

# « *Intelligence artificielle et intelligence humaine : recherches actuelles et perspectives* »

Jean-Paul Delahaye

Université de Lille, CRISTAL UMR 9189 CNRS

**Le Président :** C'est un grand honneur pour notre institution de recevoir le professeur Jean-Paul Delahaye dans le cadre de notre cycle 2018-2019 « Dépasser l'humain ».

Après la communication du professeur Michel Blay sur « l'origine de la notion de progrès », vous allez nous parler des « recherches actuelles et des perspectives autour de l'intelligence artificielle et de l'intelligence humaine ».

Vous êtes mathématicien de formation. Vous obtenez l'agrégation de mathématiques, puis un doctorat de troisième cycle sur la même matière, en mathématiques, et un doctorat d'État.

Depuis 1988, vous êtes professeur d'informatique à l'Université de Lille et chercheur au sein du *LIFL* (Laboratoire d'informatique fondamentale de Lille, CNRS) devenu récemment *CRISTAL* (Centre de recherche en informatique signal et automatique de Lille, CNRS) rattachés à cette université. Depuis quatre ans vous êtes Professeur émérite, mais toujours membre de *CRISTAL*.

Vous travaillez sur la théorie computationnelle des jeux depuis 1996 et sur le hasard depuis 2004.

Vos anciens thèmes de recherches étaient la théorie des transformations de suites (de 1977 à 1982), conclu par votre thèse d'État et la programmation logique en lien avec l'intelligence artificielle (de 1982 à 1998).

Vous êtes également :

- rédacteur des articles mathématiques de la chronique *Logique et calcul* de la revue « Pour la science »
- conseiller scientifique en mathématiques et auteur d'articles pour l'Encyclopædia Universalis<sup>3</sup>,
- auteur contributeur de la revue de culture scientifique en ligne *interstices*<sup>4</sup>,
- Enfin vous proposez aussi un blog (<http://www.scilogs.fr/complexites/>) consacré aux "Complexités".

Depuis 2016, vous êtes est membre du Comité d'éthique du CNRS.

Vous travaillez aujourd'hui sur les monnaies cryptographiques et la "technologie blockchain".

Vous avez encadré une vingtaine de thèses, et vous êtes l'auteur d'une quinzaine de livres, dont une partie est destinée à un large public.

En 1998, vous avez reçu le Prix d'Alembert de la *Société Mathématique de France* et, en 1999, le Prix Auteur de la Culture scientifique du *Ministère de l'Education Nationale et de la Recherche*.

Vos derniers ouvrages sont :

- Les mathématiciens se plient au jeu, *Éditions Belin/Pour la science*, 2018.
- Mathématiques et mystères : Cryptographie, étrangeté des nombres, conjectures, etc. *Éditions Belin/Pour la science*, 2016.
- Inventions Mathématiques, *Éditions Belin/Pour la science*, 2014.
- Merveilleux nombres premiers. *Éditions Belin/Pour la science*, Seconde édition revue, mise à jour et complétée, 2013.

- Culturomics : le numérique et la culture, *Éditions Odile Jacob*, 2013. (co-auteur Nicolas Gauvrit).
- TOUT. Les rêves mathématiques d'une théorie ultime, *Éditions Hermann*, 2011
- Jeux finis et infinis, *Éditions du Seuil*, 2010.

Vous vous êtes défini comme universitaire, passionné des sciences et des idées, tout en vous intéressant à des actions de culture scientifique.

Enfin, vous aimez la musique de Jean-Sébastien Bach.

## Une seule intelligence ?<sup>1</sup>

**Jean-Paul Delahaye :** Affirmer qu'il y a plusieurs types d'intelligence plaît beaucoup car cette idée évite à chacun de se trouver en un point précis d'une échelle absolue, et que chacun espère exceller dans l'une des formes d'intelligence dont la liste tend à s'allonger. Cette idée a son théoricien, le psychologue américain Howard Gardner qui, à partir de 1979 a proposé une classification des types d'intelligence. Dans son livre « Frame of Mind » de 1983, bible de la « théorie des intelligences multiples », il énumère huit sortes d'intelligences. Très critiquée, par exemple par Perry Klein de l'université d'Ontario qui la considère tautologique et infalsifiable, cette théorie est à l'opposé d'une autre voie de recherche qui affirme qu'il n'existe qu'une sorte d'intelligence qu'il faut concevoir mathématiquement avec l'aide de l'informatique et de la théorie du calcul.

### Les incontestables succès de l'intelligence artificielle

Avant d'en donner les principes, évoquons l'intelligence des machines et la discipline informatique appelée *intelligence artificielle*, prétexte de nombreux romans et films de science-fiction et sujet de controverses. Il faut l'admettre, aujourd'hui, les machines réussissent des prouesses qu'autrefois personne n'aurait hésité à qualifier d'intelligentes. Nous ne reviendrons pas sur la victoire définitive de l'ordinateur sur les meilleurs joueurs d'échecs, qui en 1997, à la suite de la défaite de Garry Kasparov (champion du monde) face à l'ordinateur Deep Blue, a été unanimement saluée comme un événement majeur de l'histoire de l'humanité. À cette époque, pour se consoler peut-être, certains n'ont pas manqué de remarquer que les meilleurs programmes pour jouer au jeu de go étaient d'une affligeante médiocrité. Il se trouve que, depuis quelques années, des progrès spectaculaires ont été réalisés par l'introduction de nouvelles idées dans les programmes de go. Les machines jouent maintenant très correctement à ce jeu stratégique. En mars 2013, le programme *Crazy Stone* du chercheur français Rémi Coulom de l'Université de Lille a battu le joueur professionnel japonais Yoshio Ishida qui, au départ de la partie avait laissé un avantage de 4 pierres au programme. Depuis, en mars 2016, le programme AlphaGo a battu Lee Sedol considéré comme le meilleur joueur de go. Des idées assez différentes de celles utilisées pour les échecs ont été nécessaires pour cette victoire de la machine, mais pas plus que pour les échecs on ne peut dire que l'ordinateur joue comme un humain. La victoire de AlphaGo est un succès remarquable de l'intelligence artificielle qui prouve d'ailleurs qu'elle avance régulièrement.

Parmi les succès de l'intelligence artificielle, celui concernant le jeu de dames anglaises est absolu. Depuis 1994, aucun humain n'a battu le programme canadien *Chinook*. Mieux, depuis 2007, le programme joue une stratégie optimale – impossible à améliorer – réglant définitivement la question. Pour le jeu d'échec, on sait depuis longtemps qu'il existe aussi des stratégies optimales, mais leur calcul est considéré hors d'atteinte pour plusieurs décennies encore.

L'intelligence des machines ne se limite plus aujourd'hui aux problèmes bien clairs de nature mathématique ou se ramenant à l'exploration d'un grand nombre de combinaisons, cependant il est important de noter que les chercheurs en intelligence artificielle ont découvert, même avec les jeux de tableau cités, qu'imiter le fonctionnement intellectuel humain est difficile et, de fait, impossible avec les techniques dont nous disposons. Pour les jeux de dame, d'échec ou de go et bien d'autres, les programmes ont des capacités équivalentes aux meilleurs humains, mais ils fonctionnent très différemment. Cela ne doit pas nous interdire d'affirmer que nous avons mis un peu d'intelligence dans les machines : il ne serait pas « fair-play », face à une tâche donnée, d'obliger les machines à affronter les humaines en imitant servilement leurs méthodes et modes de raisonnements.

Le cas des véhicules autonomes est remarquable aussi de ce point de vue. Il illustre d'une autre façon que, lorsque l'on conçoit des systèmes nous égalant à peu près pour les résultats, on le fait en utilisant des techniques le plus souvent totalement différentes de celles que la nature a mises en œuvre en nous, et que d'ailleurs nous ne comprenons que très partiellement : pour le jeu d'échec par exemple, personne ne sait décrire les algorithmes qui copient le jeu des champions humains.

