



LA MARCHÉ DES SCIENCES

DÉCOUVERTES, INVENTIONS, AVENTURES SAVANTES AU FIL DE L'HISTOIRE

AURÉLIE LUNEAU
CHAQUE JEUDI
16H - 17H



© RADIO FRANCE / CHRISTOPHE ABRAMOWITZ

en partenariat avec **POUR LA
SCIENCE**

Écoute, réécoute, podcast
franceculture.fr



Robot, as-tu du cœur ?

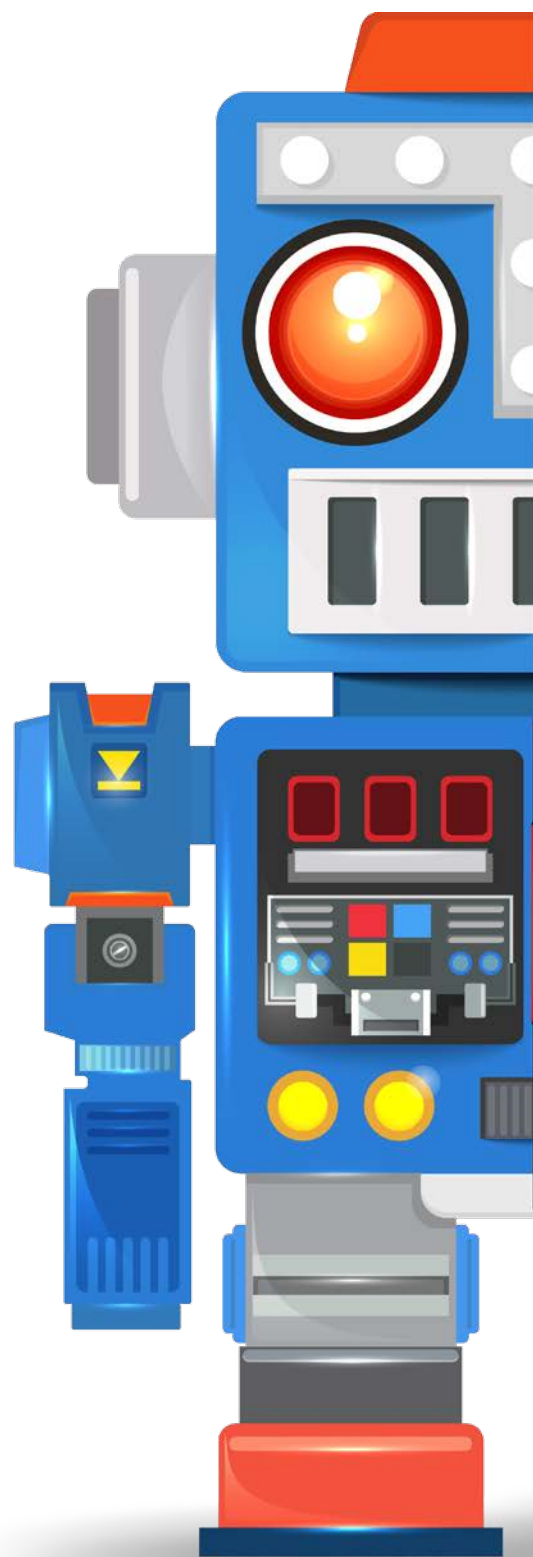
Pascale Fung

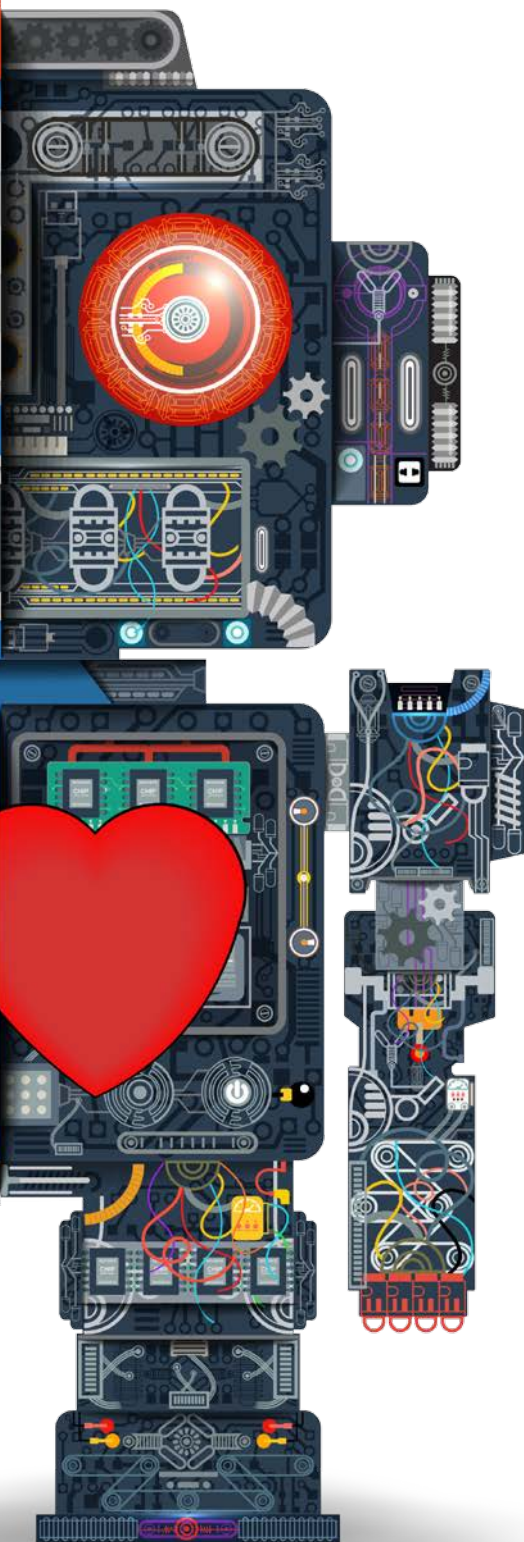
**Avant de côtoyer des robots au quotidien,
il va nous falloir leur apprendre à comprendre
les émotions humaines... Ou, du moins, à les imiter.**

Les robots intelligents sont aujourd'hui répandus et ils sont parfois étranges. Demandez à votre téléphone : « Siri, est-ce que tu m'aimes ? ». Il vous répond : « Disons que vous avez toute mon estime... ». Siri est l'agent conversationnel développé par la société Apple pour aider nos iPhones à nous trouver un restaurant, une information ou encore à appeler un ami lorsque nous avons les mains prises. Tous les téléphones de dernière génération intègrent désormais de tels agents, mais la présence de ces derniers dans nos vies n'est pas nouvelle. Depuis longtemps, que vous appeliez la Fnac ou la SNCF, vous avez d'abord affaire à une machine parlante.

Pour autant, ces intelligences artificielles ne répondent pas toujours de façon... intelligente, notamment parce que les logiciels de reconnaissance de la parole ne sont pas encore assez performants ; et elles ne saisissent pas non plus les émotions, l'humour, les sarcasmes ou encore l'ironie. Or nous allons avoir de plus en plus souvent affaire à de telles machines ; et