

BULLETIN N° 207
ACADÉMIE EUROPÉENNE
INTERDISCIPLINAIRE
DES SCIENCES
INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES



Lundi 12 septembre 2016 à 17h :
à la Maison de l'AX, 5 rue Descartes 75005 Paris

Conférence du Professeur Jean Gabriel GANASCIA
Laboratoire Agents Cognitifs et Apprentissage Symbolique Automatique
(ACASA) UMPC-LIP6
" Intelligence artificielle – des Big Data au cerveau "

Notre Prochaine séance aura lieu le lundi 3 octobre 2016 à 17h
à la Maison de l'AX, 5 rue Descartes 75005 Paris
 Elle aura pour thème:

Conférence de Jérôme SACKUR, Professeur à l'Ecole Polytechnique
Directeur d'Études à l' École des Hautes Études en Sciences Sociales,
Laboratoire de Sciences Cognitives et Psycholinguistique/
Ecole Normale Supérieure Ulm
" Temps, subjectivité et métacognition: nouvelles pistes de recherche empirique sur la conscience "

ACADÉMIE EUROPÉENNE INTERDISCIPLINAIRE DES SCIENCES INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES

PRÉSIDENT : Pr Victor MASTRANGELO
VICE PRÉSIDENT : Pr Jean-Pierre FRANÇOISE
VICE PRÉSIDENT BELGIQUE(Liège):
 Pr Jean SCHMETS
VICE PRÉSIDENT ITALIE(Rome):
 Pr Ernesto DI MAURO
SECRÉTAIRE GÉNÉRALE : Irène HERPE-LITWIN
TRÉSORIÈRE GÉNÉRALE: Édith PERRIER

MEMBRES CONSULTATIFS DU CA :
 Gilbert BELAUBRE
 François BÉGON
 Bruno BLONDEL
 Michel GONDRAN

COMMISSION FINANCES: Claude ELBAZ
COMMISSION MULTIMÉDIA: Pr. Alain CORDIER
COMMISSION ÉDITION:
 Pr Robert FRANCK et Pr Pierre NABET
COMMISSION SYNTHÈSES SCIENTIFIQUES:
 Jean-Pierre TREUIL
COMMISSION CANDIDATURES:
 Pr. Jean-Pierre FRANÇOISE

PRÉSIDENT FONDATEUR : Dr. Lucien LÉVY (†)
PRÉSIDENT D'HONNEUR : Gilbert BELAUBRE

CONSEILLERS SCIENTIFIQUES :
SCIENCES DE LA MATIÈRE : Pr. Gilles COHEN-TANNOUJJI
SCIENCES DE LA VIE ET BIOTECHNIQUES : Pr Ernesto DI MAURO

CONSEILLERS SPÉCIAUX:
ÉDITION: Pr Robert FRANCK
AFFAIRES EUROPÉENNES :Pr Jean SCHMETS
RELATIONS VILLE DE PARIS et IDF:
 Michel GONDRAN ex-Président
MOYENS MULTIMÉDIA et RELATIONS UNIVERSITÉS:
 Pr Alain CORDIER
RELATIONS AX et MÉCÉNAT : Gilbert BELAUBRE

SECTION DE NANCY :
PRÉSIDENT : Pr Pierre NABET

septembre 2016

N°207

TABLE DES MATIERES

p. 03 Séance du 12 septembre 2016
 p. 05 Annonces
 p. 06 Documents

Prochaine séance : lundi 3 octobre 2016

**Conférence de Jérôme SACKUR, Professeur à l'Ecole Polytechnique
 Directeur d'Études à l' École des Hautes Études en Sciences Sociales,
 Laboratoire de Sciences Cognitives et Psycholinguistique/
 Ecole Normale Supérieure Ulm**

***" Temps, subjectivité et métacognition: nouvelles pistes de recherche empirique sur
 la conscience "***

ACADEMIE EUROPEENNE INTERDISCIPLINAIRE DES SCIENCES
5, rue Descartes 75005 Paris

Séance du Lundi 12 septembre 2016 5 rue Descartes 75005 Paris à 17h

La séance est ouverte à 17h **sous la Présidence de Victor MASTRANGELO** et en la présence de nos collègues Gilbert BELAUBRE, Jean-Louis BOBIN, Alain CARDON, Gilles COHEN-TANNOUDJI, Françoise DUTHEIL, Claude ELBAZ, Michel GONDRAN, Irène HERPE-LITWIN, Gérard LEVY, Claude MAURY, Edith PERRIER, Jacques PRINTZ, Jean SCHMETS, Alain STAHL, Jean-Pierre TREUIL.

Etaient excusés François BEGON, Jean-Pierre BESSIS, Bruno BLONDEL, Michel CABANAC, Alain CORDIER, Juan-Carlos CHACHQUES, Daniel COURGEAU, Ernesto DI MAURO, Vincent FLEURY, Jean -Pierre FRANÇOISE, Robert FRANCK, Jacques HENRI-ROBERT, Dominique LAMBERT, Jacques LEVY, Valérie LEFEVRE-SEGUIN, Antoine LONG, Pierre MARCHAIS, Anastassios METAXAS, Jacques NIO, Pierre PESQUIES, Michel SPIRO, Jean VERDETTI.

Etait présente en tant que membre correspondant: Marie Françoise PASSINI

I. Présentation de notre conférencier Jean Gabriel GANASCIA

Le Pr Jean Gabriel GANASCIA nous a fourni les données suivantes:

- *Professeur université Pierre et Marie Curie*
- *Directeur Adjoint du Labex OBVIL (Observatoire de la vie littéraire)*
- *Directeur équipe ACASA - Laboratoire d'Informatique de Paris 6 – LIP6 (UMR 7606 CNRS – UPMC)*
- *Membre senior Institut Universitaire de France*
- *Président du COMETS (Comité d'éthique du CNRS)*
- *Membre de la CERNA (Commission de réflexion sur l'Éthique de la Recherche dans les sciences du Numérique d'Allistène)*
- *EurAI Fellow*

Professeur d'informatique à l'Université Pierre et Marie Curie et membre senior de l'Institut Universitaire de France, Jean-Gabriel Ganascia poursuit ses recherches au LIP6 (Laboratoire d'Informatique de Paris VI) et au sein du Labex OBVIL qui fait collaborer son équipe avec les équipes de littérature de l'université Paris-Sorbonne.

Spécialiste d'intelligence artificielle (EurAI Fellow – *European Association for Artificial Intelligence*), d'apprentissage machine et de fouille de données, ses recherches actuelles portent sur le versant littéraire des humanités numériques, sur la philosophie computationnelle et sur l'éthique des technologies de l'information et de la communication.

PARCOURS

Après une formation initiale d'ingénieur et de philosophe, il s'est orienté vers l'informatique et l'intelligence artificielle. Il est titulaire d'une thèse de doctorat sur les systèmes à base de connaissance obtenue à l'université Paris-Sud en 1983 et d'une thèse d'Etat sur l'apprentissage symbolique soutenue à l'université

Paris Sud en 1987. Il a été nommé assistant à l'université d'Orsay (Paris XI) en 1982, puis maître de conférence dans cette même université en 1987 et professeur d'informatique à l'UPMC en 1988.

Il a dirigé le Diplôme d'Etudes Approfondies IARFA (Intelligence Artificielle, Reconnaissance des Formes et Applications) pendant 12 ans (1992-2004). Il a aussi été chargé de mission à la direction du CNRS (1988-1992) avant de créer et de diriger le Programme de Recherches Coordonnées « Sciences Cognitives » pour le compte du ministère de la recherche (1993) puis le Groupement d'Intérêt Scientifique « Sciences de la cognition » (ministère de la recherche, CNRS, CEA, INRIA, INRETS) (1995-2000). Il a aussi coordonné, pour l'université Pierre et Marie Curie, le master Erasmus Mundus DMKM (*Data Mining and Knowledge Management* – Fouille de données et gestion de connaissances).

Il est président du COMETS (comité d'éthique du CNRS) et membre de la CERNA (commission de réflexion sur l'éthique de la recherche dans les sciences du numérique d'Allistène)

PUBLICATIONS

Au cours de sa carrière, il a publié plus de 350 articles dans les actes de conférences scientifiques, dans des livres scientifiques et dans des revues. Il est aussi l'auteur de plusieurs ouvrages destinés au grand public donc voici quelques items :

- Chapouthier G., Ganascia J.-G., Naccache L., Picq P., *Que reste-t-il du propre de l'homme?*, Les Presses de l'ENSTA, 2012
- Ganascia J.-G., *Voir et pouvoir: qui nous surveille?* Editions le pommier, 2009.
- Ganascia J.-G., *Idées reçues sur l'intelligence artificielle*, Le cavalier bleu, 2007
- Ganascia J.-G., *Les sciences cognitives*, Editions le pommier, 2006
- Ganascia J.-G., *Gédéon ou les expériences extravagantes d'un expérimentateur en chambre*, Editions le pommier, collection Roman & Plus, 2002
- Ganascia J.-G., *2001, l'odyssée de l'esprit* – Flammarion (Collection essais) 1999.
- Ganascia J.-G. - *Le petit Trésor de l'informatique et des sciences de l'information* - Flammarion 1998
- Serres M., Farouky N., *Trésor des sciences* - Flammarion 1997.
- Ganascia J.-G.- *L'intelligence artificielle*.- Flammarion (Collection Dominos), 1993
- Ganascia J.-G.- *L'Ame Machine*.- Editions du Seuil, Collection Sciences Ouvertes, 1990

II. Conférence de Jean Gabriel GANASCIA " Intelligence artificielle – des Big Data au cerveau "

Résumé en français de la présentation de notre conférencier:

Le programme AlphaGo, la voiture autonome Google, le logiciel de reconnaissance de la parole Siri, le profilage des utilisateurs, la reconnaissance faciale, ... un nombre incalculable d'applications phares de l'intelligence artificielle recourent désormais à l'apprentissage profond, c'est-à-dire à des réseaux de neurones formels, développés sur de grandes masses de données (*Big Data*). En quoi ces neurones formels peuvent-ils être mis en rapport avec les neurones de notre cerveau ? Telle est la question que nous essayerons d'éclairer en faisant un tour d'horizon de l'intelligence artificielle (IA), des ouvertures de perspective, des transformations induites et des malentendus sous-jacents

Un compte-rendu détaillé sera prochainement disponible sur le site de l'AEIS , <http://www.science-inter.com>.

Notre Président Victor MASTRANGELO procède ensuite à la clôture de cette riche séance.