### Les véhicules autonomes

---

<sup>1</sup> Ce texte est une version revue et complétée et mise à jour d'un article publié dans la revue *Pour la science* en décembre 2014.

La conduite de véhicules motorisés demande aux êtres humains des capacités qui vont bien au-delà de la simple mémorisation d'une quantité massive d'informations et de l'exploitation d'algorithmes traitant rapidement et systématiquement des données symboliques comme des positions de pions sur un damier. Nul ne doute que, pour conduire comme nous des véhicules motorisés, l'ordinateur doit être capable d'analyser de manière extrêmement fine des images variées et changeantes : où est le bord de cette rue jonchée de feuilles d'arbres ? Quelle est la nature de cette zone noire à 50 mètres au centre de la chaussée, un trou dangereux ou seulement une tache d'huile sale sans importance ? etc.

Conduire une voiture avec nos méthodes demanderait de mettre au point des techniques d'analyse d'images bien plus subtiles que celles que nous savons programmer aujourd'hui. C'est pour cela que les systèmes de pilotage automatisé comme ceux de la firme Google « conduisent » tout à fait différemment des humains. Ces fameuses *Google car* exploitent en continu un système GPS de localisation très précis du véhicule et des *cartes routières* indiquant de manière bien plus détaillée que toutes les cartes usuelles – y compris celles de Google maps ! – la forme et le dessin des chaussées, la signalisation routière, et tous les éléments importants de l'environnement. Les voitures Google exploitent aussi des radars embarqués, des lidars (light detection and ranging, systèmes optiques créant une image numérique en trois dimensions de l'espace autour de la voiture) et des capteurs sur les roues. Ayant déjà parcouru des centaines de milliers de kilomètres sans presque aucun accident, ces voitures sont un succès de l'intelligence artificielle, et cela, même si elles sont incapables de réagir à des signes ou injonction d'un policier au centre d'un carrefour, et qu'elles s'arrêtent parfois brusquement lorsque des travaux sont en cours sur leur chemin modifiant temporairement les voies de circulation. Avec ces machines, on est loin de la méthode de conduite que pratique un être humain. Lui, grâce à son intelligence générale et sa capacité à extraire de l'information des images qu'il voit à travers le pare-brise, sait piloter sur un trajet jamais emprunté, sans cartes, sans radar, sans lidar, sans capteur sur les roues, et ne se trouve pas paralysé au moindre obstacle inopiné !

### **Robots journalistes**

Les questions que nous avons évoquées jusqu'à maintenant n'exigent pas la compréhension du langage écrit ou parlé. Pourtant, contrairement aux annonces de ceux qui considèrent le langage comme une source de difficultés impossibles à surmonter pour les machines, des succès remarquables ont été obtenus dans la réalisation de tâches exigeant une bonne maîtrise des langues naturelles.

L'utilisation des robots-journalistes inquiète parfois la profession car elle est devenue courante dans certaines rédactions comme celles du *Los Angeles Times*, de *Forbes* ou de *Associated Press*. Pour l'instant, ces automates-journalistes se limitent à convertir des résultats (sportifs, ou économiques par exemple) en courts articles. Il n'empêche que parfois on leur doit de très utiles traitements dont voici un exemple. Le 17 mars 2014, un tremblement de terre de magnitude 4,7 se produisit à 6 h 25 en heure locale au large de la Californie. Trois minutes après la secousse, un petit article d'une vingtaine de lignes était automatiquement publié sur le site internet du *Los Angeles Time* donnant, sous la forme d'un texte en anglais, des informations sur l'événement : lieu de l'épicentre, magnitude, heure, comparaison avec d'autres secousses récentes. L'article était le premier à paraître sur l'événement. Il exploitait des données brutes fournies par un service spécialisé (le *US-Geological Survey Earthquake Notification Service*) et résultait d'un algorithme dû à Ken Schwencke, un journaliste programmeur. Interrogé, celui-ci précisa que, d'après lui, ces méthodes ne conduiraient à la suppression d'aucun emploi, mais rendraient au contraire le travail des journalistes plus intéressant. Il est vrai que ces programmes sont pour l'instant confinés à la rédaction d'articles courts exploitant des données factuelles faciles à traduire en petits textes, qu'un humain ne saurait sans doute pas mieux rédiger.

### **Aider Wikipédia**

Un autre exemple inattendu de rédaction automatique d'articles concerne les encyclopédies Wikipedia en langue suédoise et en langue Filipino, une des deux langues officielles aux Philippines (l'autre est l'Anglais). Le programme *Lsjbot*, mis au point par Sverker Johansson, a en effet créé plus de deux millions d'articles de l'encyclopédie collaborative et est capable d'en produire 10 000 par jour. Ces pages générées automatiquement concernent des animaux ou des villes et proviennent de la traduction dans le format imposé par Wikipédia d'informations disponibles dans des bases de données déjà informatisées. L'exploit a été salué mais aussi critiqué. Sverker Johansson, pour se justifier, indique que ces pages peu créatives sont utiles. Il fait aussi remarquer que l'encyclopédie Wikipedia est souvent biaisée par le fait que le choix des articles reflète essentiellement les intérêts des jeunes blancs, de sexe masculin et amateurs de technologies. Il a noté par exemple que l'encyclopédie Wikipedia suédoise comporte 150 articles sur les personnages du livre de

Tolkien, *Le Seigneur des Anneaux*, et seulement une dizaine sur des personnes réelles liées à la guerre du Vietnam : « Est-ce vraiment le bon équilibre ? », demande-t-il.

Johansson projette de générer une page par espèce animale recensée. Pour lui, ces méthodes doivent être généralisées, mais il précise que Wikipedia a besoin aussi de rédacteurs qui écrivent de manière plus littéraire que *Lsjbot*, et soient capables d'exprimer des sentiments, « ce que ce programme ne sera jamais capable de faire ».

### Énigmes astucieuses ?

Beaucoup plus complexe, et méritant mieux l'utilisation du terme *intelligence artificielle* est le succès du programme *Watson* d'IBM au jeu télévisé *Jeopardy !* (une sorte de *Questions pour un champion*). En 2011, le programme réussit en effet à battre les meilleurs joueurs humains du jeu. Cela surprit de nombreux observateurs qui, avant cette victoire de la machine, croyaient inconcevable la maîtrise automatisée du langage et du sens commun à un niveau permettant la résolution d'énigmes linguistiques délicates comme celles proposées dans les épreuves du jeu *Jeopardy !*, qui exigent la maîtrise d'une large culture générale.

La mise au point de programmes résolvant les mots-croisés aussi bien que les meilleurs humains confirme que l'*intelligence artificielle* sait développer des systèmes possédant des performances linguistiques. Cela oblige à reconnaître qu'il faut cesser de considérer que le langage est réservé aux êtres humains ? Malgré ces succès, on est très loin de la perfection. Pour s'en rendre compte, il suffit de se livrer à un petit jeu avec le système de traduction automatique en ligne de *Google*. Une phrase en français (a) est traduite en anglais (b), puis à partir de (b) traduite en français (c). Parfois cela fonctionne parfaitement, on retrouve la phrase initiale ou une phrase équivalente. Parfois, le résultat est catastrophique, comme ici :

- (a) La subtilité manque aux machines qui semblent stupides.
- (b) The lack subtlety machines that seem stupid.
- (c) Les machines manquent de subtilité qui semblent stupide.