bientôt, certaines auront des corps ! Nous connaissons déjà les aspirateurs robots, nous aurons probablement bientôt des robots infirmiers... C'est pourquoi il est important que ces intelligences artificielles progressent dans leur compréhension de nos mots et qu'elles nous « comprennent ». Il faut donc parvenir à les doter de moyens efficaces d'appréhender, de partager et de traiter les émotions humaines : en deux mots, nous avons besoin de robots empathiques.

Quelques robots imitant l'émotion humaine sont déjà sur le marché : Pepper, par exemple, est un petit compagnon humanoïde





L'ESSENTIEL

- Les gens interagissent de plus en plus souvent avec des machines commandées par la voix.
- Ils attendent d'elles spontanément qu'elles saisissent leurs émotions, y compris les plus complexes, tels l'humour ou la moquerie.
- C'est pourquoi des chercheurs travaillent à doter les intelligences artificielles de logiciels capables d'extraire des émotions de la voix et des mots.
- Même si ces recherches débutent, de premiers robots empathiques existent et s'améliorent de jour en jour.

fabriqué par la société française Aldebaran Robotics afin d'accueillir les visiteurs dans les points de vente de l'opérateur téléphonique japonais Softbank Mobile. Autre cas, Jibo est un assistant personnel domestique de 2,7 kilogrammes. La « robotique empathique » n'en est encore qu'à sa naissance, mais les outils et les algorithmes qui amélioreront spectaculairement les machines en les dotant d'une « empathie artificielle » de qualité sont en cours de développement.