Irène HERPE-LITWIN

Annonces

- I. L'AEIS vous rappelle la disponibilité en téléchargement gratuit au format PDF de son ouvrage sur le thème du colloque AEIS-2014 "SYSTEMES STELLAIRES ET PLANÉTAIRES- CONDITIONS D'APPARITION DE LA VIE" sur le site d'EDP-Sciences:
<http://www.edp-open.org/images/stories/books/fulldl/Formation-des-systemes-stellaires-et-planetaires.pdf>
- II. Le 4 octobre prochain à 14h se tiendra à l'Académie des Sciences , 23 quai de Conti, 75006 PARIS, une conférence sur l'Intelligence Artificielle. Pour plus d'informations , se rendre sur le site:
http://www.academie-sciences.fr/pdf/350/350_041016.pdf

Documents

Pour illustrer la riche conférence du Pr Jean Gabriel GANASCIA sur l'Intelligence Artificielle (IA) nous vous proposons les articles suivants:

1. Concernant les rapports entre TURING et l'IA:

- p.07: Issu du site du CNRS <http://www.cnrs.fr/ins2i/IMG/pdf/prade.pdf> un article intitulé "Turing et l'Intelligence artificielle" par Pierre Marquis, O.Papini, H. Prade
- p .09: Issu du journal du CNRS "hors série" du mois de mai 2012, "Sur la piste de l'intelligence artificielle" de Laurianne Geoffroy

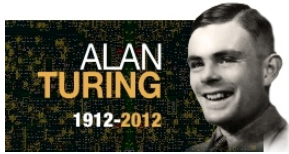
2. Divers articles sur le sujet:

- p. 15: Issu d'un ouvrage intitulé *De l'animal machine à l'âme des machines* sous la direction de Jean-Luc Guichet, publications de la Sorbonne, pp. 169-183, 2010 un chapitre de Jean-Gabriel GANASCIA intitulé "La querelle de l'âme des machines" -

(site à visiter pour lire d'autres articles: <http://www-poleia.lip6.fr/~ganascia/Publications>)

3. Un article général sur la conscience:

- p.28 : Un Editorial de René TREGOUET paru dans la revue ArtFlash du 16 septembre 2016 intitulé : "*Comment le cerveau prend-il conscience du monde* "
site: <http://www.rtflash.fr/comment-cerveau-prend-t-il-conscience-monde/article>



Turing et l'Intelligence Artificielle

P. MARQUIS, O. PAPINI, H. PRADE
(CRIL, Lens; LSIS, Marseille; IRIT Toulouse)

À l'orée des années 1950, entre la naissance de la cybernétique [4] et l'émergence quelques années plus tard de l'intelligence artificielle (IA), alors que les meilleurs esprits du temps s'interrogent sur la possibilité de construire des machines pensantes, Alan Turing propose, dès le début d'un article demeuré célèbre [3], un test pour déterminer si une machine fait montre d'intelligence. On ne peut qu'admirer l'ingéniosité du dispositif qui, en se dégageant au maximum des préjugés et en évitant l'écueil d'avoir à définir dans l'absolu ce qu'est la pensée ou l'intelligence, propose pour ce faire "le jeu de l'imitation" : un interrogateur humain dialoguant avec deux entités (sans les voir) doit déterminer laquelle est l'humain et laquelle est la machine. S'il se trompe plus souvent que quand il a à distinguer, dans les mêmes circonstances, une femme d'un homme, alors la machine franchit le test.

Ce test n'est pas juste une incitation à développer des programmes informatiques capables de tromper l'interrogateur, même si, très tôt, alors que l'IA en était encore à ses balbutiements, un programme des plus simples – ELIZA [5] – a pu, dans un contexte il est vrai très particulier, leurrer un temps des observateurs humains. Le test de Turing vaut d'abord par le but empirique qu'il assigne à l'IA – faire que la machine rivalise par ses performances avec l'humain dans différents registres réputés requérir de l'intelligence. Dans la suite de [3], Turing répond essentiellement aux différentes objections qui se posent naturellement quand on évoque la possibilité de construire un jour une machine intelligente, avant de mettre en avant l'idée d'apprentissage comme une capacité importante pour les machines à qui on veut conférer quelque intelligence, et même de suggérer le rôle du hasard dans des méta-heuristiques pour rechercher des solutions à un problème. Cet accent mis sur l'apprentissage (en particulier par renforcement) se trouve d'ailleurs déjà dans un rapport antérieur [2], davantage empreint d'une vision cybernétique.

Les considérations de Turing [3] concernent aussi les capacités de calcul (et de mémoire) nécessaires au déploiement d'une intelligence. Elles le conduisent à discuter les limitations théoriques des machines, en relation avec des questions de calculabilité et de logique formelle, elles-mêmes liées aux fondements des mathématiques. En revanche, l'article [3] ne mentionne aucunement un certain nombre de thèmes qui constituent maintenant des sujets d'étude importants de l'IA, en matière de logique pour la représentation des connaissances. En particulier, le raisonnement en présence d'information incomplète, incertaine, révisable, évolutive, incohérente, n'est nullement une préoccupation en 1950. Pour autant, on doit nécessairement s'y confronter et renoncer à traiter des informations parfaites quand on s'écarte de la modélisation d'artefacts mathématiques. Heureusement, on sait maintenant que l'information imparfaite peut être traitée rationnellement et rigoureusement.

Les ordinateurs d'aujourd'hui ont des capacités de calcul qui dépassent de beaucoup les espérances qu'on pouvait avoir à ce sujet en 1950. Ces progrès, conjugués à ceux de l'IA, font que la machine aujourd'hui est capable d'explorer des combinatoires importantes, de raisonner déductivement avec des millions de clauses, d'exploiter des représentations sophistiquées gérant l'incertitude, les exceptions, les incohérences. La machine peut être dotée de capacités, toujours en progrès, d'induction, d'apprentissage, de fouille de données, mais aussi de décision, d'argumentation, de calcul de compromis, d'éléments de négociation, de reconnaissance et gestion des émotions, etc.

Les machines actuelles offrent des capacités remarquables, leur permettant par exemple de battre des champions d'échec, d'appréhender des environnements mal connus de façon largement autonome, de réagir à des sollicitations humaines multiples et variées, de composer des textes (y compris en obéissant à de multiples contraintes littéraires). Sont-elles intelligentes pour autant ? Si une seule machine recélait toutes ces capacités et d'autres encore, et était capable de les déployer opportunément selon les circonstances, serait-elle véritablement intelligente ? Il semble bien que non. De nos

jours, une machine peut certes réviser, faire évoluer des objectifs qu'on lui a attribués. Une machine peut même être programmée pour pouvoir réécrire, restructurer sa connaissance initiale à partir d'informations reçues/perçues. Mais la machine d'aujourd'hui ne pense pas : elle n'a pas conscience d'elle-même (et en particulier de ses limites), elle ne peut pas ultimement décider de ses buts ni sans doute imaginer de nouvelles formes de représentations du monde [1].

Au fur et à mesure que se développent des machines capables d'effectuer des tâches réputées intelligentes de plus en plus sophistiquées, progressant ainsi toujours plus loin dans la direction proposée par Turing, on peut s'interroger sur les limites de la simulation de la pensée par la machine. Quel comportement peut-on exactement reproduire par calcul ? Si la thèse de Church/Turing fournit une réponse précise à la question d'un point de vue théorique, il n'est pas exclu que l'on puisse un jour concevoir et développer des machines répondant à d'autres modèles de calcul. On peut aussi s'interroger sur les limites que les lois de la Physique imposeront à de telles machines quant à ce qu'elles pourront calculer en pratique. Symétriquement, se pose la question de la nature et de la spécificité de la pensée humaine, qui sans doute ne relèvent pas seulement de capacités de raisonnement et de décision, aussi remarquables soient-elles. Y a-t-il des limites à la pensée humaine ? En laissant de côté, à dessein, tout point de vue métaphysique sur la question, la nature biologique du cerveau humain (assujetti lui aussi aux lois de la Physique) suggère une réponse positive ... mais le progrès scientifique continu montre aussi que si limite il y a nous sommes encore très loin de l'entrevoir. Par ailleurs, qui sait les surprises que l'évolution réservera ? Ainsi les interrogations de Turing sur la possibilité de construire des machines pensantes resteront sans doute d'actualité pour très longtemps et continueront à stimuler la communauté des chercheurs en IA, comme c'est le cas depuis plus de cinquante ans.

Références

- [1] M. Borillo, H. Prade, J.-L. Soubie. Les machines pensent-elles ? Propos autour du spectacle de Jean-François Peyret "Histoire naturelle de l'esprit (suite et fin)". Livret du Service Culture UPS n.1 (<http://www.irit.fr/Livret-IA/>), 2000.
- [2] A. M. Turing. Intelligent machinery. Report National Physical Laboratory, London, 1948. Reprinted in: *Machine Intelligence*, 5:3-23, Edinburgh University Press, 1969.
- [3] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433-460, 1950.
- [4] N. Wiener. *Cybernetics - or Control and Communication in the Animal and the Machine*. Wiley, 1948.
- [5] J. Weizenbaum. ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36-45, 1966.



cnrs

le journal

Hors-série
Mai 2012



L'héritage d'**ALAN TURING**

- L'inventeur de l'ordinateur
- Le pionnier de l'intelligence artificielle

cnrs

dépasser les frontières

Informatique En créant son concept de machine universelle, puis en mettant au point son fameux test visant à déterminer si une machine peut penser, Alan Turing a ouvert la voie aux travaux sur l'intelligence artificielle de manière décisive.

Sur la piste de l'intelligence artificielle

PAR LAURIANNE GEFFROY

« **Les machines peuvent-elles penser ?** » s'interroge Alan Turing en préambule de l'une de ses publications les plus célèbres¹. La question est abstraite et sans conteste philosophique. Elle ouvre néanmoins très concrètement la voie de l'intelligence artificielle et influencera des générations de chercheurs dans leurs réalisations : moteurs de recherche, réseaux sociaux, analyses d'images, agents virtuels... Mais comment la question d'une machine pensante a-t-elle pu avoir une telle influence ? Tout simplement parce qu'en y répondant Turing propose une définition de l'intelligence et donne, en quelque sorte, le mode d'emploi pour construire cette fameuse machine pensante.

ALAN TURING, L'AUDACIEUX

« *Il fallait vraiment être audacieux et visionnaire pour se poser une question pareille à une époque où les ordinateurs avaient la puissance de calcul d'une montre* », affirme Jean-Paul Haton, chercheur au Laboratoire lorrain de recherche en informatique et ses applications. Effectivement, nous sommes en 1950, l'informatique est balbutiante, et la discipline de l'intelligence artificielle ne fera son apparition officielle qu'en 1956, après la mort de Turing, à la suite de l'atelier d'été organisé par John McCarthy au Dartmouth College.

D'audace et de vision, Turing n'en manquera pas, puisqu'il répondra à cette question philosophique en mathématicien et élaborera le test qui porte aujourd'hui son nom. La première étape de son raisonnement consiste à reformuler la question de départ, abstraite, en une proposition plus facilement



16 En 1997, Garry Kasparov, le champion du monde d'échecs, est battu par l'ordinateur Deep Blue, mis au point par la société IBM.

démontrable. Ainsi, « *Une machine peut-elle penser ?* » devient « *Une machine peut-elle imiter l'homme ?* » qui, lui, sans conteste, pense...

LE TEST DE L'IMITATION

Le test de Turing sera donc un jeu d'imitation à l'issue duquel une machine pourra être déclarée intelligente ou pas. Dans un premier temps, un interrogateur doit dialoguer avec une femme et un homme ayant pour consigne de se faire passer pour une femme, puis déterminer lequel des deux est réellement une femme. Pour faire abstraction des apparences, de la voix et du visage, les échanges sont dactylographiés. Dans un second temps, l'homme est remplacé par une machine à l'insu de l'interrogateur. Désormais, la machine imite l'homme qui imite la femme. Si l'interrogateur est incapable d'identifier la femme, qu'il ne peut dissocier l'homme et la machine, alors on peut considérer que la machine est intelligente.

Deux présupposés, qui fondent la notion d'intelligence pour Turing, sont requis pour élaborer ce test. En premier lieu, Alan Turing a la certitude que toute activité intellectuelle peut se reproduire

à l'aide de nombres. « *Le calcul, qui vient du mot caillou, consiste à manipuler des objets symboliques pour décrire des opérations de la pensée*, raconte Jean-Gabriel Ganascia, chercheur au Laboratoire d'informatique de Paris 6 (LIP6)². Alan Turing s'est demandé quelle était la puissance de ces objets symboliques, autrement dit, ce que l'on pouvait faire avec des calculs. » Dès 1945, il a l'intime conviction que les fonctions de notre esprit peuvent s'écrire sous forme mathématique à partir d'un nombre fini de symboles que l'on stockerait et que l'on emploierait au fur et à mesure que le raisonnement avance. C'est-à-dire qu'une machine dotée d'une mémoire, d'une bonne base de connaissances et de règles de calcul bien choisies pourrait tout à fait mimer le raisonnement humain.

