



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, ➤

➤ d'après les messages échangés, qui est, de l'un ou de l'autre de ces interlocuteurs, une machine. Le contenu des messages est libre : il est possible de parler de tous types de sujet. Cependant, chacun des éléments du test de Turing peut être remis en question et, aujourd'hui encore, il faut se demander quels sont les biais à éviter lorsqu'on effectue un tel test. Jean-Paul Delahaye, chercheur en informatique à Lille (et auteur de la rubrique « Logique & calcul » de *Pour la Science*), explique par exemple qu'une simple question de calcul est susceptible de démasquer l'interlocuteur-programme : l'humain sera en effet mauvais pour effectuer correctement un calcul laborieux, alors que la machine va exceller. Mais, pourrait-on rétorquer, rien n'empêche de programmer la machine pour qu'elle fasse exprès de se tromper...

Toujours est-il que faire passer le test de Turing reste un graal pour tous les concepteurs d'agents conversationnels. Depuis 1990, il existe même une compétition annuelle, le prix Hugh Loebner, qui couronne les systèmes satisfaisant le mieux les critères du test de Turing, c'est-à-dire avec lesquels il est le plus difficile de déterminer s'il s'agit d'un humain ou d'une machine. Mais, répétons-le, aucune intelligence artificielle n'a encore passé avec succès le test de Turing. Les meilleurs programmes n'effectuent qu'un simulacre de conversation qui, malgré de récents progrès (systèmes capables de répondre dans des domaines ouverts, adaptables à différents interlocuteurs, dotés de certains traits de personnalité...), fait rarement illusion plus de quelques minutes.

SURVEILLER LA COÉVOLUTION DES HUMAINS ET DES MACHINES

Il reste donc à trouver des façons pertinentes d'évaluer les diverses capacités, tant cognitives qu'émotionnelles, des machines. Pour ce qui est des performances cognitives, de nombreux chercheurs imaginent des scénarios inspirés du test de Turing, mais composés de plusieurs épreuves dans des contextes différents. Dans des versions scolaires, les machines passent des examens, comme des écoliers. D'autres tests portent sur la capacité à démêler des ambiguïtés du langage naturel, sur la préhension d'objets, etc. (*voir l'article de Gary Marcus*). Quant aux capacités émotionnelles, tout ou presque reste à faire : la conception de tests sur cet aspect constitue un nouveau chantier de la recherche en intelligence artificielle.

Ainsi, avec robots munis de facultés que l'on considérerait comme le propre des humains il n'y a pas si longtemps, des variantes du test de Turing ou des batteries de tests sont à développer – qui pourront être à la base d'une sorte

JIBO, UN ROBOT BAVARD

Sur son écran rond, ce petit assistant conversationnel illustre une histoire qu'il est en train de raconter. Ce petit robot de forme épurée aide à effectuer des tâches simples et a été aussi conçu pour amuser : Jibo peut faire quelques plaisanteries.



d'autorisation de mise sur le marché et de contrôle technique pour robots tout au long de leur vie. C'est un axe prometteur des recherches actuelles.

Enfin, il est nécessaire de surveiller la coévolution des humains et des machines dans leur interaction. La vie des robots émerge à travers nous et s'inscrit dans notre monde intérieur. La simple présence d'un robot assis près peut déjà nous conduire à nous interroger : « Que pense-t-il ? » Les robots pourront nous manipuler ou nous mentir pour notre bien-être, mais encore faudra-t-il s'assurer que c'est bien le cas. Il faut se préoccuper de l'évolution, à long terme, du comportement des humains vis-à-vis des agents conversationnels et des robots, afin d'éviter l'isolement social, un attachement démesuré à la machine ou des manipulations abusives. C'est aussi dans cette direction que doit évoluer le test de Turing : non seulement tester les capacités des machines, mais aussi tester leur propension à modifier nos croyances et notre comportement vis-à-vis d'elles.

Ajoutons, pour finir, qu'une bonne information sur les capacités de ces assistants conversationnels, sur la façon dont ils fonctionnent, sur les biais qu'ils peuvent présenter, permettrait aux utilisateurs de prendre plus de recul et de se comporter de façon adéquate vis-à-vis d'eux. Cela fait aussi partie des défis sociétaux à relever pour bien vivre avec les machines. ■

BIBLIOGRAPHIE

L. Devillers,
Des robots et des hommes : mythes, fantasmes et réalité, Plon, 2017.

L. Devillers,
Rire avec les robots pour mieux vivre avec, *Journal du CNRS*, juin 2015 (<http://bit.ly/1Lnhgp2>).

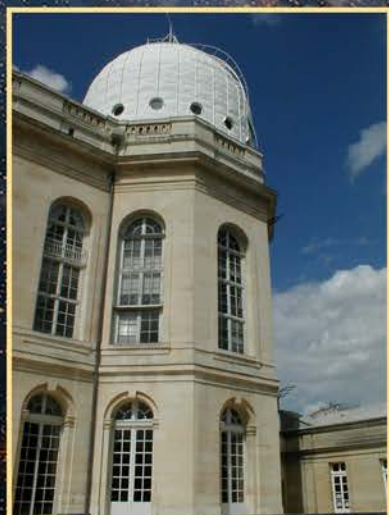
Vivre avec les robots, vidéo réalisée par le Limsi-CNRS : www.youtube.com/watch?v=p1ID-gvUnWs

Les robots en quête d'humanité, *Dossier Pour la Science*, n° 87, avril-juin 2015.