C'est ce sur quoi travaillent les équipes du laboratoire que je dirige à l'université des sciences et technologies de Hong Kong. Mon intérêt pour les robots empathiques est né il y a six ans, lorsque j'ai lancé mon équipe dans la mise au point de l'équivalent chinois de Siri. D'emblée, la façon dont les utilisateurs développent naturellement des réactions émotionnelles vis-à-vis des assistants personnels et leur frustration lorsque ces machines n'arrivent pas à les comprendre m'a fascinée. J'ai vite pris conscience que la clé de la conception de machines empathiques est la mise au point d'algorithmes efficaces de reconnaissance de la parole. Or je venais justement de passer vingt-cinq ans de ma carrière à en développer...

Le module d'empathie

Toute machine intelligente est, par essence, un ensemble de composants logiciels, tels des bases de données, des fichiers de configuration, des programmes... Chacune des composantes d'un tel « système logiciel » remplit une fonction unique. Ainsi, un robot intelligent pourra être doté d'un module de traitement de la parole, d'un programme de reconnaissance des objets présents dans les vidéos qu'il a enregistrées, etc. Toutefois, le « cœur » d'un robot empathique sera le composant logiciel conçu pour analyser les signaux faciaux et acoustiques des humains, puis le contenu du discours afin d'en dégager les émotions humaines... C'est ce « module d'empathie » qui traitera les informations émises par les humains et dictera au robot ses réactions.

Ces informations sont complexes. Deux humains qui communiquent envoient et reçoivent de nombreux signaux émotionnels. Qu'ils en aient conscience ou pas, chacun

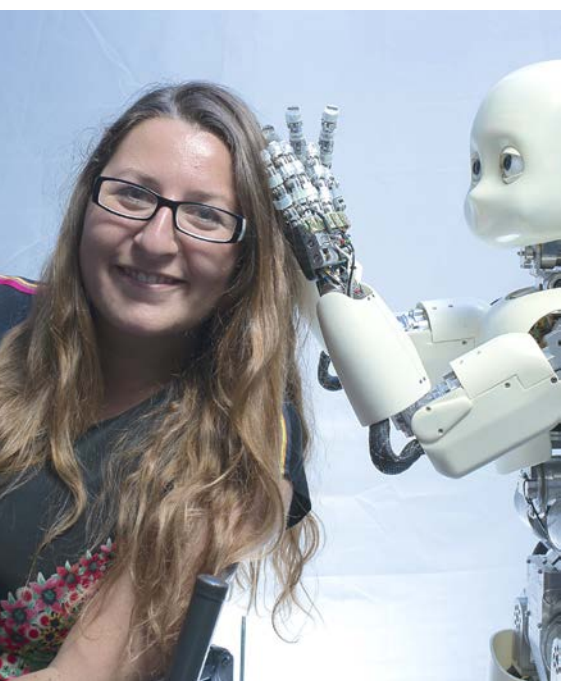
© Doggygraph / Shutterstock.com

■ L'AUTEURE



Pascale FUNG est professeure d'électronique et de génie informatique à l'université des sciences et technologies de Hong Kong.

LES ROBOTS EMPATHIQUES
seront-ils un jour sympathiques ?
La réponse dépendra de la qualité atteinte dans le développement de logiciels d'empathie artificielle et du sens du mot « amitié ». Pour certains, l'amitié avec une machine est par définition impossible. Pour d'autres, elle sera évidente. Comment vivrons-nous avec nos amis les gentils robots ?



les exploite pour saisir l'état émotionnel de l'autre et pour lui faire comprendre le sien. Ces signaux sont notamment les mimiques faciales, tous les éléments du langage corporel (attitude, gestes...), le ton de la voix, mais aussi, évidemment, le contenu du discours. La mise au point d'un module empathique implique donc d'identifier les traits de la communication humaine exploitables par les machines, puis d'« entraîner » des algorithmes d'apprentissage automatique.