Ce faisant, il dresse le portrait intuitif de ce qui deviendra l'ordinateur moderne, avec son disque dur, son processeur, son système d'exploitation et ses

L'ERREUR EST HUMAIN !

Pour se faire passer pour un humain, une machine doit avoir sa logique, mais aussi... ses défauts. Alan Turing étayera sa démonstration avec des dialogues imaginaires, précisant les erreurs et les errements humains à reproduire pour donner toutes les chances à la machine de réussir le test. Ainsi, lorsque l'interrogateur demande « *Pourriez-vous m'écrire un sonnet ?* », la machine n'aura qu'à botter en touche, comme le feraient la plupart des gens, et répondre « *Je n'ai aucun talent pour la poésie* ». De la même façon, si l'interrogateur demande « *Ajoutez 34 957 à 70 764* », la machine, après une pause de 30 secondes, pourra fournir une réponse fautive « *105 621* ».

17



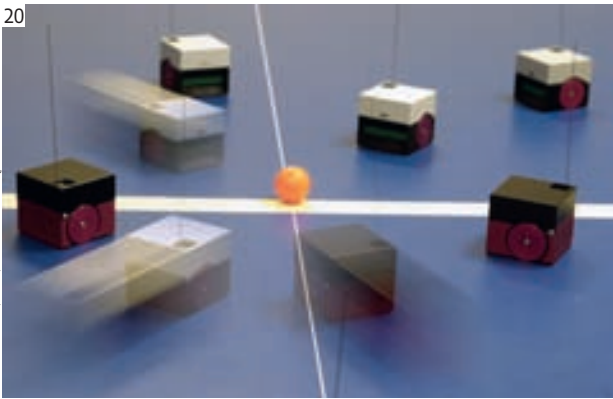
© B. KRAFFT

18



© B. KRAFFT

20



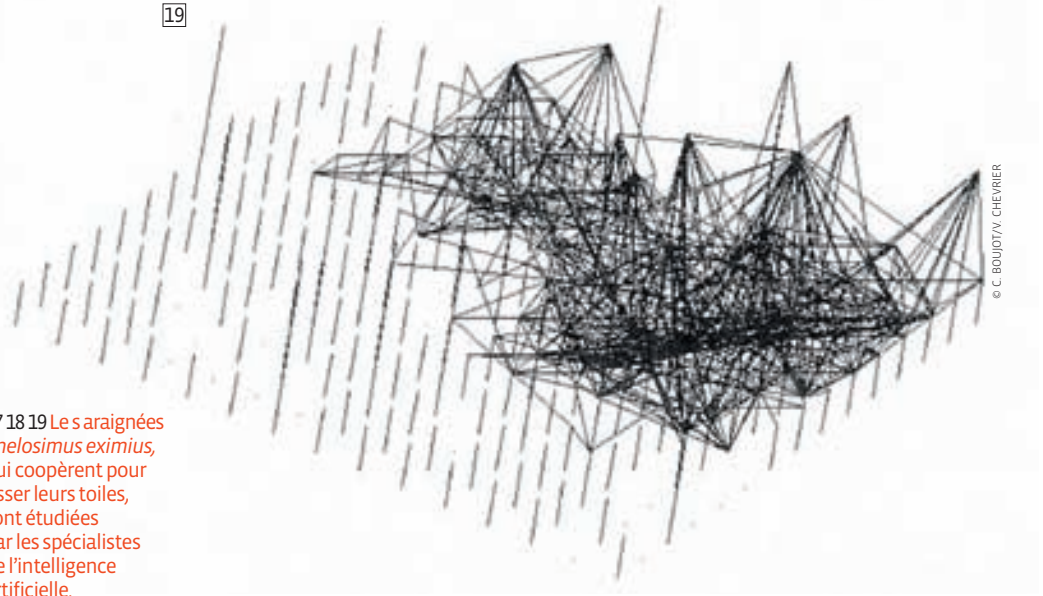
© R. LAMOURÉUX/INRIA/CNRS PHOTOTHÈQUE

programmes. « Alan Turing est le père de l'intelligence artificielle, mais aussi, et de façon indissociable, le père de l'informatique », souligne Marie-Christine Rousset, déléguée scientifique au Laboratoire d'informatique de Grenoble³. Une paternité qui lui est à ce point reconnue que certains n'hésitent pas à voir dans le célèbre logo de la "pomme croquée", apposé sur quelques millions d'ordinateurs, un hommage posthume à ce mathématicien visionnaire.

LES LIMITES DES ALGORITHMES

En décrivant sa machine, Turing envisage également ses possibilités. « Il a ouvert le domaine de l'intelligence artificielle tout en montrant qu'il existait des limites, que certaines choses pouvaient être mécanisées et automatisables dans le raisonnement humain, et d'autres pas, poursuit Marie-Christine Rousset. Ces limites, propres aux algorithmes, offrent un cadre fondamental qu'il est indispensable de connaître pour ne pas promettre la lune, ce que beaucoup de chercheurs en intelligence

19



© C. BOUJOTY CHEVRIER

17 18 19 Les araignées *Anelosimus eximius*, qui coopèrent pour tisser leurs toiles, sont étudiées par les spécialistes de l'intelligence artificielle.

artificielle ont fait au départ. » À ce premier présupposé concernant la reproduction de l'intelligence par les nombres s'ajoute un second : l'intelligence est indépendante de la forme. La différence physique entre un homme et une femme n'a donc aucune incidence sur l'intelligence proprement dite, pas plus que la différence de constitution entre un homme et une machine.

« Avec son test, Alan Turing circonscrit vraiment la portée de l'intelligence artificielle en éliminant les aspects physiques et physiologiques », résume Jean-Gabriel Ganascia. Débarassée de la forme, l'intelligence n'est donc plus le propre de l'homme. On peut ainsi la déceler dans la nature, l'évolution des espèces ou certains comportements animaux, et l'utiliser, au même titre que le raisonnement humain, pour créer des

algorithmes destinés à nourrir les machines. « Aujourd'hui, les robots en essaim s'inspirent largement des insectes sociaux, des fourmis ou des abeilles, qui font émerger de l'intelligence à partir d'interactions très simples, informe Jean-Paul Haton. Nous avons, par exemple, modélisé le comportement d'araignées sociales, de type *Anelosimus eximius*, qui coopèrent pour construire leurs toiles. Il s'agit de recherches fondamentales qui pourraient très bien, un jour, inspirer le monde industriel pour les chaînes de fabrication. » À la fin de sa vie, Turing lui-même tentera de mettre en équation la croissance des cellules biologiques, la façon dont celles-ci s'organisent et permettent à l'intelligence d'émerger. Ces travaux sur la morphogenèse devanceront les notions d'algorithme génétique et d'informatique évolutionniste.

21



© A. BADER/SCHER/ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Jean-Gabriel Ganascia, Laboratoire d'informatique de Paris 6

« Le test de Turing circonscrit la portée de l'intelligence des machines. Il évacue le débat sur la conscience ou plutôt sur la reproduction d'une conscience ; il évacue les dimensions perceptives et corporelles de l'intelligence, ainsi que les interrogations métaphysiques. Ce faisant, il a tracé la voie à l'approche pragmatique et empirique qu'a adoptée l'intelligence artificielle et qui a assuré ses indéniables succès pendant plus d'un demi-siècle. »



© C. LEBEDINSKY/CNRS PHOTOTHÈQUE

Pour Alan Turing, les connaissances sont indispensables à la réalisation d'une machine intelligente. Aucune machine ne peut jouer au jeu de l'imitation et réussir à duper l'homme si on ne lui donne pas un solide bagage culturel et des règles logiques pour l'utiliser à bon escient. Cette approche anticipe ainsi ce que seront les systèmes à base de connaissance, l'une des deux voies de l'intelligence artificielle.

LE CERVEAU DÉCORTIQUÉ

Ces connaissances, Turing envisage également que la machine puisse les acquérir toute seule et se met donc à étudier l'organisation et la dynamique de notre cerveau. Il décrit, dans une publication de 1948⁴, des réseaux de neurones artificiels – modélisés sous forme mathématique et destinés, en théorie, à devenir des réseaux électroniques – initialement connectés de façon aléatoire et capables de s'auto-organiser en fonction de l'usage que l'on en fait. C'est la deuxième voie de l'intelligence artificielle.

« Ces deux écoles, intitulées Making a mind (imiter la pensée) et Modeling the brain (imiter le cerveau) par le philosophe américain Hubert Dreyfus, se sont longtemps opposées, mais connaissent, depuis le début du XXI^e siècle, un rapprochement certain, signale Jean-Paul Haton. Une troisième voie, fondée sur des méthodes statistiques et probabilistes, a également fait son apparition pour prendre en compte la variabilité naturelle des données. » Cette approche, que l'on couple volontiers aux deux autres, permet, par exemple, de faire de la reconnaissance vocale de qualité en prenant en compte

la variabilité de la voie humaine. *Idem* pour la reconnaissance de l'écriture manuscrite ou le traitement automatique de la langue naturelle.

Quelle que soit la voie utilisée, l'intelligence artificielle a envahi notre quotidien. Elle joue notamment un rôle indispensable dans une société nourrie par un flux constant de vidéos, d'images et de messages en tout genre. Tout un pan de la recherche développe ainsi de nouveaux outils pour compiler, trier et analyser les ressources du Web, des technologies regroupées sous le terme de Web sémantique. « On souhaite développer des moteurs de requête afin de pouvoir interroger le Web par l'intermédiaire d'une question précise, telle "Où est mort Albert Einstein ?", et de pouvoir obtenir en retour une réponse claire et non une liste de documents, illustre Marie-Christine Rousset, qui dirige, pour l'Institut universitaire de France, le projet Intelligence artificielle et Web : extraire et raisonner sur les données du Web. L'une des difficultés est la mise au point d'outils qui restent performants face à la masse phénoménale

d'informations disponibles sur le Web. Le cadre formel de la calculabilité et de la complexité fondé sur le modèle de la machine de Turing est fondamental pour relever ce défi (lire p. 9). L'autre challenge est d'évaluer la qualité des données car, sur le Web, il y a toujours des erreurs ou des opinions divergentes. »

L'ANALYSE DE CONTENUS

Au LIP6, certains chercheurs se concentrent sur l'image et développent des outils pour analyser son contenu : savoir ce qu'elle représente, s'il s'agit d'un objet ou d'une personne, et même de quelle personne exactement. Sur le Web, il est d'ores et déjà possible de faire de la recherche d'images simples par similarité (forme, couleur, texture, etc.) en soumettant au moteur de recherche une requête sous la forme d'une image exemple. Mais la recherche intelligente d'images est loin d'être parfaite. « Il y a toujours un fossé sémantique entre ce que la machine peut faire et les attentes des utilisateurs, réagit Matthieu Cord, spécialiste du traitement de l'image au



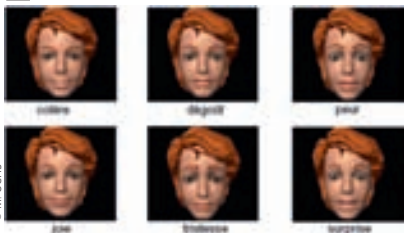
À voir sur le journal en ligne : le film *Le Qi d'iCub*.

