Turing : les machines peuvent-elles penser?

Soumis par Stephane Desbrosses

Les machines peuvent-elles penser ? Si la question fut déjà posée deux millénaires auparavant, la réponse suscite encore des débats au sein de la philosophie et des différents courants de l’Intelligence Artificielle, notamment parce qu’elle se heurte à un autre débat philosophique ancien : comment pouvons-nous nous-même savoir si l'on pense ? A partir de quel moment peux-t-on considérer un phénomène comme relevant de la pensée ?
Pensée, intelligence, simulationIl semble en effet nécessaire, à priori, de savoir définir pensée et intelligence. A priori seulement. En fait, certains tenants de l’intelligence artificielle ont proposé d’évaluer l’intelligence et la capacité d’une machine à penser selon un critère humain : certaines tâches précédemment réalisées strictement par des humains et considérées comme dénotant l’intelligence (jouer aux échec, résoudre un problème mathématique, établir une démonstration…) peuvent désormais être exécutées par des ordinateurs. Deux conclusions peuvent en être tirées : soit il nous faut admettre que la machine possède effectivement un embryon d’intelligence, soit nous devons reconnaître que les capacités que l’on considérait comme reflétant l’intelligence ne sont finalement que des compétences mineures. Mais jusqu’à quand la deuxième conclusion peut elle durer ?

Il s’agit en tout cas de l’un des problèmes majeurs posés par la réflexion sur les performances et les possibilités des machines : serions nous capables d’accepter qu’une machine puisse être considérée comme intelligente ?

D’une part, la machine est basée sur des principes physiques stricts, à la logique implacable, elle est dénuée d’émotions, de désirs, elle suit des comportements prévisibles, programmés, et ne laisse pas la place à l’interprétation ou la sensibilité aux circonstances. C’est du moins ce que nous en pensons habituellement, mais nous verrons que ces préjugés sont caduques. D’autre part, si la machine réussit à simuler un comportement jugé intelligent, alors on a tendance à considérer qu’il s’agit justement d’une preuve que ce comportement automatisable n’est pas "intelligent". Enfin dans le cas de la performance mécanique à une tâche dont on ne peut nier le caractère intelligent, nous estimerons volontiers que la machine résout le problème d’une manière différente de celle de l’homme, et que cette réussite ne constitue donc pas la re-création d’un comportement intelligent, mais simplement une simulation, froide, logique, dure.

Pourtant, ces arguments ne sont pas forcément réalistes : la machine est-elle toujours logique, implacable, dénuée d’émotions ou de désirs ? S’il est vrai que la machine se fonde sur un système physique aux principes stricts, elle n’est pas pour autant insensible à la nuance ou aux circonstances. Certains programmes prennent en compte des valeurs de l’environnement pour décider de leur fonctionnement, ce simple principe permet d’envisager la machine, non plus comme un système isolé commandé par des règles internes strictes, mais comme un système régulé à la fois par ses prédispositions (logiques, mécaniques…) et des données indépendantes tirées de l’environnement. De plus, l’humain évolue également dans un environnement régis par des lois strictes : l’intransigeance des lois électroniques se retrouve dans les lois biochimiques qui sous-tendent le fonctionnement de notre cerveau ou de notre corps. Nous sommes gouvernés et limités par des principes stricts, les lois de la thermodynamique régissent le comportement de notre système nerveux, au niveau élémentaire. Concentrations de matières et déplacements d’influx obéissent rigoureusement à des règles biochimiques stables. Et si l’on se plait à croire que les règles physiques ne font pas de nous des "machines", alors on ne peut se servir de l’argument selon lequel les principes stricts empêchent la flexibilité du raisonnement, la variabilité des comportements. La contradiction supposée entre cette souplesse de l’intelligence et ce substrat rigide gouverné par des lois immuables n’a peut être pas lieu d’être.