### Test de Turing

Les limites des systèmes actuels ont pour conséquence qu'on est très loin aujourd'hui de la mise au point de dispositifs informatiques susceptibles de passer le *Test de Turing* conçu en 1950. Alan Turing voulait éviter de discuter de la nature de l'intelligence. Plutôt que d'en rechercher une définition, il proposait de considérer qu'on aura réussi à mettre au point des machines intelligentes lorsqu'elles seront devenues indiscernables des humains.

Pour tester cette indiscernabilité, il suggérait de faire dialoguer par écrit avec la machine une série de juges qui ne sauraient pas s'ils mènent leurs échanges avec une machine tentant de se faire passer pour un humain, ou un humain véritable. Lorsque les juges ne pourront plus faire mieux que répondre au hasard pour indiquer qu'ils ont eu affaire à un homme ou une machine, le test sera passé. Concrètement, faire passer le test de Turing à un système informatique S consiste à réunir un grand nombre de juges, à les faire dialoguer aussi longtemps qu'ils le souhaitent avec des interlocuteurs choisis pour être une fois sur deux un humain, et une fois sur deux le système S ; les experts indiquent, quand ils le souhaitent, s'ils pensent avoir échangé avec un humain ou une machine. Si l'ensemble des experts ne fait pas mieux que le hasard – donc se trompe dans 50 % des cas ou plus – alors le système S a passé le test de Turing.

Turing, optimiste, pronostiqua qu'on obtiendrait une réussite partielle au test en l'an 2000 : les experts dialoguant 5 minutes et se trompant dans 30 % des cas au moins. Bien évidemment, cette réussite partielle est beaucoup plus facile que la réussite complète attestant l'indiscernabilité (sans limitations de temps pour le dialogue et avec un taux d'erreur d'au moins 50 %). Turing avait, en gros, vu juste. En effet, depuis quelques années la version partielle du test a été passée, sans qu'on puisse prévoir quand sera passé le test complet, chose à propos de quoi Turing n'émettait aucune prédiction.

Le test partiel a, par exemple, été passé le 6 septembre 2011 lors d'une rencontre technologique à Guwahati en Inde par le programme *Cleverbot* créé par l'informaticien britannique Rollo Carpenter. Trente juges dialoguèrent pendant 4 minutes avec un interlocuteur inconnu qui était, dans la moitié des cas, un humain et, dans l'autre moitié, le programme *Cleverbot*. Les juges et les membres de l'assistance (1334 votes) considèrent le programme comme humain dans 59,3 % des cas. Notons que les humains ne furent considérés comme tels que par 63,3 % des votes.

Plus récemment, le 9 juin 2014 à la *Société Royale de Londres*, un test organisé à l'occasion du soixantième anniversaire de la mort de Turing permit à un programme nommé *Eugene Goostman* de duper

dix des trente juges réunis (33 % d'erreur). Victime de la présentation biaisée que donnèrent les organisateurs de cette manifestation, de nombreux articles de presse, dans le monde entier, parlèrent d'un événement historique comme si le test complet d'indiscernabilité homme-machine avait été réussi. Ce n'était pas du tout le cas puisque le test avait d'ailleurs encore été affaibli : l'humain que le système informatique simulait était censé n'avoir que 13 ans et ne pas écrire convenablement l'anglais (car d'origine ukrainienne !).

Non, le test de Turing n'a pas été passé, et il n'est sans doute pas sur le point de l'être. Il n'est d'ailleurs pas du tout certain que les tests partiels fassent avancer vers la réussite au test complet. En effet, les méthodes utilisées pour tromper brièvement les juges sont basées sur le stockage d'une multitude de réponses préenregistrées (correspondant à des questions qu'on sait que les juges posent) associées à quelques systèmes d'analyse grammaticale pour formuler des phrases reprenant les termes des questions des juges et donnant l'illusion d'une certaine compréhension, en fait totalement absente. Quand ces systèmes ne savent plus quoi faire, ils ne répondent pas et posent une question. À un journaliste qui lui demandait comment il se sentait après sa victoire, le programme *Eugene Goostman* de juin 2014 répondit « quelle question stupide vous posez ! Pouvez-vous me dire qui vous êtes ? ».

Aujourd'hui, donc, l'intelligence artificielle réussit assez brillamment à égaler l'humain pour des tâches spécialisées (y compris demandant une certaine maîtrise du langage) ce qui parfois étonne et doit être reconnu comme des succès d'une discipline qui avance régulièrement. Cependant, elle le fait sans vraiment améliorer la compréhension qu'on a de l'intelligence humaine qu'elle ne copie quasiment jamais, ce qui a en particulier comme conséquence qu'elle n'est pas sur le point de proposer des systèmes disposant vraiment d'une intelligence générale, chose nécessaire pour passer le difficile test de Turing qui reste hors de portée aujourd'hui (si on ne le confond pas avec ses versions partielles !).

### À la recherche de l'intelligence générale

C'est cette situation qui donne son intérêt aux recherches tentant de saisir ce qu'est l'*intelligence générale*. Ces travaux sont parfois un peu abstraits, voire mathématiques, mais peut-être est-ce leur force et le seul moyen d'accéder à une notion absolue d'intelligence, indépendante de l'homme.

Quand on tente de formuler une définition générale de l'intelligence vient assez naturellement à l'esprit l'idée qu'être intelligent, c'est repérer des régularités – des structures – dans les données dont on dispose, qu'elle qu'en soit leur nature, ce qui permet de s'y adapter et de tirer le maximum d'avantages de la situation dans laquelle on se trouve. L'identification des régularités, on le sait par ailleurs, permet de compresser des données, et de prédire avec succès les données suivantes qu'on recevra.

*Intelligence, compression et prédiction* sont trois concepts intimement liés les uns aux autres. Si, par exemple, on vous communique les données suivantes : 4, 6, 9, 10, 14, 15, 21, 22, 25, 26, 33, 34, 35, 38, 39, 46, 49, 51, 55, 57, et que vous en reconnaissez la structure, cela vous permet de les compresser en « les 20 premiers produits de deux nombres premiers » et de deviner ce qui va venir : 58, 62, 65, 69, 74, 77, 82, 85, 86, 87, 91, 93, ...

Ce lien entre *intelligence, compression et prédiction* a été exprimé de manière formelle par Ray Solomonoff en 1965. Il proposait une version moderne du principe du rasoir d'Ockham « *Pluralitas non est ponenda sine necessitate* » (Les multiples ne doivent pas être utilisés sans nécessité) qu'il traduisait en un principe plus précis : entre toutes les explications compatibles avec les observations, la plus concise – ou encore la mieux compressée – est la plus probable.

Si on dispose d'une mesure de concision permettant de comparer les théories, la seconde version du principe du rasoir d'Ockham devient un critère mathématique. La *théorie algorithmique de l'information* de Kolmogorov qui propose de mesurer la complexité (et donc la simplicité) de tout objet numérique (une théorie le devient une fois entièrement décrite) par la taille du plus court programme qui l'engendre, donne cette mesure de concision et rend donc possible la mathématisation complète du principe de parcimonie d'Ockham.

L'aboutissement de cette voie de réflexion et de mathématisation est la *théorie générale de l'intelligence* développée par Marcus Hutter qui, en 2005, a publié un livre devenu une référence : *Universal Artificial Intelligence* (Springer-Verlag). En utilisant la notion mathématique de concision, une notion mathématisée d'environnement (ce qui produit les données desquelles un système intelligent doit tenter de tirer quelque chose), et le principe mathématisé d'Ockham, Hutter définit une mesure mathématique universelle d'intelligence. Elle est obtenue comme la réussite moyenne d'une stratégie dans l'ensemble des environnements envisageables. Cette dernière notion (dont nous ne formulons pas ici la version technique

avec tous ses détails mathématiques) est trop abstraite pour être utilisable directement dans des applications. Cependant, elle permet le développement d'une théorie de l'intelligence et fournit des pistes pour comparer sur une même base abstraite, non anthropocentrée et objective, toutes sortes d'intelligences. Contrairement à l'idée d'Howard Gardner, cette voie de recherche soutient que l'intelligence est unique, qu'on peut dépasser le côté arbitraire des tests d'intelligence usuels pour classer sur une même échelle tous les êtres vivants ou mécaniques qui sont susceptibles de posséder de l'intelligence.