22 Le robot humanoïde européen iCub a la taille d'un enfant de 3 ans. Ses algorithmes lui permettent d'apprendre à mesurer qu'il interagit avec son environnement.



© L. NATALE

23



23 Certains agents virtuels peuvent exprimer des émotions. 24 Mis au point dans le cadre du projet européen Semaine, l'auditeur sensible artificiel (SAL) peut dialoguer avec un humain via un personnage virtuel.

24



LIP6. En ce moment, avec le projet Asap, nous développons des algorithmes pour faire de la reconnaissance automatique du texte enfoui dans les images ou les vidéos, afin de pouvoir générer des résumés automatiques. Pour nous, c'est très simple, mais pour une machine... »

LA RECONNAISSANCE FACIALE

Les réseaux de neurones artificiels, boostés ces dernières années par l'évolution des technologies et des méthodes d'apprentissage, sont, eux, particulièrement efficaces pour la reconnaissance de visages. « Les réseaux de neurones dernière génération peuvent, à l'image de notre œil, analyser les visages, couche d'information après couche d'information, les mémoriser, puis les reconnaître lorsqu'ils les rencontrent à nouveau », raconte Ludovic Denoyer, spécialiste des réseaux de neurones LIP6.

Parmi les inventions contemporaines en matière d'intelligence artificielle, certaines sont les descendantes directes



25

25 Eliza, l'un des tout premiers agents conversationnels, imite un psychothérapeute lors d'un premier entretien avec son patient.

des machines décrites par Turing dans son test, soit des automates passés maîtres dans l'art de la conversation. Qui n'a jamais été interpellé sur Internet, dans un magasin ou un musée par l'un ces petits personnages animés : « Puis-je vous aider ? Posez-moi votre question. » On appelle ces avatars des *chatbots*, des « robots bavards » en anglais, ou encore des « agents conversationnels » en français. « Il y a deux catégories d'agents conversationnels : ceux qui comprennent

réellement ce qu'on leur dit en réalisant un traitement automatique de la langue, et ceux qui ne comprennent rien à ce qu'on leur raconte, comme Eliza, mais qui ont un système de règles suffisamment malin pour répondre, à peu près bien, à n'importe quelle question », explique Nicolas Sabouret, spécialiste des agents conversationnels au LIP6.

LE SUCCÈS DU ROBOT PSYCHIATRE

Eliza, développé par Joseph Weizenbaum en 1966 pour jouer les apprentis psychiatres, fut l'un des tout premiers agents conversationnels. Il possédait alors moins de 50 mots clés et reformulait la plupart des phrases de son « patient » : « P : Je suis angoissé. – E : Pourquoi dites-vous que vous êtes angoissé ? – P : C'est à cause de ma femme. – E : Parlez-moi de votre famille. » Eliza fut tellement convaincant que certains chercheurs l'imaginaient déjà passer le test de Turing avec succès... Mais, aujourd'hui encore, aucun programme n'a réussi cette prouesse (*lire l'encadré ci-contre*). Les agents de conversation les plus perfectionnés fonctionnent désormais avec 40 000 règles et ont recours à des modèles statistiques, ce qui améliore sensiblement la pertinence de leurs réponses, étend le champ de leur conversation et leur offre même le sens de l'humour (*lire l'encadré p. 13*).

Depuis Turing, les agents conversationnels et les robots compagnons se sont dotés de la parole et de la vision, ce qui a amélioré considérablement leurs

LES "ROBOTS BAVARDS" ONT LEUR PRIX

Tous les ans depuis 1991, le prix Loebner récompense les *chatbots* les plus convaincants, ceux qui imitent l'homme à la perfection (ou plutôt à l'imperfection). Comme dans le scénario imaginé par Turing, des juges dialoguent à l'aveugle avec des hommes et des *chatbots* et doivent deviner qui est qui. Turing avait prédit qu'en l'an 2000 les ordinateurs parviendraient à tromper 30% des juges après cinq minutes de conversation. Cette prédiction ne s'est pas encore réalisée, mais les meilleurs *chatbots* arrivent à duper plusieurs juges, tandis que les humains de la compétition se voient décerner l'étrange titre d'humain le plus humain !



interactions avec l'homme. Il est d'ailleurs étonnant de constater qu'à travers nos seules expressions faciales un robot arrive à cerner notre état d'esprit et à nous répondre de façon tout à fait pertinente. Ce type de programme est de plus en plus utilisé lors de formations, pour apprendre, par exemple, à gérer un client difficile ou un manifestant agité.

« Dans le cadre du projet européen Tardis, nous allons développer un agent virtuel pour que des jeunes en insertion sociale s'entraînent à passer des entretiens d'embauche, indique Nicolas Sabouret. Le recruteur virtuel n'utilisera que les émotions exprimées par le jeune,

de la levure de boulanger et formulait des hypothèses convaincantes sur la fonction de certains gènes.

Plus récemment, en 2011, le superordinateur Watson, d'IBM, capable de répondre en quelques secondes à n'importe quelle question de culture générale, mais aussi de s'abstenir lorsqu'il n'est pas sûr de sa réponse, remportait le jeu Jeopardy. Doté de nouveaux algorithmes et de solides connaissances médicales, Watson va rejoindre l'hôpital Cedars-Sinai, à Los Angeles, pour aider les spécialistes à affiner leur diagnostic ou à proposer un traitement contre le cancer. « On dit toujours que l'intelligence

artificielle n'a pas tenu ses promesses, qu'elle n'a pas réussi à faire un double de l'homme. Mais cela n'a jamais été l'objectif », souligne Jean-Gabriel Ganascia. Aujourd'hui, l'intelligence artificielle doit donc être jugée à l'aune de ses réalisations concrètes, et non pas des rêves fous qu'elle a pu susciter.

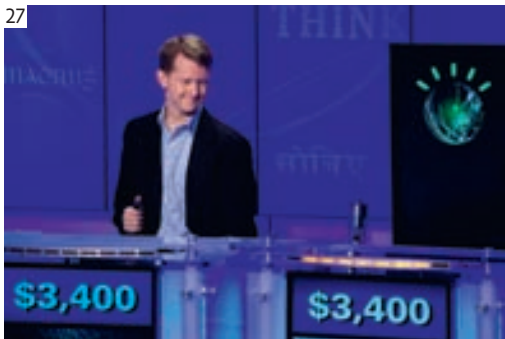
1. « Computing Machinery and Intelligence », *Mind*, 1950.
2. Unité CNRS/UPMC.
3. Unité CNRS/Grenoble INP/Inria/UJF/UPMF/ Université Stendhal-Grenoble-III.
4. « Intelligent Machinery », *National Physical Laboratory*, 1948.

26



© G. EVANS/ABERYSTWYTH UNIVERSITY

27



© GETTY IMAGES/ATP

CONTACTS :

Matthieu Cord
> matthieu.cord@lip6.fr
Ludovic Denoyer
> ludovic.denoyer@lip6.fr
Jean-Gabriel Ganascia
> jean-gabriel.ganascia@lip6.fr
Jean-Paul Haton
> jean-paul.haton@loria.fr
Marie-Christine Rousset
> marie-christine.rousset@imag.fr
Nicolas Sabouret
> nicolas.sabouret@lip6.fr

à travers la voix et la reconnaissance des expressions faciales, pour réagir de façon adaptée. Si un jeune se révèle trop désinvolte, l'agent pourra ainsi se montrer cassant dans la question suivante. » D'autres laboratoires se focalisent sur la voix. « Aujourd'hui, il est courant de pouvoir donner des ordres à son téléphone portable mais, d'ici à cinq ans, grâce à un système de traduction instantanée, on pourra discuter en français avec un ami japonais qui parle japonais, illustre Jean-Paul Haton. Ce n'est plus de la science-fiction ! »

UN DÉBAT TOUJOURS OUVERT

Alors, les machines peuvent-elles penser ? Le débat reste ouvert. Mais force est de constater que les programmes imaginés par les spécialistes de l'intelligence artificielle réalisent des performances de plus en plus époustouflantes. En 2005, le robot scientifique Adam, de l'université d'Aberystwyth, au pays de Galles, réalisait des expériences sur le génome

26 Le robot Adam émet des hypothèses sur la fonction des gènes. 27 En 2011, l'ordinateur Watson a remporté le jeu Jeopardy. 28 Dans son film 2001, l'odyssée de l'espace, Stanley Kubrick a imaginé un superordinateur doté d'une intelligence artificielle.

28



© MGM/RANG-IMAGES

La querelle de l'âme des machines

Jean-Gabriel Ganascia

Avertissement

Qu'il puisse y avoir une dispute sur l'âme des machines apparaîtra vraisemblablement bien insolite, voire incongru et quelque peu effrayant, à l'érudit familier des études classiques. Aussi, convient-il de le rassurer : rien ne laisse penser que les machines actuelles qui peuplent si ardemment notre univers quotidien ressentent vraiment des émotions ou, *a fortiori*, disposent de l'équivalent d'une conscience. En dépit de l'insistance dont il fait preuve, le téléphone portable ne s'impatiente jamais, il ne s'offusque pas lorsqu'on lui coupe le sifflet et l'ordinateur ne manifeste aucune irritation devant nos errements malhabiles. Certes, les industriels et les institutions européennes qui financent la recherche évoquent parfois l'informatique affective ou le calcul des émotions. Mais, à y regarder de près, les ingénieurs ne s'intéressent pas directement à ce que ressentent les machines : s'ils explorent les mécanismes biologiques de l'émotion et ce qui la suscite en nous, ce n'est pas tant pour réifier la conscience sur des machines, que pour fabriquer des « machines animistes », c'est-à-dire des machines animées à qui nous prêtons, pour un temps, et dans des circonstances limitées, une âme.

Ces précautions étant prises, soulignons que les mythes contemporains suggèrent de tels possibles. A titre d'illustration, mentionnons deux événements de ce début de XXI^e siècle. Le premier fût la sortie, en 2001, d'un film de science fiction intitulé *AI, Artificial Intelligence* et dont le scénario a été écrit par Steven Spielberg ; il se présente comme une réplique moderne de Pinocchio : on y voit un monde où se croisent des êtres humains semblables en tous points à nos contemporains et des doublures parfaites d'hommes, de femmes ou d'enfants faites pour se substituer aux absents, à nos proches, à ceux que nous avons perdus, ou aux autres, à ces amantes et amants de passage que nous désirons éperdument, sans espoir de les conquérir jamais. Or, l'un de ces androïdes ne se contente pas de provoquer des émotions chez les hommes, il en possède... Et, c'est là que l'intrigue se noue. Le second épisode de la vie publique tient à une partie d'échecs. Tous se souviennent qu'en 1997 le champion du monde d'échecs en titre, Garry Kasparov, fût battu par l'ordinateur *Deep Blue* après un tournoi épique comprenant six parties : la première a été gagnée par Garry Kasparov, la seconde par la machine, puis il y eut trois parties nulles avant une finale tragique à l'issue de laquelle l'homme déclara forfait. Six ans plus tard, en 2003, une seconde finale eut lieu entre un nouvel ordinateur, le dénommé *X3D-Fritz*, et Garry Kasparov. L'innovation tenait à l'emploi de réalités

virtuelles, qui reconstituaient un échiquier en trois dimensions, et à la mise en scène : on invita la plus belle femme des Etats-Unis, Katie Horn, qui venait d'être élue Miss New-York, à jouer le premier coup dans l'espoir que X3D-Fritz en perdrait ses moyens... Et – oh, déception ! – il n'en fût rien : Kasparov fût de nouveau battu. Dans ces deux cas, comme en bien d'autres, nous constatons que nos contemporains semblent se préparer à accueillir des machines douées de conscience. Mais, ces mythes sont-ils plus prégnants aujourd'hui que ne le furent ceux de Pygmalion ou du Golem en leur temps ? Rien ne le dit, si ce n'est que les exploits éclatants de la technique les rendent apparemment plus vraisemblables.