Il n y a alors aucun obstacle théorique à la construction d’une intelligence sur la base de mécanismes figés, par exemple, électroniques. Les tenants de cette école de pensée, dite "maximaliste", ne voient que des raisons pratiques qui puissent empêcher la machine de simuler la pensée humaine. La complexité d’une telle entreprise force pour le moment les ingénieurs de l’IA à se focaliser sur la simulation de comportements relativement simples, dénotant une certaine dose d’intelligence, de jugement ou de décision. C’est également sur ces tâches que l’autre courant de l’IA fonde sa conception de l’intelligence d’une machine : pour lui, on peut certainement recréer par simulation des comportements nécessitant de l’intelligence, mais on ne pourrait envisager la possibilité de sentiments ou de pensée à l’intérieur d’une machine. Il semble qu’on ne saurait vraisemblablement admettre l’intelligence d’une machine que le jour ou elle égalera l’homme ?
Le test de TuringC’est sur ce facteur humain que Turing base en 1950(1) son test devenu célèbre, le Test de Turing, une épreuve issue d’une question déjà très présente à l’époque : Les machines peuvent-elles penser ? Turing considérait cette question mal posée et ambiguë : comment pouvons-nous nous-même savoir si l'on pense, qu’est-ce que penser ? Il tente alors de se poser une question plus concrète, qui ne nécessiterait pas d’avoir à définir intelligence ou pensée, mais dont la réponse suffirait à qualifier aux yeux de tous, une machine comme entité disposant d’une forme d’intelligence : une machine peut-elle se faire passer pour un humain ? En substance, cela revient à se demander si les machines sont capables de faire ce que les hommes, en tant qu’entités pensantes, sont capables de faire. Si l’homme ne parvient pas à distinguer une machine

d’un humain lorsqu’il leur parle, par exemple, il est alors raisonnable d’attribuer à cette machine la capacité à simuler l’intelligence, et implicitement, la pensée.

Turing propose donc comme réponse le test suivant : un homme (le juge) discute, par clavier interposé et sans les voir, avec deux autres "personnes", l'un étant humain et l'autre une machine, lesquelles ont pour but de se faire passer pour un humain. Si l'interlocuteur n'arrive pas à déterminer qui est la machine (et par conséquent qui est l'humain) à travers les conversations qu'il entretient avec les deux "personnes", alors on pourra dire de la machine qu'elle simule correctement l’intelligence et est capable d'égaler certaines performances des humains.