Malgré la difficulté à mettre en œuvre pratiquement la théorie (par exemple pour concevoir de meilleurs tests d'intelligence, ou des tests s'appliquant aux humains comme aux dispositifs informatiques), on a sans doute franchi un pas important avec cette théorisation complète d'un problème qui semblait échapper à toute mathématisation. Conscients de son importance pour la réussite du projet de l'intelligence artificielle, les chercheurs ont maintenant créé un domaine de recherche nouveau consacré à ce thème de *l'intelligence artificielle générale*. Il dispose de sa propre revue spécialisée, le Journal of General Artificial Intelligence : <http://www.degruyter.com/view/j/jagi>

Pour éviter à la discipline de se satisfaire du développement de sa partie mathématique, un concours informatique a été créé par Marcus Hutter. Il utilise l'idée que plus on peut compresser plus on est intelligent. Pour gagner et empocher une partie de 50 000 euros mis en jeu, il faut compresser au mieux le contenu de l'encyclopédie Wikipedia considéré comme une image miniature de la richesse de notre univers.

On espère que la nouvelle discipline aidera les chercheurs à réaliser cette *intelligence générale* qui manque tant à nos machines actuelles et les oblige à n'aborder que des tâches spécialisées, le plus souvent en contournant les difficultés qu'il y aurait à employer les mêmes méthodes que les humains, qui eux disposent de cette *intelligence générale*.

### 1. Des tests d'intelligence où la machine gagne

Les tests d'intelligence inventés par le psychologue français Alfred Binet (1857-1911) ne sont sans doute pas un bon moyen de mesurer l'intelligence des machines. En effet des programmes, dont il est impossible de soutenir qu'ils sont réellement intelligents, obtiennent d'assez bons scores à certains tests. En 2003, Pritika Sanghi et David Dowe (Pritika Sanghi, David Dowe, A computer program capable of passing IQ tests. *Proc. 4th ICCS International Conference on Cognitive Science*, 2003) ont par exemple écrit un programme qui obtient un score proche du score humain moyen dans une grande variété de tests du type suivant :

Considérer la suite



Quel dessin parmi ces quatre dessins poursuit la suite



Compléter le tableau

2	4	8
3	6	12
4	8	?

Ce programme n'est pas très compliqué et se base sur quelques principes simples fréquemment utilisés par des créateurs des tests. En perfectionnant ce programme, il atteindrait probablement les scores des meilleurs humains. Cela prouve non pas que les programmes sont intelligents, mais plutôt que ces tests sont insuffisants pour caractériser et mesurer l'*intelligence générale*.

Une catégorie particulière de tests est d'ailleurs bien mieux réussie par la machine que par l'homme, et vous pouvez vous-même en faire l'expérience. Considérons la suite de nombres entiers : 3, 4, 6, 8, 12, 14, 18, 20, 24, 30, 32, 38, 42

La question est : quels sont les trois nombres venant logiquement derrière ?

La réponse est assez facile pour cet exemple, et vous aurez sans doute reconnu la suite des nombres premiers augmentés d'une unité chacun,  $2+1$ ,  $3+1$ ,  $5+1$ ,  $7+1$ ,  $11+1$ ,  $13+1$ ,  $17+1$ , etc. Les nombres qui suivent sont donc :  $43+1$ ,  $47+1$ ,  $53+1$ , soit 44, 48, 54.

Il existe un site internet qui répond parfaitement à ce type de tests. Il s'agit du site <https://oeis.org/> de l'« Encyclopédie des suites numériques en ligne » de Neil Sloane. En lui soumettant la question, il trouve instantanément la bonne réponse. Il propose d'ailleurs d'autres réponses que vous n'auriez pas imaginées. Celle qu'il indique en second est : nombres  $n$  tels que pour tout entier  $k$  premier avec  $n$  et plus grand que  $k^2$ ,

le nombre  $n - k^2$  est premier. Cette seconde réponse logique conduirait à proposer les trois entiers : 48, 54, 60. Elle est beaucoup plus compliquée que la première et ne serait bien évidemment pas la réponse que donnerait un système programmé à partir de l'Encyclopédie à qui on demanderait de fournir une seule réponse.

Ce que réalise ce système informatique est bien supérieur à tout ce que peut espérer obtenir un être humain. Si vous en doutez, essayez de résoudre les cinq problèmes suivants :

A : 11, 12, 14, 16, 20, 21, 23, 25, 29

B : 11, 31, 71, 91, 32, 92, 13, 73

C : 3, 7, 14, 23, 36, 49

D : 1, 2, 4, 5, 10, 20, 29, 58, 116

E : 1, 4, 5, 7, 8, 11, 13, 14, 16, 22, 25, 28, 31, 34

### Réponses

A : nombres à partir de 10 dont la somme des chiffres est premier (30, 32, 34).

B : nombres premiers écrits à l'envers à partir de 10 (14, 34, 74).

C : le nombre en position  $n$  est la somme de  $n^2$  et du  $n$ -ième nombre premier (66, 83, 104).

D : diviseurs de 580 (145, 290, 580).

E : nombres qui, précédés de 2 et suivis de 1, donnent des nombres premiers (35, 37, 38).

Toutes ces réponses sont trouvées instantanément par le programme de l'Encyclopédie. Il n'est cependant pas raisonnable de considérer que le programme qui produit ces réponses est en quoi que ce soit intelligent. Il se contente d'aller rechercher la suite qu'on lui propose dans une base de données (soigneusement complétée depuis des années), et il rend comme réponse la plus simple des réponses (la notion de simplicité étant déterminée par le contenu de l'encyclopédie qui classe les suites). Le succès du programme provient de sa mémoire et de sa capacité à la parcourir rapidement sans erreur, alors qu'un humain, face à ce type de problèmes, mène des calculs de tête et tente directement de repérer des structures. Comme c'est le cas pour de nombreuses applications d'*intelligence artificielle* (voir l'article), la réussite – extraordinaire ici – du logiciel est obtenue par une méthode sans aucun rapport avec celle que met en œuvre une intelligence humaine.

## 2. Watson est-il intelligent ?

Récemment on a réussi à programmer un système informatique de telle façon qu'il joue et gagne à un jeu demandant apparemment de maîtriser un grand nombre de connaissances de culture générale et, en même temps, de comprendre le langage naturel (l'anglais).

Le jeu est *Jeopardy !*, très populaire aux États-Unis, il existe depuis 1962 et ressemble au jeu *Questions pour un champion* de la chaîne France 3 de la télévision française. Les questions portent sur la géographie, la littérature, les arts, le sport et les sciences. Elles sont formulées en anglais courant et les réponses doivent aussi être formulées en anglais. La machine reçoit les questions par écrit et donne ses réponses par synthèse vocale. La rapidité des réponses intervient dans certaines phases du jeu. Le programme *Watson* créé par la firme IBM a été confronté à deux champions humains du jeu en février 2011. Le programme l'emporta, montrant qu'un système informatique convenablement programmé pouvait s'attaquer à des défis comportant à la fois un traitement du langage naturel et la maîtrise de connaissances sur des sujets variés et étendus. Aujourd'hui le programme *Watson* sert à développer des outils d'expertise en médecine et il est utilisé pour la formation des médecins. Voir : Ferrucci D., et al. *Watson : Beyond Jeopardy ! Artificial Intelligence*, 199, 93-105, 2013.