Enjeux

Indépendamment de ces légendes, la question de l'âme des machines suscite toujours de vives réactions. Certains récusent d'emblée l'hypothèse comme vaine, contre-intuitive ou réductionniste ; d'autres y voient l'avènement d'une technologie toute puissante condamnant l'homme à devenir esclave de ses propres créatures. Et des débats enflammés sur la question de la conscience des machines, toute proche de celle de l'âme des machines, animent le cercle des philosophes d'inspiration analytique. Bref, la controverse demeure vive. Qui plus est, à bien des égards, on trouve un parallélisme entre les arguments déployés aujourd'hui, par les philosophes qui traitent de l'âme ou de la conscience des machines, et ceux des philosophes de l'époque moderne qui disputèrent de l'âme des animaux. Tout cela incite indubitablement à aller plus avant et à se plonger dans les discussions autour de l'âme – ou de la conscience – des machines, pour les mettre en regard des controverses autour de la question de l'âme des bêtes. Le contexte diffère indubitablement : les machines ne sont pas les animaux ; elles font des calculs, elles reconstruisent sans trop de difficultés des raisonnements abstraits, elles s'adaptent aux conditions environnantes et elles manipulent de grandes quantités de mots qu'elles combinent pour produire des énoncés si pertinents qu'ils font parfois illusion ; en revanche, telles l'albatros sur le pont du navire, elles peinent à se mouvoir dans le monde et à discerner les objets et les êtres qui le peuplent. Les animaux, quant à eux, perçoivent mieux que nous et *a fortiori* que les machines ; ils se déplacent avec plus d'agilité ; ils font preuve de ruse et d'aptitude à chaparder, à courir ou à chasser leur proie. Mais, leurs capacités à maîtriser un langage demeurent assez piètres, en dépit des efforts pour le leur apprendre. On connaît les exploits des primates. Rappelons ceux d'Alex, le perroquet mort en 2007, au grand dam de sa tutrice Irene Pepperberg¹ : il comptait jusqu'à six, nommait une cinquantaine d'objets, de formes et de

¹ « R.I.P. Alex, the parrot that learnt to count and communicate », *The Independent*, Andrew Gumbel in Los Angeles, 12 September 2007 (remarque : R.I.P. signifie Requiescat In Pace en latin ou Rest in Peace en Anlais).

couleurs et possédait la maturité affective d'un enfant de deux ans. Pour exceptionnelles qu'elles fussent, ces capacités cognitives demeurent assez dérisoires comparées à celles de nos ordinateurs contemporains. Nous nous trouvons donc aux deux extrémités : l'animal perçoit et ressent, sans disposer de rationalité ; la machine raisonne, sans discernement ni conscience. Selon que l'on place l'âme du côté de la conscience ou de celui de la raison, on accordera plus facilement une âme aux animaux ou aux machines. Ainsi, tant la question de l'âme des machines que celle de l'âme des animaux révèle l'ambivalence du concept d'âme qui suscite aujourd'hui, comme à l'époque moderne, querelles et controverses.

A ces indéniables différences entre les hommes et les machines, il faut ajouter un changement de contexte : lorsqu'il introduisit sa théorie des animaux-machines, Descartes souhaitait résoudre deux questions d'ordre théologique, celle de la présence des animaux dans l'au-delà et celle de la souffrance animale, que seul un Dieu méchant aurait pu tolérer. De telles interrogations ne sont plus à l'ordre du jour, et beaucoup de travaux contemporains poursuivis par les éthologistes montrent que les animaux possèdent des capacités cognitives troublantes : indubitablement, ils perçoivent et ils ressentent des émotions. Symétriquement, rien dans la science actuelle des machines ne laisse supposer que celles-ci ressentent une quelconque émotion. Cependant, la question qui nous intéresse ici ne porte pas sur les faits, mais sur les possibilités logiques : pourrait-il se faire qu'une machine possède un jour une âme ? Prise au sérieux et traitée en termes philosophiques, cette question nous renseigne d'abord sur les débats qui traversent la communauté philosophique d'inspiration analytique, en particulier sur les controverses entre tenants du cognitivisme orthodoxe, fondé sur des manipulations de symboles, et représentants du neuro-mimétisme, qui cherchent à naturaliser la philosophie en l'ancrant sur une science physique et physiologiste de l'esprit. Nous retrouvons sans aucun doute là des oppositions semblables à celles qui traversèrent la communauté philosophique à l'époque moderne, entre cartésiens et mécanistes. De ce fait, l'évocation des disputes actuelles est susceptible d'éclairer d'un jour neuf l'ancienne querelle de l'âme des animaux, du moins tel est l'objectif de cet article.

Pour clarifier les enjeux, commençons par les généralités : bien évidemment, l'existence d'une âme dans la machine viderait de sa substance la querelle de l'âme des bêtes, puisque cette même querelle naquit de la distinction entre les hommes, qui auraient une âme, et les animaux qui, réduits à de simples machines, ne seraient pas dotés d'âmes. C'est d'ailleurs l'argument des mécanistes du XVIII^e qui, comme Julien de La Mettrie², radicalisèrent la

² J. O. de, La Mettrie, *L'homme machine*, in *Œuvres Philosophiques*, texte revu par Francine Markovits, 2 t., Paris, Fayard, 1987 (« Corpus des Œuvres de Philosophie en Langue

position de Descartes au point d'admettre que l'homme lui-même ne serait qu'un automate réductible à des mécanismes élémentaires. Dans cette éventualité, soit on dénie à l'âme toute réalité, soit on admet que l'âme serait le produit de l'activité d'une machine et qu'il existerait donc une âme chez les animaux aussi bien que chez les hommes.

Notre but n'est pas de prendre parti et de convaincre du bien fondé de la position de La Mettrie, mais de montrer l'actualité de cette problématique. Certes, selon toute vraisemblance, la querelle de l'âme des bêtes s'est éteinte ; l'intérêt qu'on y porte semble essentiellement historique. Il met en lumière les débats qui ont traversé la scène philosophique française à l'âge des Lumières. En revanche, la querelle de l'âme des machines apparaît toute contemporaine et c'est ce que nous allons essayer de montrer en évoquant les débats qui eurent lieu dans le champ des sciences cognitives et des différents courants philosophiques cognitivistes qui s'en inspirent. Bien évidemment, comme nous allons le voir, le terme d'âme n'est plus beaucoup employé. On lui préfère aujourd'hui ceux d'agent, de conscience ou d'esprit. Les mots diffèrent ; la querelle demeure et, comme nous le verrons plus loin, les arguments sont analogues.

Ames et agents

Pour commencer, disons-le tout de go : la plupart des spécialistes d'intelligence artificielle sont aristotéliens, sans le savoir ! Plus exactement, ils pensent que les machines disposent d'une âme au sens aristotélien. L'affirmation apparaît assez péremptoire ; pourtant, en lisant attentivement les écrits des pères fondateurs de l'intelligence artificielle, en particulier d'Herbert Simon³, d'Alan Newell⁴ ou de John McCarthy⁵, il apparaît que le fondement de cette discipline repose sur ce postulat. Simplement, le mot *âme* a été remplacé par un terme plus contemporain, issu de la théorie économique : le terme d'agent.

La suite de cette partie explicite le rapprochement entre âme, au sens aristotélien, et agent. Profitons-en pour rappeler que jusqu'à Descartes, personne ne contestait l'existence d'une âme animale. Et, si nous reprenons le mot *âme* dans le sens qu'Aristote lui attribuait, il apparaît plus difficile encore de refuser aux machines une âme, même si, au regard de la

française »).

³ H. Simon, *Les sciences de l'artificiel*, trad. de l'anglais par Jean-Louis Le Moigne, Paris, Gallimard, col. folio/essais, 2004.

⁴ A. Newell, 1982, « The Knowledge Level », *Artificial Intelligence Journal*, 18, p. 87-127.

⁵ J. McCarthy, et P. Hayes, 1969. « Some philosophical problems from the standpoint of artificial intelligence », *Machine Intelligence 4* (textes réunis par Bernard Meltzer et David Michie), New York, American Elsevier, p. 463-502.

signification que le terme a progressivement acquise, l'idée en choque plus d'un.

Etymologie

Commençons par la signification des mots : si l'on se réfère à Aristote, en particulier au *Traité de l'âme*⁶ et à l'*Ethique de Nicomaque*⁷, l'âme caractérise les êtres animés ; l'âme c'est, sinon le souffle qui anime, tout au moins le moteur, la source du mouvement. Plus exactement, selon Aristote, les deux propriétés communes à tous les êtres animés sont le *mouvement* et la *sensation*.

Il en va de même en intelligence artificielle avec la notion d'*agent*. Un agent est, par définition, une entité qui agit, et donc qui est *en mouvement*. Deuxième propriété, un agent reçoit des informations par l'intermédiaire d'un certain nombre de capteurs ; autrement dit, il possède l'équivalent de *sensations* qui lui procurent des *connaissances* sur le monde.

Fonctions de l'âme

Venons en maintenant aux fonctions de l'âme. Aristote en reconnaît quatre : ce sont les fonctions *nutritive*, *sensitive*, *cogitative* et *de mouvement*.

Tous les êtres vivants, même les plantes, possèdent la fonction nutritive qui recouvre, pour Aristote, à la fois la faculté de se reproduire et celle de s'alimenter.

Les animaux détiennent une fonction sensitive ; le toucher et le goût, qui se présente selon Aristote comme le sens de la nourriture et comme une modalité particulière du toucher, appartiennent à tous. Et, avec ces sens, viennent les sentiments de plaisir et de douleur ainsi que l'appétit qui correspond, toujours pour Aristote, à une attirance vers ce qui cause le plaisir et à une répulsion devant ce qui produit la douleur.

Parmi les animaux, c'est-à-dire parmi les êtres vivants doués de sensations, certains possèdent une faculté de mouvement. Et un très petit nombre possèdent le raisonnement et la réflexion, autrement dit, la fonction cogitative.

Avec les machines contemporaines, il semble qu'il en aille de même : toutes ont une fonction « végétative » ou nutritive. Un courant électrique les alimente et elles subsistent au rythme de la pulsation de leur unité centrale.

Parmi ces machines, certaines se couplent avec des capteurs qui recueillent de l'information relative au monde environnant et agissent en fonction de ces données. Comme elles agissent, on parle d'agents. Parmi ceux-ci, quelques

⁶ Aristote, *De l'âme*, Paris, Flammarion, collection GF-Flammarion, 1993.

⁷ Aristote, *Ethique de Nicomaque*, Paris, Flammarion, collection GF-Flammarion, 1992.

uns se meuvent dans le monde extérieur, comme les robots ou les entités virtuelles qui explorent le web.

Ces agents eux-mêmes se divisent en deux catégories.