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

LUMIÈRES SUR L'UNIVERS

FORMATION EN LIGNE
EN ASTROPHYSIQUE - NIVEAU (L1-M1) - DIPLOMANTE



LES PARCOURS

Des étoiles aux planètes : niveau L1-L2

Cosmologie et Astrophysique extragalactique : niveau L2

Mécanique Céleste : niveau L3

Fondamentaux pour l'astronomie et l'astrophysique : niveau L3

Sciences Planétaires : niveau L3

Fenêtres Sur l'Univers : niveau M1

Instrumentation, chaîne de mesure et projets : niveau M1

Ouvert à tous (niveau bac minimum) : passionnés, animateurs scientifiques, étudiants ...

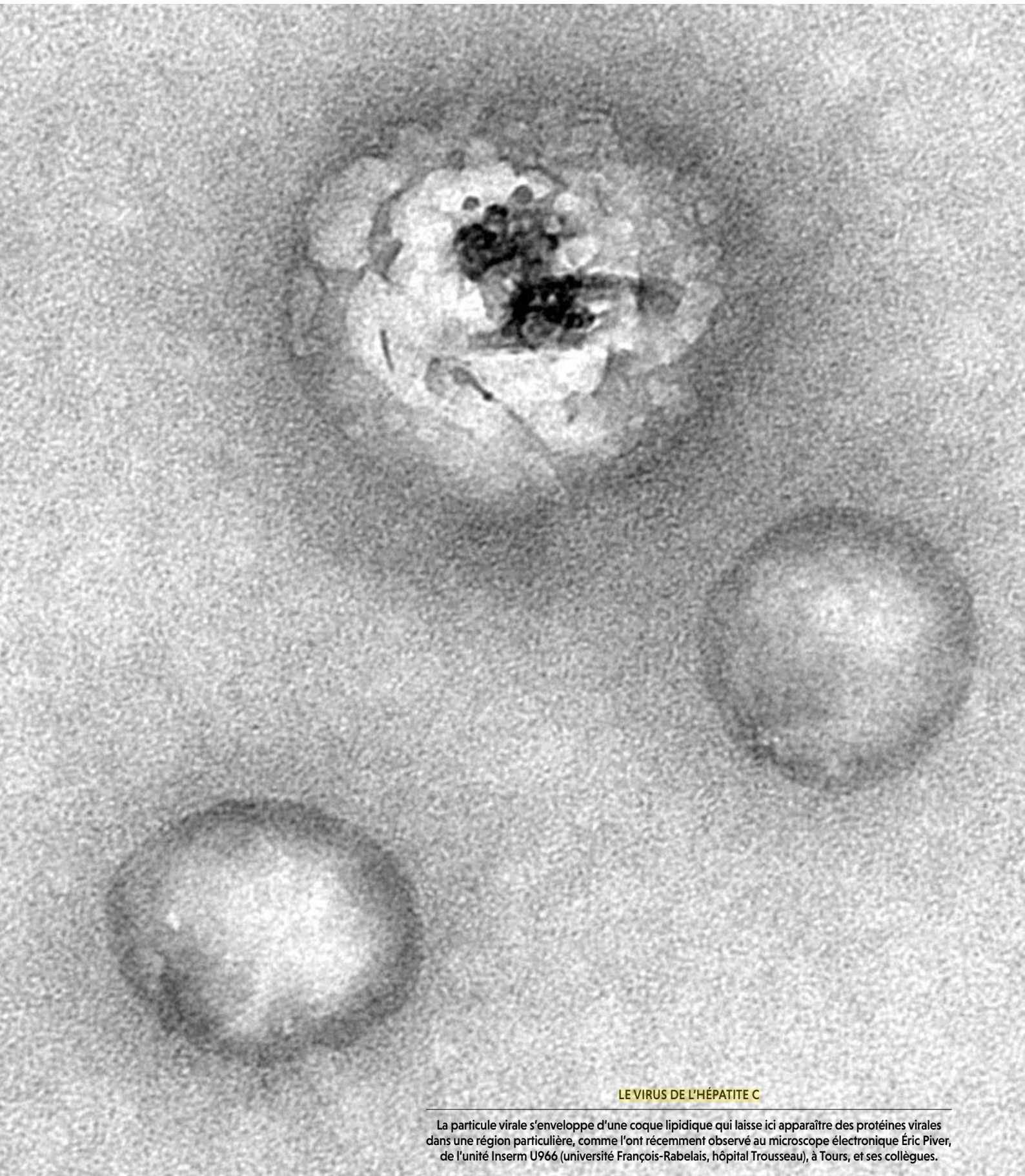
Cours et exercices interactifs assurés par des astronomes professionnels.

Formation pouvant être validée sous forme de Crédits de Licence ou de Master (ECTS)

Inscriptions 2017-2018 ouvertes jusqu'au 19 juin 2017 !

<https://media4.obspm.fr/dulu>

contact.dulu@obspm.fr



LE VIRUS DE L'HÉPATITE C

La particule virale s'enveloppe d'une coque lipidique qui laisse ici apparaître des protéines virales dans une région particulière, comme l'ont récemment observé au microscope électronique Éric Piver, de l'unité Inserm U966 (université François-Rabelais, hôpital Trousseau), à Tours, et ses collègues.