compréhensible », admet Murray Campbell. Ce problème d'opacité commence déjà à poser des difficultés aux chercheurs qui travaillent sur les réseaux de neurones convolutifs (ou systèmes d'« apprentissage profond »).

NIVEAU DE DIFFICULTÉ : Cela dépend. Quelques-uns des systèmes actuels pourraient bien réussir certaines épreuves d'I-athlon, telles que

comprendre une image ou traduire d'une langue à l'autre. D'autres épreuves, telles que l'explication du contenu narratif d'une vidéo ou le tracé d'un diagramme à partir d'une description verbale, relèvent encore de la science-fiction.

À QUOI CELA SERT : À réduire l'impact des biais cognitifs humains sur la mesure de l'intelligence des machines et à quantifier les performances de celles-ci, plutôt qu'à simplement les identifier.

J. P.



BIBLIOGRAPHIE

G. Marcus et al. (éd.), **Beyond the Turing Test**, numéro spécial de *AI Magazine*, vol. 37(1), 2016.

G. Marcus, **What comes after the Turing test ?** *The New Yorker*, en ligne le 9 juin 2014.

Défi du schéma de Winograd : <http://commonsensereasoning.org/winograd.html>

A. M. Turing, **Computing machinery and intelligence**, *Mind*, vol. 59, n° 236, pp. 433-460, 1950.

d'intelligence artificielle et de traitement du langage chez Nuance Communications, a proposé une épreuve de construction qui évaluerait la perception et l'action physique, deux importantes composantes de l'intelligence comportementale qui sont totalement absentes du test de Turing initial (voir l'encadré « Test 3 »). Et Peter Clark, de l'institut Allen pour l'intelligence artificielle, aux États-Unis, a suggéré de faire passer à des machines les mêmes épreuves standardisées de sciences et autres matières que celles passées par les écoliers (voir l'encadré « Test 2 »).

Parallèlement, les participants à la conférence ont débattu des principes généraux que devrait respecter un bon test. Guruduth Banavar et ses collègues d'IBM, par exemple, ont souligné que les tests eux-mêmes devraient être générés par ordinateur. Stuart Shieber, de l'université Harvard, a mis l'accent sur la transparence : pour faire progresser le domaine, les récompenses devraient être réservées aux systèmes ouverts – accessibles à l'ensemble de la

communauté de l'intelligence artificielle – et reproductibles.

Quand des machines seront-elles capables de relever les défis que nous leur avons fixés ? On l'ignore. Mais les chercheurs prennent déjà au sérieux ces épreuves.

LE TEST DE TURING ÉTAIT UN BON DÉBUT

Par exemple, un robot qui aurait passé l'épreuve du défi de la construction pourrait monter des camps temporaires pour réfugiés. Une machine ayant réussi le Défi du schéma de Winograd et à un examen de biologie de CM1 nous rapprocherait de machines capables d'assimiler l'immense littérature médicale et de proposer de nouveaux traitements efficaces contre des cancers ou d'autres maladies graves.

Le domaine de l'intelligence artificielle, comme les autres, a besoin d'objectifs clairs. Le test de Turing était un bon début ; il est maintenant temps de construire une nouvelle génération de défis. ■





L'ESSENTIEL

> Nous interagissons de plus en plus avec des machines, ou des robots, qui nous parlent.

> Ces « agents conversationnels » ont, ou auront, des capacités élaborées sur le plan cognitif, mais aussi sur le plan émotionnel.

> 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.

L'AUTEURE



LAURENCE DEVILLERS
professeure en intelligence artificielle à l'université Paris-Sorbonne
et chercheuse
au Limsi-CNRS, à Orsay

Tester les robots pour bien vivre avec

DIALOGUER, APPRENDRE, DÉTECTER L'ÉTAT ÉMOTIONNEL DE SES INTERLOCUTEURS, FAIRE DE L'HUMOUR... : CERTAINES MACHINES LE FONT DÉJÀ. MAIS IL SERA INDISPENSABLE DE BIEN ÉVALUER CES CAPACITÉS SI L'ON VEUT UTILISER ET CÔTOYER DES ROBOTS EN TOUTE CONFIANCE.

L'époque où une demande de renseignements par téléphone exigeait de taper sur des touches de son appareil avant d'obtenir l'information enregistrée ou de joindre le bon interlocuteur humain s'achève. De plus en plus, nous tombons sur des répondeurs « intelligents » auxquels on s'adresse en parlant naturellement, et qui nous renseignent aussi en langage naturel. Et la conversation avec une machine est loin de se limiter à des échanges téléphoniques. On peut aujourd'hui parler à son téléphone portable ou à son ordinateur pour lui demander de retrouver l'adresse d'un ami,

d'envoyer un courriel, d'ouvrir telle ou telle application, de se connecter à tel ou tel site, voire de converser, tout simplement.

Plus généralement, interagir et dialoguer avec une machine, qu'il s'agisse d'un téléphone, d'un appareil domestique ou d'un robot industriel devient de plus en plus banal. Ce qui sous-entend certaines capacités élaborées de la machine, ou plus exactement des logiciels dont elle est munie. Les systèmes les plus avancés de ce type sont aujourd'hui composés de plusieurs modules, tels que des modules de reconnaissance de la parole, de reconnaissance de quelques expressions émotionnelles, de compréhension, de dialogue, de génération de >