Agents réactifs

Certains d'entre eux, qualifiés de réactifs, agissent en réponse à leurs sensations ; ils visent à se nourrir ou, plus exactement, à optimiser une fonction de récompense qui résume la quantité de plaisir – ou de déplaisir – qu'ils ont accumulée. Un bon exemple nous est fourni par les robots aspirateurs « Roomba » que l'on trouve désormais dans le commerce ; ils possèdent des testeurs mécaniques et infrarouges qui leur évitent de se cogner contre les murs ou de tomber dans les escaliers. D'autres capteurs mesurent la quantité de poussière qu'ils ont aspirée. Ces robots se nourrissent d'électricité et se délectent de poussière. Ils apprennent, au fur et à mesure de leurs pérégrinations, à explorer les lieux où ils ont plus de chances de trouver de la poussière. Leur fonction de récompense se fonde sur la quantité de poussière ingérée. Ils aspirent la poussière et ils aspirent à la poussière, sauf lorsqu'ils n'ont plus suffisamment d'électricité dans leur batterie, auquel cas, de peur de compromettre, pour une satisfaction immédiate, leurs aspirations futures à la poussière, ils rentrent d'eux-mêmes dans leur station d'alimentation afin de se recharger. Et, grâce à cela ils se régénèrent et poursuivent leur course à la poussière.

Agents rationnels

D'autres agents, dits intelligents ou rationnels, possèdent l'équivalent d'une volonté, ou ce que l'on appelle en termes techniques, un but. Ils procèdent alors à un calcul pour déterminer la séquence d'actions qui leur permettra d'atteindre le but ou, tout au moins, de se rapprocher le plus possible de sa réalisation.

Notons qu'il existe aussi des agents doués de la fonction cogitative qui ne raisonnent pas à proprement parler, car ils ne calculent pas la séquence d'actions qui leur permet d'atteindre un but prédéfini. Ce sont des *agents réflexifs* visant à poursuivre leur spéculation. Ainsi en va-t-il d'un agent que nous avons conçu dans mon équipe afin de composer les improvisations d'une ligne de basse dans un trio rythmique, en jazz⁸. Sa finalité tient simplement à jouer de la musique, autrement dit à engendrer une séquence de notes qui obéit aux contraintes de la grille d'accords et du jeu de ses compagnons.

⁸ G. Ramalho, P.-Y. Rolland, J.-G. Ganascia, « An Artificially Intelligent Jazz Performer », *Journal of New Music Research*, Amsterdam, v. 28, n. 2, p. 105-129, 1999.

Autonomie des agents

Une question subsiste toutefois : les agents fabriqués à l'aide des techniques d'intelligence artificielle parviennent-ils à acquérir une forme d'autonomie ? Autrement dit, en reformulant la question en termes anciens, ont-ils un « libre arbitre » ?

Nous savons à quel point cette question fit l'objet de polémiques dans la passé, aussi nous nous garderons bien d'y répondre de façon tranchée. Disons, cependant, que si l'autonomie d'un agent se mesure à son imprédictibilité comme le suggère Luciano Floridi et Jeff Sanders⁹, il ne fait aucun doute que les agents programmés par les techniques d'intelligence artificielle sont autonomes.

A titre personnel, je travaille dans mon laboratoire sur des agents qui évoluent dans le temps, en fonction de leur histoire, et qui apprennent soit à mieux servir leur maître, soit à défier leurs adversaires à des jeux le plus efficacement possible, en anticipant leurs réactions¹⁰. Ces agents m'apparaissent intéressants pour notre propos car ils n'ont pas d'essence, au sens où ils ne possèdent pas de personnalité bien définie, mais ils s'en forgent une à partir de leur histoire. On peut dire, d'une certaine façon, que ces agents sont existentiels, au sens où leur existence précède leur essence. J'appelle ces agents du terme un peu impertinent, je le confesse, de sartriens ; et ceux-ci apparaissent parfaitement imprévisibles, tout en se présentant comme doués d'une âme toute rationnelle, dans l'acception aristotélicienne du terme. Ce sont, au sens où nous l'avons défini, des agents autonomes.

Bref, ces quelques exemples illustrent parfaitement la thèse que nous défendons ici, à savoir que les machines possèdent bien une âme au sens aristotélicien. Et, de même que, tant que l'on prenait le concept d'âme dans son acception aristotélicienne, personne ne songeait à dénier aux animaux une âme, de même il n'y a aucune raison de refuser d'attribuer une âme aristotélicienne aux machines, à raison de leurs facultés cognitives.

Au reste, Aristote admettait l'existence d'une gradation naturelle entre les êtres, dans le règne du vivant. D'après lui, les plantes disposaient uniquement d'une âme nutritive ; les animaux possédaient en outre une âme qualifiée de sensitive et leurs actions répondaient aux sensations de façon à satisfaire au mieux leurs appétits ; et les hommes avaient une âme délibérative, capable de calculer les actions réalisant leurs volontés. Ajoutons à cela qu'il y a aussi, parmi les animaux et les hommes, une gradation, selon

⁹ L. Floridi, et J. Sanders, « On the Morality of Artificial Agents », *Minds and Machines*, Kluwer Academic Publishers, Hingham, MA, USA, 14.3, p. 349–379, 2004.

¹⁰ Ch. Meyer, J.-G. Ganascia, J.-D. Zucker, « Learning Strategies in Games by Anticipation », actes de la 15^e conférence internationale d'intelligence artificielle – International Joint Conferences on Artificial Intelligence, IJCAI'97, Morgan Kaufman, 1997.

que les facultés perceptives ou cogitatives apparaissent plus ou moins développées.

Symétriquement, on a identifié une gradation dans l'univers des agents artificiels : ceux-ci se distinguent et s'ordonnent selon le nombre plus ou moins élevé d'états internes qu'ils manifestent, selon le perfectionnement de leurs organes de perception, selon leurs plus ou moins grandes capacités d'action et selon leurs aptitudes à la ratiocination. Ainsi, Daniel Dennett¹¹ établit l'existence d'une hiérarchie d'agents, partant de simples thermostats, qui possèdent deux états internes – Etat 1 : « la température est supérieure au seuil fixé », Etat 2 : « la température est inférieure au seuil fixé » –, une seule action à leur disposition – activer la commande de chauffage – et un capteur élémentaire évaluant la température, à des robots extrêmement perfectionnés, possédant des millions, voire des centaines de millions, d'états, des désirs, une volonté et des capacités de raisonnement.

La conscience des machines

Toutefois, la question ne se résume pas à celle de l'âme au sens aristotélicien ; sinon la querelle de l'âme des animaux n'aurait pas eu d'objet. Et, à examiner certains débats actuels autour de la condition animale, cette querelle ne s'est pas complètement éteinte, même si les termes du débat ont changé : la dispute ne porte plus aujourd'hui sur le concept d'âme à proprement parler, mais sur la conscience et surtout sur la souffrance animale. Et cela reconduit les querelles anciennes. Comme nous l'avons précédemment évoqué, la souffrance des bêtes, réduites par les hommes à l'esclavage pour satisfaire leurs propres besoins, fût à l'origine des débats autour de l'âme animale : comment imaginer qu'un Dieu bon put tolérer et même encourager ainsi le mal ? Et, aujourd'hui, il en va de même : c'est la souffrance animale dans les expérimentations médicales qui motive la réflexion. Comment tolérer que, pour le bien de l'humanité, on sacrifie aussi systématiquement des êtres innocents ?

Or, la question se reformule aussi dans le registre de l'artificiel : est-ce que les automates androïdes, qui apparaissent si semblables à nous, et les autres créatures que l'homme fabrique possèdent une conscience ? Et, si oui, sont-ils susceptibles de souffrir du sort que nous leur réservons ? Cela fit l'objet de nombreuses discussions dans les cinquante dernières années et le débat court toujours. Il suffit, pour s'en convaincre, de lire l'édition du bulletin de l'*American Philosophy Association* publiée en novembre 2007 : il s'y trouve un article de Riccardo Manzotti intitulé *Towards Artificial Consciousness*¹²,

¹¹D. C. Dennett, 1994, « The practical requirements for making a conscious robot. », Londres, Royaume-Uni, *Philosophical Transactions, The Royal Society* (349), p. 133-146.

en français, « Vers une conscience artificielle », autrement dit, vers une machine douée de conscience.

Hésitation des scientifiques

Notons, tout d'abord, que les scientifiques et les techniciens restèrent, pour beaucoup, assez prudents. Ainsi, Alan Turing éluda la question de la conscience des machines dans le célèbre article qu'il écrivit en 1950 sur l'intelligence des machines¹³. Pour lui, un ordinateur doué d'intelligence ne possède pas nécessairement une conscience ; d'ailleurs, selon lui, nous ne sommes pas certains que nos semblables soient tous dotés d'une âme, même si nous leur concédons certaines facultés intellectuelles ; et il en va de même pour les animaux en général et, *a fortiori*, pour les machines.

Dans un ordre d'idées analogue, les pionniers de l'intelligence artificielle que furent Herbert Simon et Alan Newell¹⁴ firent, en 1958, des prédictions jugées, par la suite, hâtives et inconsidérées. En particulier, ils annoncèrent qu'avant dix ans :

- une machine serait championne du monde au jeu d'échecs,
- les théories de psychologie prendraient toutes la forme de programmes informatiques,
- des théorèmes mathématiques originaux seraient démontrés par des machines,
- les machines composeraient de la musique douée de grande valeur esthétique...

En dépit de l'audace un peu inconsidérée dont ils firent preuve, et qui leur fût sévèrement reprochée par la suite, aucun d'entre eux n'imagina doter les machines d'une conscience.

Or, là où les scientifiques esquivent les difficultés, les philosophes relèvent le gant et s'affrontent. Parmi eux, le combat fait rage. Il ne s'agit pas de dresser ici la chronique de la grande querelle de l'âme des machines, car on déborderait très largement la place qui nous est impartie ici. L'histoire de ces disputes mériterait certainement qu'on la relate en détail, car cela donnerait un aperçu assez complet des débats qui animent le champ philosophique contemporain. Nous nous contenterons simplement d'évoquer ici quelques uns des épisodes les plus imagés de cette querelle.

¹² R. Manzotti, « Towards Artificial Consciousness », *APA Newsletter on Philosophy and Computers*, rédacteur Peter Boltuc, Newark, DE, USA, vol. 7, Number 1, 12-15, 2007.

¹³ A. Turing, « Computing Machinery and Intelligence », Oxford, Royaume-Uni, *Mind*, vol. 49, p. 433-460, 1950.

¹⁴ H. Simon et A. Newell, « Heuristics Problem Solving : the Next Advance in Operations Research », Cambridge, MA, USA, *Operation Research*, vol. VI, p. 1-10, 1958.

Les zombies de Chalmers

Le premier tient à la notion de zombies philosophiques qu'introduisit David Chalmers¹⁵. On sait que dans la mythologie haïtienne, un zombie est un être mort, dont l'âme a quitté le corps, et qui s'anime soudain d'une vie animale, toute mécanique. On voit quelquefois ces zombies dans les champs où ils travaillent gentiment. Ils se confondent alors avec les autres hommes et vaquent aux occupations ordinaires des paysans. Apparemment, rien ne les en distingue, si ce n'est qu'ils ne possèdent plus d'âme.

Avec les machines, il se pourrait qu'il en aille identiquement : elles s'animent d'une vie mécanique et ressemblent en tous les points de leurs activités à des êtres doués de conscience ; et, au cas – fort hypothétique – où elles feraient illusion et seraient dépourvues de conscience, elles se présenteraient à nous exactement comme les zombies. Certes, beaucoup doutent que de tels zombies philosophiques existent matériellement aujourd'hui, même si les machines nous surprennent parfois. Mais qu'en vaut-il de la possibilité logique de tels zombies ? Y a-t-il une caractéristique qui distingue ces zombies, sans conscience, d'êtres conscients ? Et, si c'est le cas, cela nous permettrait-il de décider si oui ou non les machines ont une âme ? Telles sont les questions posées par David Chalmers et auxquelles beaucoup¹⁶ ont essayé d'apporter une réponse.

