

L'intelligence artificielle et le test de Turing

Par **Jean-Paul DELAHAYE**

Professeur émérite à l'Université Lille 1
Chercheur au Laboratoire d'Informatique Fondamentale de Lille

En conférence le 15 avril

En 1950, Turing proposa un test d'intelligence pour les machines. Il pensait que les progrès de l'informatique permettraient de le satisfaire en cinquante ans. Même si les avancées de l'Intelligence artificielle ont été remarquables, aujourd'hui, aucune machine ne passe le test.

Est-ce qu'une machine avec laquelle nous échangeons des messages écrits sans contrainte de sujet peut tenir son rôle d'une façon telle qu'on ne réussisse pas à savoir si nous dialoguons avec un humain ou un système informatique ? C'est le test proposé par Alan Turing en 1950. Il pensait qu'on saurait construire de tels systèmes à la fin du XX^{ème} siècle. La réalité est qu'en 2014, malgré d'immenses progrès dans le domaine de l'*Intelligence artificielle*, on ne sait pas réaliser de programmes possédant une intelligence générale permettant de passer le test avec succès. Les progrès ont porté sur des capacités spécialisées (jeux d'échecs, traitement de problèmes mathématiques, dialogues en langue naturelle sur des sujets limités, etc.). Récemment, avec l'ordinateur Watson de la firme IBM, un pas a été franchi concernant ce qui était l'obstacle principal : la programmation du *sens commun*.

Le jeu de l'imitation

Définir l'intelligence est difficile et il n'est pas certain qu'on puisse y arriver un jour d'une façon satisfaisante. C'est cette remarque qui poussa le mathématicien britannique Alan Turing, il y a soixante ans, à proposer le *jeu de l'imitation* qui fixait un objectif précis à la science naissante des ordinateurs qu'on n'appelait pas encore *informatique*.

Le *jeu de l'imitation* consiste à mettre au point une machine impossible à distinguer d'un être humain. Précisément, Turing suggérait qu'un juge **J** échange des messages dactylographiés avec, d'une part, un être humain **H** et, d'autre part, une machine **M**, ces messages pouvant porter sur toutes sortes de sujets. Le juge **J** ne sait pas lequel de ses deux interlocuteurs (qu'il connaît sous les noms **A** et **B**) est la machine **M** et lequel est l'humain **H**. Après une série d'échanges, le juge **J** doit deviner qui est la machine et qui est l'être humain.

Turing pensait que si un jour nous réussissons à mettre au point des machines rendant impossible l'identification correcte (c'est-à-dire conduisant le juge **J** à un taux de mauvaise identification de 50 % identique à ce que donnerait une réponse au hasard), alors nous pourrions affirmer avoir conçu

une *machine intelligente* ou – ce que nous considérerons comme équivalent – *une machine qui pense*. Sa procédure aujourd'hui appelée *test de Turing* a donné lieu à de nombreuses discussions intéressantes et est à l'origine d'une série de réalisations informatiques concrètes mises en compétition annuellement lors d'un prix organisé à l'initiative de Hugh Loebner. Les résultats des logiciels participant au jeu de l'imitation progressent chaque année..., mais assez lentement.

Est-ce que le test de Turing caractérise l'intelligence ?

Il faut mesurer à quel point réaliser un programme qui passe le test de Turing est difficile. Mettons-nous dans la peau du juge **J** dialoguant, par le biais d'un terminal d'ordinateur, avec les deux interlocuteurs **A** et **B**.

Une première idée, pour reconnaître la machine, consiste à poser une question du type « quelle est la valeur de 429 à la puissance 3 ? ». Si **A** répond 78953589 au bout d'une seconde et que **B** refuse de répondre, ou attend trois minutes pour proposer un résultat, il ne fera pas de doute que **A** est la machine et **B** l'humain. Cependant, les spécialistes qui conçoivent les programmes pour passer le test de Turing ne sont pas idiots et ils prévoient cette ruse grossière. Leur programme, bien que capable sans mal de mener le calcul de 429^3 en une fraction de seconde, refusera de répondre ou demandera dix minutes avant de fournir un résultat, ou même proposera une réponse erronée. Tout ce qu'un humain ne sait pas très bien faire et qu'un ordinateur réussit sans mal – la mémorisation d'une suite de cent chiffres aléatoires est un autre exemple – sera traité de la même façon : l'ordinateur fera semblant de ne pas réussir. Pour reconnaître l'humain, la méthode du juge doit donc s'appuyer sur des tâches que les humains traitent sans mal et sur lesquelles les ordinateurs buttent.