Du moins est-ce ainsi que s'y est prise mon équipe pour développer des machines capables non seulement d'identifier les mots et leur sens, mais aussi d'extraire plusieurs caractéristiques fondamentales du signal acoustique qu'est la voix. Ainsi, même si nous la considérons peu sous cet angle, la communication entre humains est aussi une forme de traitement du signal. Le cerveau est entraîné à détecter les émotions convoyées par la voix à partir des caractéristiques des signaux acoustiques qui leur sont associés, qu'il s'agisse de stress, de joie, de peur, de colère, de dégoût... Une personne joyeuse parlera plus vite et sur un ton aigu ; stressée, elle adoptera au contraire une voix plate et sèche. Doté d'un algorithme d'apprentissage automatique convenablement entraîné, un ordinateur peut, comme un cerveau, détecter ce genre de caractéristiques, un peu comme un détecteur de mensonges mesure la tension artérielle, les pulsations cardiaques et la conductivité de la peau.

Pour obtenir une machine ayant, par exemple, la capacité de détecter le stress, nous avons supervisé l'entraînement d'un algorithme en lui faisant analyser des enregistrements de la voix de personnes stressées.

Déceler la détresse dans la voix

C'est ainsi que nous avons commencé à apprendre aux machines à déceler la détresse dans des extraits vocaux d'étudiants de notre université, lesquels s'y connaissent puisque, entre eux, ils surnomment notre établissement l'« université du stress et de la tension de Hong Kong ». Pour obtenir ces extraits, nous leur avons posé une série de 12 questions de plus en plus stressantes et avons obtenu ainsi un premier corpus de signaux vocaux stressés en anglais, mandarin et cantonais. Dix heures d'enregistrement ont suffi pour que nos algorithmes reconnaissent le stress dans 70 % des cas où il est présent

dans la voix. Ce niveau de performance est comparable à celui des auditeurs humains.

Pendant que nous effectuions ces travaux, une autre équipe du groupe apprenait aux machines à reconnaître l'humeur exprimée par la musique d'un morceau de musique de variété, indépendamment des paroles. Contrairement à l'émotion, l'humeur est une ambiance qui reste généralement pendant toute la durée d'un morceau. Nos chercheurs ont commencé par rassembler quelque 5000 morceaux en tous genres dans les principales langues européennes et asiatiques. Les musicologues avaient d'ailleurs déjà classé plusieurs centaines de ces morceaux en 14 catégories d'humeur.

À l'aide d'ordinateurs, nous avons extrait de ces chansons 1000 caractéristiques fondamentales – par exemple l'énergie ou des propriétés acoustiques telles la fréquence fondamentale, les harmoniques... –, puis nous avons utilisé ce corpus pour entraîner 14 programmes « classificateurs », chacun spécialisé dans une humeur particulière. L'un d'eux décèle par exemple l'humeur mélancolique d'une chanson, un autre l'humeur joyeuse, etc. Comme ces 14 classificateurs peuvent être employés en même temps, leurs résultats peuvent être combinés, mais il arrive qu'ils se contredisent, par exemple que le détecteur d'humeur joyeuse trouve joyeuse une chanson mélancolique. Dans un tel cas, nous le réentraînons au cycle d'apprentissage suivant. D'anomalie en anomalie, nous avons ainsi fini par rendre performant le système formé des 14 détecteurs d'humeur. Nous avons ainsi obtenu une machine capable d'extraire l'humeur de n'importe quel morceau de musique juste en « l'écouter ».

Depuis, nous nous sommes appuyés sur ces recherches pour fonder avec d'anciens étudiants la société Ivo Technologies de développement de machines empathiques utilisables à domicile. *Moodbox*, son premier produit, sera un centre de contrôle d'ambiance, qui réglera la musique et la lumière de chaque pièce en fonction des émotions de ses utilisateurs.

Toutefois, pour saisir l'humour, le sarcasme, l'ironie et d'autres caractéristiques complexes de la communication verbale interhumaine, une machine aura besoin de plus que juste un logiciel de reconnaissance des émotions véhiculées dans le son. Il lui faudra aussi comprendre la signification sous-jacente d'un discours et en comparer le contenu avec l'émotion qui le baigne.

Depuis les années 1980, des chercheurs ont développé des systèmes perfectionnés de reconnaissance de la parole, de sorte que cette technologie atteint sa maturité. Pour autant, il y a loin entre transcrire un discours et le comprendre ! Songez à la complexité de la chaîne d'événements cognitifs, neurologiques et musculaires qui se combinent quand deux personnes se parlent. L'un des interlocuteurs formule ses pensées, choisit des mots puis articule, tandis que l'autre reçoit les sons et les autres signaux envoyés, avant d'y déchiffrer le message qui y est enfoui.