La substitution progressive des neurones

Parmi ceux qui s'y frottèrent, figure Zénon Pylyshyn. Ce philosophe imagina une curieuse expérience de pensée : la substitution progressive des neurones. Ainsi, il prit virtuellement une personne en pleine santé, il lui ouvrit le scalpe puis la boîte crânienne, et il substitua, à son insu, l'un de ses neurones par un dispositif matériel équivalent. S'en apercevrait-elle ? Cela changerait-il quelque chose à la conscience qu'elle aura du monde ? Un neurone sur 100 milliards ne représente pas grand-chose, surtout si on le remplace par un dispositif qui a les mêmes fonctions matérielles. Il y a donc tout à parier que cette personne ne ressent aucune modification dans son état de conscience. Si l'on procède le lendemain de la même façon, adviendrait-il quelque chose ? Deux neurones, cela ne fait pas beaucoup plus qu'un. Il ne devrait donc rien se passer de plus. Imaginons alors que l'on continue régulièrement à remplacer un à un des neurones naturels par des dispositifs matériels

¹⁵ D. Chalmers, *The Conscious Mind : In Search of a Fundamental Theory*, New York and Oxford, Oxford University Press, 1996.

¹⁶ D. Dennett, « The Unimagined Preposterousness of Zombies », Exeter, Royaume-Uni, *Journal of Consciousness Studies*, 2 (4). 322-326, 1995.

T.C Moody, « Conversation with Zombies », Exeter, Royaume-Uni, *Journal of Consciousness Studies*, 1 (2). 196-200, 1995.

équivalents. Y a-t-il un moment où la conscience s'en trouvera modifiée ? Cette modification sera-t-elle brusque ou progressive ? Comment se manifestera-t-elle ? Sans entrer dans les détails, si la conscience ne disparaît pas lorsque tous les neurones auront été substitués et si elle demeure équivalente à ce qu'elle était au début de l'expérience, nous disposerons alors d'une machine totalement artificielle, puisqu'elle ne sera constituée que de composants matériels fabriqués par l'homme, et douée de conscience. Conduire dans l'imagination une telle expérience de pensée porte sur la possibilité logique d'une machine dotée artificiellement de conscience. Aucune et aucun d'entre vous d'ailleurs ne saurait affirmer, avec certitude, qu'il n'a pas subi de telles interventions chirurgicales. Indiquons toutefois, pour rassurer ceux qui risqueraient de s'en inquiéter, que cela demeure une possibilité logique mais qu'en pratique il y a cent milliards de neurones et deux cents milliards de cellules gliales. Cela supposerait donc beaucoup d'interventions chirurgicales. De plus, chaque neurone étant connecté à entre 1000 et 10000 autres neurones par des liaisons synaptiques, cela rendrait les opérations de raboutage assez délicates et, en conséquences, assez lentes... Bref, l'état actuel de la technique chirurgicale rend très improbable de telles interventions, et encore plus improbable leur innocuité.

L'argument de la chambre chinoise

L'un des pourfendeurs officiels de l'intelligence artificielle, John Searle, juge qu'il n'y a aucune impossibilité logique à ce qu'une machine possède une conscience. Néanmoins, d'après lui, pour fabriquer une telle machine, on ne pourrait se contenter de mimer des manipulations de symboles, comme le font les techniques d'intelligence artificielle. Sa démonstration repose sur la célèbre expérience de pensée de la chambre chinoise¹⁷. Searle enferme un américain typique, qui comme tout américain qui se respecte ne parle qu'une seule langue, l'anglais, dans une prison chinoise, sans aucun contact direct avec quiconque. Cet américain est censé répondre à des messages écrits en chinois qu'il observe de la lucarne de sa cellule. Il dispose de carreaux de céramique sur lesquels sont inscrits des caractères chinois. Il doit les présenter à la lucarne de sa cellule en obéissant à des règles de manipulations consignées dans un grand registre et qui disent : « si telle séquence de caractères m'est présentée, alors je montre telle séquence de carreaux à la lucarne ». On ne le nourrit qu'au prix de telles manipulations ; il s'attèle donc à sa besogne avec ardeur. De l'extérieur, si le catalogue de règles est bien conçu, on pourrait imaginer que l'américain comprend le chinois, car il répond de façon totalement pertinente aux questions posées. Mais d'après Searle il n'en va pas ainsi et il n'en ira jamais ainsi, même après un temps

¹⁷ J. Searle, « Minds, Brains and Programs », *Behavioral and Brain Sciences*, 3, Cambridge, Cambridge University Press, 1980, p. 417-457, <http://www.bbsonline.org/documents/a/00/00/04/84/bbs00000484-00/bbs.searle2.html>

très long : notre américain ne comprendra jamais le chinois. Et, comme les machines conçues par l'intelligence artificielle se comportent identiquement, puisqu'elles manipulent des séquences de symboles conformément à un catalogue de règles, elles ne comprennent et ne comprendront jamais ce qu'elles disent, même si elles simulent parfaitement les capacités intellectuelles humaines au point de nous tromper.

Revenons sur l'argument employé par John Searle pour justifier sa réponse : d'après lui, l'intelligence artificielle met en œuvre des règles syntaxiques sur des jeux de symboles sans y associer de sémantique, et cela ne suffit pas pour reproduire les mécanismes d'attribution de signification. Or, toujours selon Searle, il se pourrait qu'une machine parvienne à reproduire les mécanismes physiologiques à l'origine de la sémantique, mais à condition d'avoir recours à d'autres techniques que celles de l'intelligence artificielle. Le rapprochement avec la querelle de l'âme des bêtes tient essentiellement ici à l'argument employé par Searle pour discréditer l'intelligence artificielle : à y bien réfléchir, il apparaît qu'il ressemble fort à celui de Descartes lorsqu'il distingue les animaux-machines, qui n'ont pas d'âme, des hommes doués d'âme. En effet, pour Searle, il existerait, tant chez l'homme que chez les animaux doués de conscience, l'équivalent d'une glande pinéale produisant des humeurs subtiles. Plus exactement, pour ce philosophe, la conscience ne se résume pas à de simples manipulations de symboles ; elle s'ancre sur un niveau de finesse équivalent à celui de la chimie, et notamment sur la simulation des processus physiologiques.

En résumé, l'actuelle querelle de l'âme des machines fait écho à la querelle de l'âme des bêtes. Le malin génie de Descartes, qui fût à l'origine de la dispute dans les temps modernes, se trouve de nouveau invoqué à l'époque contemporaine dans la querelle de l'âme des machines. Et la mystérieuse glande pinéale, par le truchement de laquelle l'âme s'accrocherait au corps, se trouve remplacée par une non moins mystérieuse théorie de l'émergence qui ferait surgir la conscience de processus chimiques élémentaires.

Au-delà

En conclusion de cette contribution, il convient de revenir sur l'objet ultime de la querelle à son origine : la vie éternelle. Cette question paraît bien incongrue aujourd'hui, à l'heure où l'existence d'un au-delà habité par les âmes des morts ne fait plus vraiment l'objet de disputes philosophiques. Pourtant, elle motiva Descartes, qui s'inquiétait de voir des êtres aussi répugnants que les cafards, les rats et les vipères peupler le paradis aux côtés des hommes. Pour évacuer la question gênante de l'immortalité de l'âme des animaux, il assimila les animaux à des machines matérielles dépourvues d'âme.

Or, en dépit des apparences, cette question de la survie de l'homme et de son immortalité devrait toujours soucier le philosophe, ou à tout le moins, l'interroger. Certes, la plupart des hommes de sciences se résignent aujourd'hui à admettre la finitude de la vie : sans aucun doute, nous sommes mortels. Non seulement les individus le sont, nous en convenons tous depuis longtemps, mais à l'heure où l'équilibre économique et écologique planétaire se trouve menacé de toutes parts, la survie de l'espèce humaine apparaît elle-même de plus en plus problématique. Il viendra certainement un temps où notre descendance s'éteindra. Les machines nous aideront peut-être à reculer quelque peu ce terme inéluctable. Pourtant, ce ne sera jamais qu'un petit délai accordé à l'espèce humaine.

Or, qu'advient-il une fois qu'elle aura disparu ? Que léguerons-nous à ceux qui viendront après ?

Il y a toutes les chances que des espèces plus résistantes comme les rats, les cafards ou les serpents prennent notre place. Et, vu le mépris dans lesquels nous les avons tenues, ce dont atteste d'ailleurs la querelle de l'âme des bêtes, il n'est pas certain qu'elles songent à conserver des témoignages de l'existence que nous avons menée et de notre culture.

Il semble donc que si nous voulons voir l'espèce humaine perdurer au-delà de sa disparition, il faut s'en préoccuper dès maintenant. Or, si les êtres naturels ne prennent pas soin de nous, nous devrions prendre nous-mêmes notre destin en main et fabriquer des êtres artificiels qui transmettront notre héritage. Et ces êtres artificiels ne peuvent être que des machines fabriquées par nous pour veiller à la survie de nos âmes.

Mais, là encore, deux cas de figure doivent être envisagés : soit ces machines ont une âme, soit elles n'en ont pas. Dans le premier cas, ayant été conçues, élevées et choyées par nous et pour nous, nous pouvons espérer qu'elles nous en seront définitivement reconnaissantes et qu'elles transmettront notre souvenir pour l'éternité. Dans le second, elles conserveront notre mémoire pour les siècles des siècles en attendant que d'autres adviennent, s'en emparent et s'en émeuvent. Dans tous les cas, la question de l'immortalité de l'âme apparaît comme un enjeu majeur de la querelle de l'âme des machines. Sans aucun doute, cet enjeu demanderait à être creusé plus profondément – et plus sérieusement – qu'il ne l'a été jusqu'ici, car pour l'heure, seule la science fiction semble s'en être vraiment emparée.

Citons, en guise d'épilogue, et pour attester de l'évocation de cette question par la science fiction, l'une des répliques du film *AI – artificial intelligence* de Steven Spielberg, prononcée par un robot après qu'il ait découvert un enfant androïde ayant passé plus de 2000 ans sous la glace : « Les êtres humains ont créé dans les arts, dans la poésie, dans les formules mathématiques, un million d'explications sur la signification de la vie.

Certainement, les êtres humains doivent être la clef de la signification de l'existence, mais les êtres humains n'existent plus»¹⁸

¹⁸ « Human beings had created a million of explanations of the meaning of life in art, in poetry, in mathematical formulas. Certainly, human beings must be the key to the meaning of existence, but human beings no longer existed » (début de la troisième partie du film, après que David ait passé plus de 2000 ans sous la glace – paroles prononcées en voix off et attribuées à un robot).

Edito : Comment le cerveau prend-t-il conscience du monde ?

Vendredi, 16/09/2016 - 01:00



Il y a quelques semaines, une passionnante étude publiée dans le Plos (voir [PLOS](#)) est venue relancer le vieux débat sur la nature de la conscience et de sa production par notre cerveau. Ces travaux, menés par les équipes de Michael Herzog (Ecole polytechnique fédérale de Lausanne) et de Frank Scharnowski de l'Université de Zurich, montrent qu'il faut en moyenne 400 millisecondes à notre cerveau pour transformer un stimulus en une perception consciente.

S'appuyant sur leur découverte, ces chercheurs proposent un modèle en deux phases pour expliquer comment la conscience procède. Face à un stimulus sensoriel, le cerveau commence par analyser de façon inconsciente les caractéristiques physiques et spatio-temporelles de l'objet de la sensation (couleur, forme, poids...). A ce niveau inconscient de traitement des informations sensorielles, ce modèle exclut la perception du temps.