Le programme *Watson* est d'une très grande complexité et surtout il fonctionne avec une puissante machine dont nous ne sommes pas près de disposer comme ordinateur personnel. La configuration matérielle est en effet composée de 90 machines de base ayant chacune 4 processeurs puissants disposant chacun de huit cœurs. Cela fait un total de 2 880 cœurs. Cette puissance est nécessaire pour répondre rapidement. D'après ses concepteurs, *Watson* est capable de parcourir presque instantanément l'équivalent de 200 millions de pages à la recherche d'informations utiles pour répondre à une énigme. Ces pages occupent 4 téra-octet ( $4.10^{12}$  octets) de disque et incluent le contenu complet de Wikipedia, ainsi que bien d'autres dictionnaires et bases de données.

Le principe de travail de *Watson* au cours d'une partie de *Jeopardy !* est d'analyser les termes apparaissant dans l'énoncé des questions et de partir à la recherche de termes liés. Une multitude de modules l'aide à organiser sa recherche et à formuler sa réponse. À nouveau, on doit considérer que la méthode de jeu de *Watson* est très différente de celle d'un humain. Celui-ci n'apprend pas, mot à mot, plusieurs encyclopédies, mais dispose, autant qu'on puisse le savoir, d'un système de classification plus efficace que *Watson* des informations (peu nombreuses comparées celles de la machine) qu'il a en tête. L'*intelligence artificielle* réussit certes à égaler et même à battre des humains, mais elle le fait en remplaçant l'intelligence (liée à l'organisation des informations dans le cerveau) par de la mémorisation massive et du calcul (le parcours rapide de quantités colossales de données). De plus, la machine *Watson*, conçue spécialement pour gagner à *Jeopardy !*, demande de longues et délicates adaptations quand on veut lui faire aborder d'autres tâches. Elle n'est pas capable en particulier de passer le test d'indiscernabilité de Turing.

### 3. Le concours de compression de Marcus Hutter

Face à une série de données, être intelligent permet d'en prévoir la suite mieux que ne le ferait le hasard. On montre que cette capacité de prévision est équivalente à la capacité de compresser les données reçues. À condition d'imaginer des jeux de données très variés, cette conception de l'intelligence est considérée par certains comme absolue, et c'est elle qui sert d'ailleurs de base à la *théorie générale de l'intelligence* de Marcus Hutter.

Ce dernier, pour illustrer cette idée et parce qu'il est convaincu que la compression de données est nécessaire et suffisante pour caractériser et mesurer l'intelligence d'un être animal, humain ou mécanique, a créé un concours fondé sur une épreuve de compression de données. Ce concours est ouvert à tous et permet de gagner tout ou partie d'une somme de 50 000 euros que Marcus Hutter a lui-même engagés.

Le prix (dénommé Hutter prize : <http://prize.hutter1.net/>) propose une épreuve unique de compression d'un texte numérique énorme. Un fichier de 100 millions de caractères extrait de l'encyclopédie Wikipedia est proposé aux candidats et doit être compressé au mieux par les programmes qui sont soumis. Au départ, le fichier, grâce à une méthode de compression classique, a été réduit à 18 324 887 caractères. Cela correspond à un gain d'environ 81 %. Chaque candidat, proposant un programme réalisant une amélioration de  $n$  % par rapport au gagnant précédent, remporte la somme de  $500n$  euros. Si par exemple vous concevez et programmez un compresseur qui fait gagner 5 % par rapport au dernier gagnant, vous emportez 2 500 euros. La compression dont il s'agit ici est bien sûr une compression sans perte : à partir du fichier compressé, le programme de décompression (associé au programme de compression) doit reconstituer exactement les 100 millions de caractères du fichier initial proposé par Hutter. Des limites concernant la taille du programme de compression et son temps de calcul sont imposées. Voir les détails des règles du concours sur la page internet décrivant le prix.

Depuis le lancement du concours, plusieurs programmes ont réussi à améliorer le taux de compression. On en est aujourd'hui à un fichier compressé de 15 949 688 caractères. Claude Shannon, le créateur de la théorie de l'information, a évalué que le langage naturel porte en gros un bit d'information par caractère, ce qui correspond dans notre cas à un fichier compressé de 12 500 000 caractères. Cela signifie qu'il y a encore de la marge, surtout qu'il se pourrait bien que l'exploitation fine des régularités de notre monde qui se reflète dans l'encyclopédie Wikipedia permette une compression sensiblement meilleure que celle évaluée par Shannon.

### Bibliographie

Goertzel B., Artificial General Intelligence : Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence* 5.1 : 1-46, 2014 :

Hutter M., 50 000 euros prize for compressing human knowledge : <http://prize.hutter1.net/>

Tesauro G., et al. Analysis of Watson's Strategies for Playing Jeopardy ! arXiv : 1402.0571, 2014 : <http://arxiv.org/pdf/1402.0571.pdf>

Dowe D., Hernández-Orallo J., « How universal can an intelligence test be », *Adaptive Behavior*, 22:1, 51–69, 2014.

Wikipedia, Voiture sans conducteur de Google : [http://en.wikipedia.org/wiki/Google\\_driverless\\_car](http://en.wikipedia.org/wiki/Google_driverless_car)

Hernández-Orallo J. et al. Project « Anytime Universal Intelligence », 2009.

<http://users.dsic.upv.es/proy/anynt/>

Hutter M., *Universal Artificial Intelligence : Sequential Decisions Based on Algorithmic Probability*, Springer, 2004.



## Échange de vues

**Joseph Thouvenel :** Une remarque, il me semble qu'une voiture autonome a fait sa première victime il y a quelques mois, en écrasant une brave femme qui traversait la rue, et une question, peut-on parler d'intelligence et la notion d'intelligence a-t-elle un sens si on ne prend pas en compte la sensibilité ?

**Jean-Paul Delahaye :** Pour l'accident mortel, en fait ce n'est pas le premier mais le troisième impliquant des voitures autonomes en période de test. Pour la question, je la tournerais autrement : est-il inconcevable qu'une machine possède une forme de sensibilité ? On ne doit pas partir du préjugé que les machines ne sauront jamais rire, n'auront jamais d'humour, ou ne seront jamais consciente d'elles-mêmes — c'est souvent un argument avancé dans les débats sur l'IA. Il faut être prudent et ne pas décréter impossible ceci ou cela sans vraiment de bonnes raisons. Aujourd'hui, on a su faire des machines qui produisent des œuvres d'art. On peut dire que ce qu'elles produisent est assez médiocre, cependant on n'arrive pas forcément à distinguer une image produite par un bon algorithme d'une image produite par un humain, et c'est encore plus vrai pour la musique. Pour être intelligent au sens plein, il faut peut-être une forme de sensibilité, mais la question à ce moment-là devient : est-il vraiment impossible de doter une machine d'une forme de sensibilité ? Ma position sur ces questions, est qu'aujourd'hui en matière d'IA, il y a toutes sortes de choses qu'on ne sait pas faire, mais que je ne vois pas d'obstacles fondamentaux à ce que, par exemple, on dote une machine de réelle sensibilité, ou d'humour, ou même d'une conscience d'elle-même. On ne sait pas le faire, mais pourquoi décréter que c'est impossible ?

**Rémi Sentis :** Je voudrais revenir sur la notion d'autonomie. À un moment dans votre exposé, vous avez dit que les machines ont été programmées pour faire ceci ou pour avoir telle action ; et vous avez bien insisté sur le fait qu'il y a des programmeurs qui écrivent des programmes, des algorithmes (en général ces algorithmes sont fondés sur des éléments mathématiques assez profonds). Donc il y a tout un travail des programmeurs pour programmer la machine. En fait, toutes les machines ont été programmées par des hommes, (ou éventuellement par une machine qui elle-même a été programmée par des hommes), donc il y a toujours un programmeur ou une équipe qui est derrière les actions « décidées » par la machine.