La chaîne qui s'établit lors d'un dialogue entre un humain et une machine est plutôt la suivante : les ondes sonores associées à la parole sont d'abord numérisées, puis de nombreux paramètres caractérisant le discours sont extraits de la série numérique obtenue ; ensuite, un logiciel de reconnaissance de la parole convertit ces paramètres en mots et un décodeur sémantique transforme la série de mots ainsi extraite en sens.

Au début de nos recherches sur les robots empathiques, nous nous sommes aperçus que des algorithmes similaires à ceux conçus pour analyser les émotions exprimées par des utilisateurs dans leurs commentaires en ligne pourraient nous aider. Ces algorithmes d'apprentissage automatique traquent la présence, dans leur contenu, de signaux révélateurs de la présence d'émotions. Des mots-clés tels que « chagrin » et « crainte » suggèrent la solitude. L'utilisation répétée d'expressions familières (« *Let's go!* », par exemple) révèle qu'une chanson est énergique. L'information donnée par le style du discours est aussi à prendre en compte : les réponses d'une personne sont-elles assurées et claires, ou hésitantes, parsemées de silences et de formulations non compromettantes ? Élaborées et détaillées, ou brèves et sèches ?

Au cours de nos recherches sur la reconnaissance de l'humeur dans la musique, nous avons entraîné des algorithmes à repérer des signaux émotionnels dans les paroles. Au lieu d'extraire des signatures sonores de chaque morceau de musique, nous avons tiré, à partir des paroles, des séries de mots – des *n*-grammes – dont nous nous sommes servis pour entraîner 14 classificateurs, chacun responsable de la détection d'une des 14 humeurs définies précédemment.

« Nous avons obtenu une machine capable d'extraire l'humeur de n'importe quel morceau de musique »

Nous avons aussi défini ce que nous nommons des « étiquettes de partie de discours », que nous avons prises en compte dans nos définitions des signatures de telle ou telle humeur. Sur la base d'une collection suffisante de jeux de ces paramètres extraits de discours humain, un ordinateur peut inférer des approximations statistiques des règles grammaticales de n'importe quelle langue. Ce sont ces règles élaborées au cours d'un entraînement qui permettent à des programmes tel *Siri* de comprendre à leur manière un discours ou encore à un programme tel *Google Translate* de traduire un texte.

Dès qu'une machine peut comprendre le contenu d'un discours, elle peut comparer ce contenu avec la façon dont il est émis. Un algorithme pourra détecter la contradiction entre le sens d'un discours et les signaux émotionnels qui l'accompagnent. C'est le cas, par exemple, quand la phrase « Je suis si heureux de passer le week-end à travailler » est accompagnée d'un soupir. Dans un tel cas, l'ordinateur calculera la probabilité que la déclaration humaine soit en réalité un sarcasme.

Intentions complexes

Afin de détecter les intentions plus complexes, une machine capable d'appréhender le contenu d'un discours et les émotions qui le baignent pourra aussi associer les informations recueillies à des données extérieures susceptibles d'aider à son interprétation. Si quelqu'un s'exclame « J'ai faim ! », elle tiendra compte, pour déterminer la meilleure réponse à apporter, de l'heure ou du lieu ainsi que de sa mémoire des usages et préférences de la personne. Au milieu du Sahara, elle s'abstiendra de proposer la recherche d'une « pizzeria dans le quartier » ; à la maison à l'heure du déjeuner, en revanche, elle pourrait répondre : « Souhaitez-vous que je vous prépare un en-cas ? »

Début 2015, des étudiants et des chercheurs postdoctoraux de mon laboratoire ont commencé à réunir nos différents modules de reconnaissance de la parole et de reconnaissance de l'émotion pour créer une machine empathique prototype que nous appelons *Zara la Supergirl*. Il a fallu des centaines d'heures de données pour entraîner *Zara*, mais aujourd'hui, ce programme tourne

Quels obstacles à l'insertion de robots en société ?

Membre de l'équipe Larsen (de l'anglais *Life-long Autonomy and Interaction Skills for Robots in a Sensing Environment*) à Nancy, Serena Ivaldi travaille sur les interactions humains-robots. Elle nous parle des principaux obstacles déjà identifiés s'opposant à l'acceptation des robots par les humains.



Serena IVALDI, roboticienne, travaille chez Inria dans l'équipe Larsen, à Nancy.

POUR LA SCIENCE

Pensez-vous que les robots vont se multiplier dans la société ?