Selon cette hypothèse, c'est seulement dans un deuxième temps que notre cerveau rendrait simultanément conscients tous les éléments, pour présenter, in fine, une image cohérente et globale à notre conscience. Ces chercheurs soulignent que ce décalage temporel de 400 millisecondes entre ces deux étapes est tout à fait considérable d'un point physiologique. Selon eux, l'importance de ce délai s'expliquerait par le fait que notre cerveau, avant de rendre accessible une information à notre conscience, veut la rendre la plus cohérente et la plus heuristique possible.

Ces recherches sont passionnantes car elles montrent que la conscience n'est pas abolie par ces phases de traitement inconscient, et elles permettent de dépasser l'opposition entre ceux qui pensent que la conscience est un phénomène continu et les tenants d'une conception discontinue de la conscience. Ces travaux éclairent également d'une lumière nouvelle la façon dont le cerveau traite le temps et intègre cette dimension temporelle dans notre perception du monde.

Cette découverte est à mettre en relation avec trois autres récentes avancées en neurosciences. La première a été réalisée en juillet 2014 par Mohamad Koubeissi et son équipe de l'Université de Washington qui ont réussi, pour la première fois, à éteindre et rallumer la conscience d'une femme, atteinte d'épilepsie, en stimulant son claustrum (Voir [GW TODAY](#)).

Cette découverte renforce l'hypothèse selon laquelle une seule zone cérébrale, le claustrum, une fine couche de matière grise, serait fortement impliquée dans l'organisation et la mise en cohérence de nos pensées, sensations et émotions que nous percevons, in fine comme une expérience unifiée et non comme des perceptions isolées.

Dans ces expériences, les chercheurs ont disposé et utilisé des électrodes près du claustrum, une zone qui n'avait jamais été stimulée. Ils ont observé qu'en stimulant cette zone précise du cerveau, la patiente perdait immédiatement conscience : elle devenait incapable de répondre à une commande visuelle ou auditive et voyait sa respiration ralentir. Mais ce processus semble réversible à volonté car, lorsque les scientifiques cessaient cette stimulation électrique, cette femme reprenait immédiatement conscience, sans avoir aucun souvenir de son état précédent. Mohamad Koubeissi pense donc que le claustrum joue un rôle majeur dans l'activation des expériences conscientes. Ces travaux ont également montré que la perte de conscience s'accompagnait d'un renforcement de la synchronisation des ondes cérébrales dans les régions frontale et pariétale, connues pour être impliquées dans l'activation de la sensibilité.

La seconde découverte très intéressante, elle aussi, vient d'être réalisée par l'équipe de Thomas Andrillon, de l'ENS, à Paris. Ces chercheurs ont voulu comprendre pourquoi notre cerveau avait besoin à ce point de se déconnecter de manière aussi nette de son environnement pendant notre sommeil.

Dans leur expérience, les chercheurs ont enregistré par ECG (électroencéphalographie), l'activité cérébrale de 23 volontaires soumis à des listes de mots désignant des animaux ou des objets. Tant qu'ils étaient éveillés, les sujets avaient pour consigne d'appuyer sur un bouton blanc lorsqu'ils entendaient un nom d'animal et sur un bouton rouge, lorsqu'ils entendaient un nom d'objet. Mais ce que les chercheurs voulaient comprendre, c'est ce qui se passait dans le cerveau de ces volontaires quand ils s'endormaient, tout en continuant à entendre ces listes de mots.

En analysant les ECG des sujets, les chercheurs ont observé qu'au cours du stade qui suit l'endormissement, la phase dite de sommeil lent léger, le signal cérébral montrait une forme caractéristique d'un passage à l'action dans la région du cortex moteur contrôlant la main gauche ou droite, selon le type de mot entendu. Cela voulait donc dire que le cerveau continuait son travail d'analyse des mots et préparait le geste visant à appuyer sur le bon bouton, mais sans aller toutefois jusqu'à accomplir réellement ce geste.

Dans les deux phases suivantes, le sommeil lent profond et le sommeil paradoxal, les aires cérébrales traitant les propriétés physiques du signal (intensité, fréquence...) continuaient à s'activer mais les signaux ainsi générés n'étaient pas communiqués dans les autres régions du cerveau. Au cours de ces deux phases, le cerveau semblait donc essayer de maintenir un isolement assez strict, pourquoi ? Sans doute, selon les chercheurs, parce que ces phases de sommeil lent et paradoxal jouent un rôle essentiel dans la consolidation des souvenirs et qu'il est donc très important – et prioritaire pour notre cerveau à ce moment là – que des stimuli sensoriels ne viennent pas le « parasiter » et l'empêcher de réaliser ce travail fondamental de fixation et de mémorisation profondes d'informations utiles.

La troisième découverte, elle aussi tout à fait fascinante, a été réalisée récemment par une équipe de chercheurs de l'Université de Californie à San Francisco (États Unis). Ces chercheurs ont pu montrer que notre cerveau est tout à fait capable de réaliser des tâches, même complexes, sans qu'il y ait eu de décision volontaire de la part des sujets de cette expérience. Pour parvenir à ces conclusions, ces chercheurs ont demandé à 32 étudiants volontaires d'apprendre le pig latin, une sorte d'argot principalement utilisé en anglais et qui combine l'inversement des syllabes (comme dans le « Verlan ») et l'ajout d'une syllabe (comme dans le « Javanais »).

Pour un mot qui commence par une consonne, on déplace cette consonne à la fin du mot et on ajoute — ay. Par exemple, le mot « raison » devient « aison-ray ». Pour un mot qui commence par une voyelle, on ajoute — way à la fin du mot. Dans ce cas, le mot « oiseau » devient « oiseau-way ». Au cours de ces expériences, les sujets ne devaient surtout pas traduire volontairement les mots qui s'affichaient sur leur écran de contrôle. Et si l'un des étudiants le traduisait quand même malgré lui, il devait appuyer sur une touche dédiée de son ordinateur. Le résultat de cette expérience fut pour le moins surprenant et montra que dans 43 % des cas, les étudiants avaient traduit les mots affichés, même lorsqu'ils avaient décidé de ne pas effectuer cette traduction...

A la lumière de ces résultats, les chercheurs font l'hypothèse que notre conscience, si elle peut sans doute produire de manière intrinsèque des informations, jouerait également – et peut-être essentiellement – un rôle de « passerelle » permettant l'accès à des pensées déjà formées dans notre inconscient. Reste que ce phénomène de la conscience recouvre des niveaux de perception et d'action différents qui ne se laissent pas facilement enfermer dans une approche théorique unique et globale.

Comme le rappelle Alain Berthoz, Professeur au collège de France, « Le cerveau ne cesse de produire des modèles internes du corps et du monde » et cet éminent scientifique souligne que cette conscience que nous avons du monde, qui nous semble naturelle, globale et immédiate, n'est jamais entièrement séparée de notre corps.

Pour Alain Berthoz, cette interaction et cette confrontation continues entre notre individualité physique et psychique et le monde entretiennent et produisent une interactivité permanente entre, d'une part, nos capacités et fonctions cognitives et, d'autre part, nos possibilités d'action. Dans ce cadre conceptuel, la conscience devrait donc être considérée comme un processus à plusieurs niveaux nous permettant de prendre des décisions logiques, pertinentes et articulées dans un ensemble cohérent intégrant des mécanismes inconscients et conscients.

Pour mieux illustrer sa théorie, Alain Berthoz prend souvent l'exemple des réflexes sensorimoteurs. Il souligne que ceux-ci, contrairement à l'opinion dominante qui a longtemps prévalu, peuvent être contrôlés si nécessaire. C'est ainsi que, lorsque nous bougeons la tête, nos yeux se déplacent de manière réflexe dans le sens contraire du mouvement de la tête, ce qui permet une stabilisation de la vision. Mais nous pouvons tout à fait, si nous le voulons, annuler, débrayer en quelque sorte, ce mouvement

réflexe, par exemple pour continuer à voir le bout de notre nez même en bougeant la tête. En revanche, si nous voulons observer un point très éloigné dans l'espace, nous saurons alors utiliser au maximum ce mouvement réflexe. Cet exemple simple mais très clair montre bien que nous avons la capacité de contrôler un réflexe en fonction de l'action que nous visons et du niveau de contrôle conscient que nous exerçons sur cette action.

Reste que la nature et la genèse de ce lien entre le corps, les émotions et la réflexion reste un mystère et demeure un sujet de débat passionné depuis l'Antiquité, tant chez les philosophes que chez les scientifiques. Le grand neurobiologiste américain Antonio Damasio propose pour sa part sa théorie des marqueurs somatiques qui tente d'articuler les processus physiques, émotionnels et cognitifs. Selon lui, notre cerveau, pour prendre la décision la plus adaptée à une situation, ne cesserait de combiner, dans une boucle neuronale incluant les changements corporels liés aux émotions, des processus cognitifs rationnels et des processus de reconstruction de souvenirs liés aux émotions passées.

Selon Damasio, notre cerveau, lois d'être un simple « transducteur » conçu pour nous fournir une image cohérente du monde et guider notre action, serait un véritable émulateur reconstruisant sans cesse la « réalité », en s'appuyant sur notre expérience, nos affects et notre réflexion et en produisant des liens d'unification des processus conscients et inconscients et l'incorporation - au sens propre du terme - des lois fondamentales de la nature régissant l'espace, le temps et le mouvement.

C'est ainsi que, très tôt, un enfant comprend instinctivement que pour mieux recevoir un objet lourd qui lui est lancé, il a intérêt à accompagner avec ses bras et son buste la réception de cet objet par un mouvement qui lui permettra de mieux absorber l'énergie cinétique de l'objet en question. De même, si cet enfant saute d'une grande hauteur sur le sol, il comprendra très vite qu'il doit bien plier les jambes et éventuellement faire une roulade pour ne pas se blesser. Dans les deux cas, cet enfant aura incorporé, sans en avoir conscience, deux lois fondamentales de la physique et du monde dans lequel il vit.

La théorie de Damasio se veut clairement en rupture avec la conception cartésienne de la nature et s'appuie sur la philosophie de Spinoza, pour qui une perception est une affection de notre corps par le monde extérieur. Par exemple, explique le philosophe hollandais, quand nous percevons le soleil comme un disque jaune, nous percevons en réalité une modification de nos yeux et cette image nous en apprend plus sur la nature de l'œil que sur celle du soleil. Par conséquent, toute connaissance est produite et s'inscrit d'abord dans une dimension corporelle et subjective, même si elle ne s'y réduit pas : c'est la fameuse affirmation de Spinoza : « L'esprit ne se connaît lui-même qu'en tant qu'il perçoit les idées des affections du corps. »

Il est intéressant de constater que ces avancées scientifiques remarquables dans le domaine de la neurobiologie vont dans le sens des intuitions de Spinoza mais également d'Husserl, le père de la phénoménologie. Ces deux grands penseurs, pourtant très éloignés sur de nombreux points, se rejoignent sur la conviction que la conscience est toujours intentionnelle et porte toujours sur un objet : il n'y a pas de conscience pure et comme le dit Husserl : « Toute conscience est conscience de quelque chose ».

Ce que nous apprennent finalement les neurosciences, dans leurs dernières avancées, c'est que ce mystérieux phénomène, à la fois si évident et si insaisissable, que nous appelons conscience, ressemble fort à un processus par lequel tout notre être, dans ses dimensions corporelles, spirituelles, intellectuelles et affectives, coproduit la réalité et réinterprète en permanence le monde, affirmant ainsi sa subjectivité et sa puissance d'exister, dans un dialogue infini et inépuisable avec lui-même, les autres et la Nature.

René TRÉGOUËT

Sénateur honoraire

Fondateur du Groupe de Prospective du Sénat