Ainsi, l'autonomie d'une machine - me semble-t-il – est une notion floue et assez commerciale (idéale pour vendre des voitures). Mais cette autonomie n'est pas absolue, elle sera toujours subordonnée à des hommes ; il y aura toujours un homme qui sera à l'origine des actions « décidées » par les machines.

**Jean-Paul Delahaye :** C'est ce que Alan Turing a appelé l'objection de Lady Ada Lovelace, qui était une mathématicienne au XIX<sup>e</sup> siècle. Elle a contribué à écrire des programmes pour la machine de Charles Babbage, une machine mécanique à engrenages. L'objection formulée par Lady Ada Lovelace est que la machine ne fait que ce qu'on lui a dit de faire et qu'elle ne sera donc jamais vraiment intelligente. Oui, elle ne fait que ce qu'on lui a dit de faire, mais il faut toujours garder à l'esprit qu'une machine peut être conçue et programmée pour avoir de l'autonomie et que cette autonomie n'est pas limitée. L'exemple le plus simple qui permet de comprendre ce que cela signifie est lié au jeu d'échecs : ceux qui programment ces jeux informatiques d'échecs se font battre par leurs programmes. Cela veut dire qu'une fois les programmes écrits et mis dans la machine, et une fois mise en fonctionnement, elle travaille de manière autonome sans même que vous compreniez le détail de ce qu'elle fait et pourquoi. Elle joue où se comporte comme un être détaché de ceux qui lui ont donné naissance et qui leur échappe... et les bat. On lui a appris à jouer aux échecs, et maintenant elle est meilleure que ceux qui l'ont conçue et mis en marche. Affirmer, ce qui est vrai, que la machine ne fait qu'obéir n'interdit pas que les machines deviennent autonomes et très compétentes pour toutes sortes d'activités : se déplacer sur certaines routes, dialoguer, produire de la musique, etc. L'objection de Lady Ada Lovelace finalement n'est pas recevable et ne constitue pas un argument permettant d'affirmer qu'il n'y aura jamais de machines autonomes et intelligentes égalant les humains pour une multitude de choses.

**Rémi Sentis :** Oui, mais la machine ne sera pas autonome totalement...

**Jean-Paul Delahaye :** Pourquoi ? Aujourd'hui, on sait faire des véhicules autonomes qui ont une véritable capacité à réagir et à s'adapter à leur environnement quand on les laisse se débrouiller seules. Il faut

pour cela que les environnements ne soient pas trop complexes, et par exemple qu'il n'y ait pas d'autres véhicules dans les espaces où se font les déplacements. Cependant, dès maintenant, il existe des véhicules autonomes et une fois lancés, ils sont autonomes comme vous et moi quand nous nous déplaçons dans la rue.

**Patrick de Saint-Louvent :** On évoque actuellement la maladie d'Asperger, pour les capacités intellectuelles supérieures des personnes atteintes de cette forme d'autisme. L'homme n'utiliserait qu'une petite partie de son intelligence, 10 % avec d'immenses réserves. Ces personnes à Haut Potentiel Intellectuel n'ouvrent-elles pas des perspectives de recherche au moins autant que les machines ? Le mathématicien a-t-il sa place dans cette recherche ?

**Jean-Paul Delahaye :** Concernant les autistes qui ont des capacités mathématiques exceptionnelles de calcul ou de mémorisation, il me semble que ce ne sont pas des gens possédant une intelligence générale, en tout point supérieure à celle d'un individu comme vous ou moi. Ils ont un cerveau spécialisé, qui s'est concentré sur certaines tâches, qui réussit beaucoup mieux que nous, voire qui réussit mieux que nous pourrions réussir même en y travaillant avec détermination. Certaines personnes réussissent à mémoriser 100 000 décimales du nombre  $\pi$ . C'est le record. Pour y arriver il leur faut des années d'entraînement quotidien. Je ne crois pas cependant que nous n'utilisons que 10 % de notre cerveau et qu'eux utilisent pleinement un potentiel que nous ne savons pas mettre en œuvre. Pourquoi la nature aurait-elle mis en nous ces 90% que nous n'utilisons pas ? Les exploits des autistes ne semblent pas utiles pour augmenter l'intelligence humaine. Ce qu'ils font ressemble plutôt à des numéros de cirque : fascinants, mais inutiles. Maintenant, est-ce que ce type d'intelligences particulières pourrait être utilisé pour concevoir des machines ? Je ne le crois pas. J'ai insisté dans mon exposé sur le fait qu'aujourd'hui les chercheurs en intelligence artificielle, n'utilisent que très peu ou pas du tout les connaissances dont on dispose sur le fonctionnement du cerveau humain. On a parfois évoqué l'idée qu'il faudrait comprendre comment fonctionne le cerveau humain pour faire des machines intelligentes, mais aujourd'hui ce n'est pas du tout la méthode qu'on utilise ! On programme des réseaux de neurones, mais quand on parle de neurones, il faut bien comprendre que ce sont des *neurones formels*. Ils ne sont pas organisés comme sont organisés les neurones de notre cerveau. Donc aujourd'hui, même quand on utilise des réseaux de neurones, on n'imité pas ce qu'on sait du fonctionnement du cerveau humain. D'ailleurs, on n'en sait pas suffisamment pour qu'on puisse en tirer quelque chose de directement utilisable. Les réseaux de neurones ne doivent pas faire illusion. Ce ne sont pas des neurones comme ceux de notre cerveau. L'idée de ces neurones formels provient effectivement de ce qu'on sait du cerveau humain, mais la façon dont on les utilise aujourd'hui ne consiste pas à les faire fonctionner comme nos vrais neurones dont on ne sait pas avec assez de précision comment ils apprennent, se souviennent et raisonnent.

**Marie-Joëlle Guillaume :** Et pourquoi reprendre le mot alors ?

**Jean-Paul Delahaye :** Parce que quand on parle de *neurone formel*, l'idée est quand même un peu inspirée des vrais neurones. Seulement on les mathématise pour les mettre dans nos machines. Ce sont des réseaux bien nets, bien propres, bien définis et régulièrement connectés. Parfois on les fabrique physiquement, mais c'est assez rare. Le plus souvent, on les fait fonctionner virtuellement, sous la forme de logiciels. Ces différents neurones formels fonctionnent en général selon un principe de seuil. Ils sont connectés à d'autres neurones. Un neurone sera excité ou au repos. Quand il est excité, il envoie des signaux à ses voisins. Quand l'un de ses voisins atteint un certain seuil d'excitation, lui-même est excité, et envoie des signaux à ses voisins. On met au point des méthodes pour que ces réseaux apprennent quelque chose. Ils apprennent par exemple à reconnaître les diverses lettres manuscrites qu'on leur présente. On leur montre un 'A', on indique au réseau qu'il doit reconnaître un 'A', il ajuste ses paramètres pour prendre en compte cette donnée. On recommence cet apprentissage un grand nombre de fois. Petit à petit, à chaque présentation les paramètres du réseau évoluent et s'ajustent. Quand on a fait ça avec un très grand nombre de lettres et que le réseau se stabilise alors il devient utilisable. On lui présente un 'A' manuscrit et il reconnaît seul que c'est un 'A', même s'il est différent de tous ceux qu'il a vu dans la phase d'apprentissage. L'idée de ces méthodes est effectivement venue de ce qu'on imagine être le fonctionnement d'un neurone vivant, mais on en a fait un objet mathématique. Les méthodes d'apprentissage de ces réseaux de neurones sont des méthodes mathématiques dont il n'est pas du tout certain qu'il imite ce qui se passe dans notre cerveau. Il y a donc bien une idée provenant des neurones humains, vivants, mais au total la façon dont ces réseaux apprennent et fonctionnent est une invention mathématique. Je crois tout simplement que les vrais cerveaux humains

apprennent par des processus bien plus complexes et variés que ceux qu'on programme avec nos « petits » neurones formels.