Depuis 1991, un tournoi de bots parlants, le Prix Loebner(2), est organisé chaque année afin de tester sur le principe du test de Turing, les meilleurs programmes de tchat. Turing estimait qu’en 2000, l’avancée de l’informatique permettrait à un ordinateur de bluffer environ 30% des juges lors d’une conversation de 5 minutes. Pour le moment, aucun programme n’a passé le test avec succès, cependant, en 2008, le programme Elbot réussissait à convaincre 25% des juges (3 sur 12). Si nous sommes encore loin de la simulation décrite par Turing, il faut reconnaître que de nombreux bots en activité réussissent à duper l’humain régulièrement, pour peu que celui-ci pense dès le départ, converser avec un humain.
Limites et objectionsLa Pièce ChinoiseDans son article(1), Turing proposait également des objections à son approche, pressentant les critiques qu’elle ne manquerait pas de susciter, afin de les discuter. C’est cependant du philosophe John Searle(3) que provient la critique considérée comme la plus pertinente. Celui-ci propose une expérience de pensée connue sous le nom de la Pièce chinoise. Imaginez vous dans une pièce contenant des symboles Chinois, disposant uniquement d’un mode d’emploi vous permettant, à partir d’une entrée formée de ces symboles, de répondre la séquence de symboles adéquats. Sans en saisir le sens, vous pourriez ainsi dialoguer avec un interlocuteur et passer le test de Turing « chinois » en vous faisant passer pour un chinois. Aucune compréhension n’est alors nécessaire, seules des règles d’associations entre l’entrée et la sortie, suffisent. Searle expose à la suite de cette réflexion, une série d’axiome et de conclusions qui semblent démontrer que le programme informatique sera à jamais condamné à la non conscience, et l’absence d’intelligence. Parmi ces conclusions, celle selon laquelle l’ordinateur manipule des symboles sans en comprendre le sens, jouant des règles de syntaxe sans jamais pouvoir leur attribuer une signification (sémantique), des relations avec l’environnement. Egalement, une simulation ne saurait être qu’une intelligence artificielle sans impact sur la réalité, au même titre qu’une simulation d’un moteur ne saurait faire avancer une voiture. La première conclusion est critiquée par Douglas Hofstadter(4), pour qui ce que l’on nomme la sémantique est relativement illusoire :
Si un programme simulait un cerveau comprenant le chinois, il comprendrait le chinois. Supposez que l’on simule un cerveau jusqu’au niveau cellulaire, ce système comprendrait le chinois tout autant qu’un chinois.
Cette position rejette la sémantique comme caractéristique de l’entité intelligente, cette sémantique ne serait qu’une facette de l’âme chère aux dualistes, abstraite et commode pour refuser l’intelligence aux machines manipulant les symboles. Or, il n’existe encore aucune théorie convaincante de la compréhension et rien ne permet de rejeter l’hypothèse selon laquelle des effets émergents de processus simples permettraient de rendre compte de cette compréhension. On associe généralement à la sémantique à des capacités telles que la catégorisation ou le prototypage, mais certains exemples d’architecture neuronale se prêtent tout à fait à la simulation de ce type de caractéristique.

Quant au fait que l’intelligence artificielle soit une intelligence simulée, cela ne la rend pas inexistante et sans impact sur la réalité. Le domaine de la robotique, notamment, offre de nombreux arguments en faveur de la réalité physique des simulations. Dès les années 50, de petits robots commandés par des programmes simples étaient capables de comportements étonnamment ressemblants à ce que l’on pourrait qualifier de comportements intelligents. L’argument de Searle a néanmoins mené à une conception de la simulation s’appuyant sur la similarité de processus sensitifs et moteurs avec ceux des êtres dotés d’intelligence. Cette conception postule que toute tentative efficace de simulation de l’intelligence animale, notamment humaine, doit s’accompagner de substrats sensitifs et moteurs (i-e, un système auditif, un système visuel, un système moteur, etc…) proches de ceux qui accompagnent l’intelligence que l’on souhaite simuler.
Les compétences des jugesLes juges ne sont pas définis, ni leur capacités, notamment de jugement. Lorsque Turing évoque ces fameux juges, il les présente sous le nom de juges « moyens » (average interrogators). Cela suppose que le juge moyen puisse être capable de différencier l’homme de l’ordinateur, or les juges à qui l’on confie cette tache ont généralement des facultés particulières (psychologues, spécialistes de l’informatique théorique, etc…) tandis que des programmes déjà expérimentés comme les programmes ELIZA et ALICE dupent régulièrement des internautes, lorsqu’ils n’ont aucune raison à priori, de penser qu’ils conversent avec un bot. Dans un sens, donc, de tels programmes ont dors et déjà passé le test de Turing sur des publics non avertis. Certains ont même réussi à bluffer des juges réputés particulièrement compétents en la matière(5), mais le fait de savoir ou non qu’un ordinateur est susceptible d’être son interlocuteur, change notablement les résultats. Il faut cependant tenir compte d’une tendance naturelle des hommes à personnaliser les objets qui l’entourent (biais anthropomorphique). Un juge compétent doit être capable de s’abstraire au maximum de ce biais.
Pratique et pertinenceLe test est un exemple utile à la philosophie de l’esprit, mais en pratique, il est peu pertinent pour mesurer les comportements supposés intelligents. Les praticiens de l’IA préfèrent tester leurs programmes