Serena IVALDI : Il est certain que ce phénomène est pour bientôt. La situation est comparable à celle qui a précédé l'introduction des ordinateurs et des téléphones portables : leur usage par chacun a paru superflu jusqu'à ce qu'il s'impose de façon explosive, et aujourd'hui plus personne ne s'interroge sur l'utilité de ces technologies.

PLS

Comment testez-vous l'acceptation des robots par les humains ?

S.I. : Nous étudions d'abord les interactions entre robots et humains au sein de notre laboratoire en faisant appel à des personnes ordinaires, puis nous recommençons dans le milieu où les robots doivent être insérés, où les gens ont des particularités. Par exemple, nous étudions l'insertion de robots aides soignants dans un établissement d'hébergement pour personnes âgées dépendantes.

PLS

Quels sont les principaux obstacles à l'insertion de robots dans la société ?

S.I. : Il y en a de plusieurs sortes. À ce stade, le principal obstacle est simplement le coût prohibitif des robots ! Sinon, le problème principal est qu'il nous faut apprendre à bien organiser les interactions des personnes avec les robots que l'on va introduire. Un robot peut être vu comme un ordinateur doté d'un corps. Toutefois, quand un ordinateur est personnifié par un corps, les interactions humains-robots deviennent comparables aux interactions humaines, qui sont multimodales et complexes. En particulier, un corps c'est physique, et cela entraîne des interactions physiques !

PLS

Quels éléments interviennent au cours d'une interaction humain-robot ?

S.I. : Quand un humain est en présence d'un robot – surtout humanoïde – maints détails influencent profondément la façon dont il va tenter de communiquer avec lui : sa distance par rapport à lui ; l'apparence de la machine ; son « regard » ; les attentes de l'humain, notamment sur le rôle de la machine ; les forces développées par chacun d'eux dans la situation ; etc.

PLS

De quelles forces parlez-vous ?

S.I. : Traditionnellement, les robots industriels étaient séparés des ouvriers, car ils développent des forces considérables ; aujourd'hui on tolère davantage de proximité, afin de rendre possibles des gestes, voire du travail collaboratif. De là l'importance de parfaitement contrôler les forces en jeu.

PLS

Pouvez-vous décrire une expérience que vous avez menée pour caractériser les difficultés de la communication humain-robot ?

S.I. : À l'Isir, l'Institut de robotique de l'université Pierre-et-Marie-Curie, nous avons fait interagir iCub – un petit robot en forme d'enfant – avec une cinquantaine d'adultes non versés en robotique. Nous leur avons demandé par exemple de collaborer avec iCub pour désigner le plus lourd de deux objets ou encore l'objet le plus approprié (le plus utile par exemple) dans une certaine situation. Nous avons constaté que les humains adoptent avec le robot des attitudes variables – détachées, pédagogiques, parentales... –, qui traduisent à quel degré ils sont prêts à le considérer comme un véritable agent social.

PLS

Qu'est-ce qui détermine l'attitude de l'humain ?

S.I. : Soit un humain tend à considérer un robot comme « un porteur de capteurs », auquel cas il tend à accepter ses décisions comme on accepte le verdict du thermomètre. Soit quelque chose fait qu'il considère plus ou moins inconsciemment le robot comme un agent social et, dans ce cas, il tend à penser que, la machine ne pouvant avoir son expérience sociale, ses décisions sont invalides. Par exemple, si l'on demande à une personne et à un robot : « Que faut-il mettre dans son sac pour aller à la piscine ? », l'humain, qui tendra à ressentir que son compagnon « n'allant pas à la piscine, ne peut pas savoir », rejettera sa décision. Son attitude change quand il a l'impression d'enseigner au robot quel comportement adéquat avoir en société ou avec lui... Nous avons même constaté avec stupéfaction que certaines personnes se mettent à essayer de convaincre le robot, comme s'il n'avait pas encore appris...

PLS

Ainsi, la question essentielle semble être celle de la confiance ?

S.I. : Oui, beaucoup se méfient encore des machines qui jouent un rôle social.

PLS

Dès lors, comment leur faire mieux accepter les robots dans la société ?

S.I. : En leur donnant assez de contrôle sur la façon dont les robots se comporteront autour d'eux, tout particulièrement si ces derniers jouent un rôle social et encore plus s'ils sont humanoïdes !

Propos recueillis par François Savatier