**Jean-Luc Bour :** Vous avez fait remarqué que le grand progrès provient de l'idée d'apprentissage. Vous avez également insisté en donnant un exemple sur la difficulté à trouver à qui, dans une phrase, se rapporte un pronom « il » ? En fait, c'est exactement les problèmes que nous avons à traiter à l'école élémentaire. Je me souviendrai toujours une dictée : *La diligence roule sur l'étroit chemin*, et comme j'avais appris ce qu'était que le chiffre trois, mais pas le mot « étroit », j'avais écrit « trois », sans comprendre. Donc pourquoi est-ce que tous ceux qui travaillent en intelligence artificielle ne se penchent pas davantage sur la manière dont un enfant avance dans son éducation et dans sa compréhension ?

**Jean-Paul Delahaye :** L'idée des méthodes d'apprentissage, est très ancienne. Les premiers programmes d'apprentissage concernaient le jeu de dames. On utilisait un système de récompense : quand le programme avait fait les bons choix et gagnait le programme était récompensé et il avait une punition « numérique » quand le choix était mauvais et le faisait perdre —je simplifie. On faisait jouer un programme contre lui-même. Petit à petit, il progressait. Mais cette progression n'a pas été très loin, elle a rapidement cessé. Les meilleurs programmes de jeux de dames ne sont plus obtenus comme cela. L'idée de faire apprendre aux programmes est étudiée sous une multitude de formes et aujourd'hui encore on en invente de nouvelles. C'est un domaine important de la recherche en intelligence artificielle qui occupe des centaines de chercheurs dans le monde. Il faut cependant préciser que toutes ces méthodes, aujourd'hui ne permettent pas de faire l'équivalent d'un cerveau d'enfant qui apprend tout seul comme un enfant apprend à force de dialogue, d'images, de lectures, de conseils, etc. Les méthodes d'apprentissage qu'on est capable aujourd'hui de concevoir et de programmer n'égalent pas ce que fait un enfant. Peut-être qu'un jour on saura le faire, c'est l'idée qu'on a parfois dénommée « œuf logiciel ». On créerait une structure minimale capable de recevoir et de traiter de l'information ; ensuite elle serait "alimentée" par toutes sortes de données venant du monde réel : elle serait nourrie d'images, de textes, de sons, etc., Petit à petit cet œuf logiciel deviendrait plus intelligent, assimilerait des connaissances, etc. C'est un rêve ! Aujourd'hui, on ne sait pas faire un tel œuf logiciel. Sinon on aurait résolu le problème du test de Turing. Cette voie a été explorée en particulier par Jacques Pitrat, un chercheur français pionnier de l'IA, mais il n'a pas réussi à atteindre ce but bien plus difficile qu'on ne l'imagine.

**Nicolas Aumonier :** Ce qui est intéressant dans la manière dont vous avez présenté les choses, c'est que cela a un côté rassurant. Le système médiatique dans lequel nous vivons semble suggérer que nous sommes presque foutus face aux machines ! Vous avez battu cela en brèche. Cependant, il existe un petit robot humanoïde nommé Nao qui est assez troublant. On le branche, il peut dire « bonjour », « au revoir », « est-ce que vous avez bien pensé à prendre vos pilules ? », et tout un tas de choses. Même si le partenaire de ce petit Nao lui prête spontanément des intentions, il n'y a évidemment rien de tel. Le mot qui m'intéresse est le mot « émergence ». Qu'est-ce qui permettra qu'on assiste à l'apparition d'une propriété émergente ?

**Jean-Paul Delahaye :** Je le répète, faire « émerger » une intelligence à partir d'un programme qui serait confronté à des sources d'informations multiples est une idée séduisante, mais qu'on ne sait pas faire fonctionner aujourd'hui, sauf pour des problèmes limités et certainement pas pour arriver à quelque chose qui ressemblerait à une intelligence humaine et au développement d'un enfant. Nao, ne peut apprendre de choses que dans un domaine très limité de la conversation, il ne deviendra pas intelligent même si on interagit avec lui pendant des siècles.

**Nicolas Aumonier :** Les gens se rassurent à bon compte, et notamment dans le domaine du diagnostic. Je me demande cependant : est-ce qu'il est plus rassurant d'être diagnostiqué et traité par un algorithme qui fait bien le boulot, n'oublie rien, produit un conclusion assez bonne, plutôt que par quelqu'un de fatigué, de pressé, qui n'a pas tout en tête, et qui va finalement faire une magistrale erreur médicale ?

**Jean-Paul Delahaye :** C'est une très bonne question. Concernant le traitement par l'intelligence artificielle des problèmes médicaux, je vais mentionner deux situations et les opposer. On a conçu des programmes d'analyse images qui savent reconnaître mieux les mélanomes que ne le fait un médecin dermatologue. On comprend bien pourquoi. C'est un exemple où les réseaux de neurones et leurs systèmes d'apprentissage fonctionnent à merveille. On apprend à ce réseau de neurones en lui donnant la bonne réponse pour de centaines d'images de boutons sur la peau qui sont des mélanomes dangereux ou bénins.

Grâce à cet apprentissage le réseau de neurones est devenu capable de faire aussi bien que ce que fait un spécialiste humain. Paraît-il. C'est tout à fait concevable et certains articles l'affirment. Personnellement, je veux bien me faire examiner par un tel système. Cependant, il ne faut pas grossir trop un tel succès. C'est très limité et on n'a pas un système général de diagnostic, ni même un système général de diagnostic dermatologique. Dans un autre domaine, on continue à faire des recherches qui posent eux un autre type de questions. Il s'agit des systèmes experts médicaux. On leur donne un certain nombre d'informations et bien programmés par de bons médecins spécialistes qui leur confient leurs expertises, ils fournissent des diagnostics satisfaisants pour traiter certaines catégories de problèmes. On peut aussi utiliser cette technologie pour du conseil en investissement boursier, ou pour de l'aide à la réparation des machines en panne. En médecine le système reçoit un certain nombre d'informations sur un patient, le système parfois pose lui-même des questions ajustées à ce qu'il sait déjà, et puis au bout d'un certain temps, le système propose un diagnostic, voire une prescription, par exemple un certain antibiotique précis. Il y a cependant à ce sujet une question qui va rejoindre la vôtre et qui m'intrigue depuis longtemps. Ces systèmes existent depuis plus de 20 ans. On en fait des très bons en particulier un système qui s'appelait MYCIN. Il était spécialisé dans le diagnostic des maladies infectieuses et la prescription d'antibiotiques. Il faisait aussi bien que les meilleurs experts. Et pourtant on ne l'a pratiquement pas utilisé. La raison ? Je crois qu'elle est simple : on n'est pas prêt à tolérer qu'une machine se trompe. Alors que pour un médecin humain, on accepte qu'il se trompe, qu'il ne prescrive pas le meilleur antibiotique vu les symptômes décrits. C'est normal, ce n'est pas grave. En revanche, si un système automatique commet une erreur, et que la personne meurt, on ne l'acceptera pas. Tout à l'heure, on évoquait le cas de la voiture autonome qui a percuté une passante et l'a tuée. Il y a eu trois cas de décès liés aux voitures autonomes. Ces cas ont toujours été considérés comme des drames et ont conduit à l'arrêt de plusieurs expérimentations de voitures autonomes. C'est un peu la même situation aujourd'hui, tous les jours, il y a centaine de personnes qui meurent dans des accidents d'auto causés par des êtres humains et cela est accepté. Personne ne trouve que c'est un scandale. Mais, si c'est une voiture qu'on essaie de rendre autonome qui provoque un accident mortel, on considère alors que c'est grave. Cela me fait d'ailleurs penser que même quand on aura avancé dans la conception des voitures autonomes, il sera très difficile de convaincre les autorités politiques d'autoriser ces véhicules. Pour les systèmes experts médicaux, comme pour les voitures autonomes, il y a là une difficulté qui freinera l'utilisation des systèmes d'IA. Il est difficile d'anticiper ce qui se passera.