sur la tâche qu’ils sont censés effectuer. Par ailleurs, la simulation de l’intelligence présente un intérêt limité, dans le sens où la recréation d’une intelligence humaine n’est que partie mineure de la recherche en IA. Il n y a pas plus d’intérêt à recréer une intelligence humaine que de créer un avion qui bat des ailes pour voler(6) : dans cet exemple, la recherche s’est inspiré du vol des oiseaux sans pour autant nécessiter sa re-création.
Stupidité artificielle et intelligence surhumaineUn autre problème s’ajoute à la mesure de l’intelligence : le test propose d’évaluer un comportement humain comme signe de l’intelligence, or, l’intelligence peut être surhumaine, et bien entendu, tous les comportements humains ne sont pas forcément intelligents…

Le test de Turing affirme implicitement que l’ordinateur doit pouvoir reproduire tous les comportements humains, y compris les erreurs de ceux-ci, comme le mensonge, les erreurs d’orthographe ou de grammaire, et même le comportement consistant à vouloir se faire passer pour un ordinateur, par exemple. Turing indique lui-même que pour mimer l’intelligence humaine, l’ordinateur devra montrer des signes de faillibilité. Le premier gagnant du Loebner Prize obtint sa victoire en partie grâce à sa capacité à commettre des erreurs.

A l’opposé, le test place une barrière à l’importance de l’intelligence évaluée. Un programme qui se montrerait supérieurement intelligent le serait visiblement trop pour qu’un juge ne se rende pas compte de son caractère mécanique, le programme éventuel répondant à cette description, bien qu’intelligent, échouerait donc tout de même au test de Turing.
La course vers l’intelligenceOn réalise alors combien complexe serait l’opération consistant à mimer l’humain, sans pour autant qu’il soit nécessairement complexe de mimer des processus reconnus comme requérrant de l’intelligence. A la question "les machines peuvent elles penser ?", le débat dualisme/matérialisme a toujours cours. Sa solution ne repose vraisemblablement que sur l’hypothétique démonstration selon laquelle un niveau supérieur de symbolisme (notamment sémantique), indépendant du substrat biologique, prendrait naissance dans notre cerveau, notre corps ou notre âme. Cette hypothèse est chaque jour mise à mal par les avancées techniques et théoriques : les comportements émergents de systèmes multi-agents permettent d’envisager, par exemple, la création d’une structure d’organisation supérieure à partir d’éléments simples comme les neurones. La complexification des simulations rend en pratique inexplicable, leur fonctionnement interne, parfois même par leurs propres créateurs. Tandis qu’au bout de la chaîne de la pensée, on explique de mieux en mieux le fonctionnement mental humain, celui des machines se complexifie au point que les hommes s’y trompent et leur prêtent des caractéristiques fondamentalement humaines, comme c’est le cas pour des robots de compagnie. Des systèmes informatiques sont désormais capables d’apprentissage, d’associations, de catégorisation, d’inférence…

Turing estimait qu’en l’an 2000, l’expression « machine pensante » ne nous choquerait plus. Combien de temps encore les préjugés sur les machines nous empêcheront-ils d’envisager la possibilité d’une « machine qui pense » ?
(1) Turing, A.M. (1950). Computing machinery and intelligence . Mind, 59, 433-460.
(2) http://loebner.net/Prizef/2008_Contest/loebner-prize-2008.html
(3) Searle, John (1980), "Minds, Brains and Programs", Behavioral and Brain Sciences 3 (3): 417–457
(4) Hofstadter, D. (1985). Mathemagical Themas. Questing for the essence of mind and pattern. Bantam Books
(5) Shah, H;, Warwick, K., (2009c). Hidden Interlocutor Misidentification in Practical Turing Tests. (A paraitre) Kybernetes Turing Test Special Issue
(6) Russell, S. J., Norvig, P., (2003), Artificial Intelligence: A Modern Approach (2nd ed.), Upper Saddle River, NJ. Prentice Hall