**Nicolas Aumonier :** J'ai été très étonné de vous entendre dire qu'une intelligence artificielle ne peut pas produire une bonne analyse grammaticale. Pourquoi ? La langue, est quand même très logique !

**Jean-Paul Delahaye :** Je me suis peut-être mal exprimé, je parlais de l'analyse grammaticale dans des situations complexes. Reprenons un exemple utilisé par Jacques Pitrat pour faire prendre conscience de la difficulté de la traduction automatique. Il s'agit d'un début de phrase suivi de trois possibilités.

*Le professeur envoya le cancre chez le proviseur*

(a) *parce qu'il lançait des boulettes, ou*

(b) *parce qu'il voulait avoir la paix, ou*

(c) *parce qu'il l'avait convoqué.*

Pour savoir que le « *il* », est le cancre (cas a), ou le professeur (cas b) ou le proviseur (cas c) il faut comprendre la situation. C'est immédiat pour nous, mais interrogez-vous sur ce que nécessite comme connaissances cette analyse dite « grammaticale ». L'analyse n'est possible que si on dispose d'informations qui justement ne sont pas de nature grammaticale. L'identification du « *il* » exige de comprendre la scène, et d'avoir en tête un certain nombre de schémas décrivant des situations de la vie courante : dans une classe, il est habituel que le cancre lance des boulettes, que le professeur se fâche, que le proviseur convoque les cancre etc. Il faut mettre ces schémas dans la machine pour qu'elle puisse identifier correctement ce qui se rapporte aux différents « *il* ». C'est très difficile.

**Nicolas Aumonier :** Est-ce que vous êtes favorable ou défavorable au fait que les robots aient une personnalité juridique ?

**Jean-Paul Delahaye :** Aujourd'hui, je suis défavorable, parce qu'ils n'ont pas un niveau d'intelligence tel qu'on puisse considérer raisonnablement que ce sont des personnes méritant de se voir attribuer une personnalité juridique. Je précise, et je sais que je vais choquer certaines personnes : je crois à l'intelligence artificielle : je crois qu'on arrivera à faire des systèmes de plus en plus perfectionnés. Peut-être en copiant l'intelligence humaine, peut-être pas. Je crois qu'à force de progrès, on créera des programmes qui non

seulement joueront bien aux échecs et conduiront des voitures aussi bien que les être humains, qui sauront commenter un poème, etc. Quand on saura faire des machines de ce type, peut-être dans dix ans, peut-être dans cinquante ans ou plus, je crois qu'alors nous devons les respecter et considérer qu'elles ont une forme de dignité, et donc leur attribuer certains droits. On devra par exemple considérer comme un acte répréhensible de les détruire. À ce moment-là, il ne sera pas absurde de leur attribuer une personnalité juridique. Aujourd'hui, on n'y est pas du tout, envisager cela c'est faire de la science-fiction comme quand on évoque le téléchargement d'un esprit dans une machine, dont on ne peut même pas formuler la moindre hypothèse concernant une date où ce sera possible, et pour ce problème particulier, j'ajoute, si un jour c'est possible. Certains journalistes semblent croire imminentes des réalisations invraisemblables et en réalité très lointaines, et pour certaines largement au-delà de tout avenir prévisible. Cependant si un jour on y arrive, je ne vois pas pourquoi on ne respecterait pas les entités intelligentes qu'on aura créées et qui seront comme nos enfants.

**Marie-Joëlle Guillaume :** Là, vous venez de tenir des propos qui nous ouvrent des abîmes ! J'aurais tendance à vous répondre que tout de même, l'âme existe, et que même la construction d'une machine ultra perfectionnée n'approchera pas de son mystère... Mais là n'était pas ma question. Elle touche à la place de l'homme. En fonction de tout ce qu'on lit, dans la presse, etc., on a l'impression qu'avec l'intelligence artificielle il y a des pans entiers de l'activité humaine – là, je passe vraiment dans le domaine des applications -, des pans entiers d'activités professionnelles qui seraient *remplacées* par l'intelligence artificielle. Par exemple, j'ai lu récemment un article à propos des professions juridiques. On y expliquait qu'il y a énormément de tâches qui seraient mieux prises en compte par l'intelligence artificielle, capable d'analyser des milliers de documents, que par des notaires, par exemple, forcément limités par le temps et l'étendue de leurs connaissances... Alors je voudrais vous poser cette question pour finir : en fonction de tout ce que vous nous avez dit, est-ce que vous voyez tout de même des limites nettes, et pouvant être posées rapidement, à ce qui serait concédé à l'intelligence artificielle au détriment de l'être humain ? Ou au contraire, avez-vous l'impression que ce périmètre peut effectivement être très important ?

**Jean-Paul Delahaye :** Ma réponse est que sur le très long terme, on peut imaginer tout ce que l'on veut, mais pas dans l'immédiat. Il n'y aura pas rapidement de voitures totalement automatiques circulant dans les rues de Paris et partout sur les routes. Pour moi, il n'y aura pas non plus de machines traitant les problèmes juridiques aussi bien que les spécialistes qui comprennent les dossiers dans le détail. Il n'y aura pas non plus de machine capable de traduire convenablement en anglais un roman de Victor Hugo ou de Marcel Proust. Ces tâches demandent une intelligence générale qu'on ne sait pas « mettre en programmes » !

**Jean-Pierre Brulon :** C'est déjà le cas dans certains cabinets d'avocats. J'ai pu avoir accès à des documents dans les archives, des notes de jurisprudence qui ont été faites par des machines d'une technicité telle que c'est vraiment affolant.

**Jean-Paul Delahaye :** Beaucoup de choses faites par des programmes ne doivent pas être considérées comme de l'intelligence artificielle, même si les programmes concernés sont assez complexes. Aujourd'hui, par exemple, les réseaux informatiques sont gérés par des algorithmes qui déterminent les flux de données et les redirigent, les avions atterrissent grâce à des dispositifs qui prennent de nombreux paramètres en compte et les utilisent au mieux ; le pilote ne touche à rien et l'avion se pose parfaitement. Dans le cas des analyses juridiques, je ne doute pas que des programmes peuvent exécuter certains travaux. Je ne suis pas compétent sur les sujet et je ne connais pas les situations que vous évoquez, mais je reste persuadé que les analyses qui nécessitent du sens commun pour démêler les situations ne peuvent pas être confiées totalement à des systèmes d'IA. Les systèmes d'intelligence artificielle ne comprennent pas assez bien les histoires qui peuvent intervenir dans les dossiers, ni les détails d'un accident et leurs interactions, etc. Des êtres humains doivent valider le travail, et souvent le corriger. En traduction automatique la situation est la même, l'IA réussit partiellement, elle simplifie le travail, mais il est inconcevable de lui faire confiance totalement. Il ne faut pas confondre les applications particulières, qui se situent dans des cadres déterminés (par exemple la traduction de notices techniques), et les travaux nécessitant des analyses fondées sur le sens commun ; c'est le cas de la traduction des romans, c'est le cas de la conversation sur des sujets non étroitement délimités, etc. Une Intelligence artificielle totale — équivalente ou supérieure à celle d'un humain — aujourd'hui n'existe pas, et c'est bien pourquoi passer le test de Turing reste un défi.

*Séance du 15 novembre 2018*