L'humour est un point d'attaque envisageable : racontez une histoire drôle à vos interlocuteurs **A** et **B** et demandez-leur d'expliquer où il faut rire et pourquoi ? Les questions d'actualités, sur lesquelles nous sommes tous informés, peuvent aussi servir de base à une tentative d'identification. L'association des deux difficultés sera obtenue, par exemple,

en demandant un commentaire sur le titre de la première page du dernier numéro du *Canard Enchaîné*. Faire tenir à un ordinateur une conversation traitant de toutes sortes de sujets (de science, d'histoire, d'art, de spectacles, de musique, de gastronomie, etc.) est un objectif extrêmement difficile. L'idée de Turing semble bonne : si on réussit à tromper un juge avec un programme, cela signifie indubitablement qu'on a mis dans l'ordinateur quelque chose ressemblant à ce que nous appelons de l'intelligence. Aujourd'hui, nous sommes encore assez loin d'avoir réussi. Ce qui est obtenu avec les meilleurs programmes est un simulacre de conversation qui, malgré des progrès récents, ne réussit que rarement plus de quelques minutes à entretenir l'illusion. Le système informatique Watson de la firme IBM, qui a gagné au jeu Jeopardy (du même type que le jeu « Questions pour un champion »), marque un progrès vers une forme d'intelligence générale, mais il n'a pas été conçu pour passer le test de Turing.

John Searle pense que réussir ne prouverait rien

Malgré l'évidente difficulté et l'échec relatif dans lequel nous sommes tenus, certains philosophes, dont John Searle, pensent que, même si nous réalisons un programme réussissant le test de Turing, cela ne prouvera pas que nous avons mis de l'intelligence dans notre ordinateur. Son raisonnement peut se résumer de la façon suivante :

- les programmes informatiques sont syntaxiques : ce ne sont que des outils à manipuler des symboles ;
- les pensées humaines ont un contenu sémantique : un sens est attaché aux mots que nous utilisons, qui ne se réduisent pas à de la syntaxe, ce sens provient de propriétés biologiques particulières de nos neurones qu'un programme ne peut pas posséder ;
- donc, les programmes ne penseront jamais et ne seront jamais intelligents.

Cette façon de voir les choses permet, en dehors de toute tentative concrète, de conclure que les ordinateurs programmés ne penseront jamais, et qu'ils ne posséderont jamais aucune intelligence authentique. Et, cela, même si on réussissait à leur faire passer le test de Turing. Pour Searle, la réussite au Test de Turing n'est pas une condition suffisante pour qu'on puisse parler d'ordinateur intelligent.

Nous n'entrerons pas dans les controverses qu'engendre ce type d'analyses. L'article original où Searle exposait en 1980 son point de vue était accompagné de 26 commentaires critiques et, depuis, une multitude d'autres discussions se sont déroulées. Remarquons cependant qu'il paraît un peu étrange de nier qu'un système que rien ne distinguerait, dans un dialogue, d'un être humain possède cette qualité – l'intelligence – qu'on accepte d'attribuer aux humains. Il y aurait comme un privilège accordé, par décret, à une catégorie particulière d'êtres – les humains – qu'on refuserait définitivement à une autre catégorie d'êtres – les ordinateurs.

On est tenté de dire que John Searle manque de fair-play : aujourd'hui, les ordinateurs sont incapables de nous imiter de manière satisfaisante, refuser de les qualifier d'intelligents se justifie mais, si un jour ils réussissent à nous imiter fidèlement, nous devrions alors avoir l'élégance de reconsidérer notre position.

Trop difficile ?

Une autre critique inattendue a été formulée par Robert French à l'encontre du test de Turing. Alors que Searle considère que le test de Turing n'est pas une condition suffisante pour qu'on puisse affirmer qu'un ordinateur pense, Robert French soutient que passer le test Turing n'est pas une condition nécessaire d'intelligence.

Il imagine l'histoire d'un peuple qui ne connaîtrait qu'une seule espèce d'oiseaux : les mouettes. Ce peuple se poserait le problème de réaliser une machine volante et, pour savoir s'il a réussi, utiliserait le *test de la mouette* : une machine sera dite *volante* s'il est impossible de la distinguer d'une mouette dont le comportement est observé à l'aide d'un radar.

Le radar limite la précision de la demande d'imitation comme le dialogue, par le biais d'échanges dactylographiés, limite la précision de la demande d'imitation dans la mise en œuvre du test de Turing. Bien sûr, les avions, les hélicoptères, les montgolfières et même les oiseaux autres que les

Le jeu de Deep Blue ressemble à un jeu humain

Le programme Watson d'IBM vainqueur au jeu Jeopardy

Les succès de *l'Intelligence artificielle* ne sont pas négligeables et conduisent à des situations où des formes affaiblies du test de Turing peuvent être passées par des systèmes informatiques spécialisés. En 1997, l'ordinateur Deep Blue, conçu spécialement pour le jeu d'échecs, a battu Garry Kasparov alors champion du monde et, depuis, d'autres machines ont atteint des résultats comparables. Même les bons joueurs d'échecs reconnaissent que, face à ce genre de programmes et sans information particulière, il leur est impossible de savoir s'ils jouent contre une machine ou contre un être humain. C'est une forme de test de Turing (où le dialogue est limité à des échanges de coups) et donc, dans ce cas, la prédiction de Turing qu'on réussirait avant l'an 2000 a été réalisée.

Plus récemment, en 2011, le programme nommé Watson, développé par la firme IBM, a réussi à gagner un concours contre les champions du jeu Jeopardy (où des questions générales sont posées en langage naturel). Cette victoire a été analysée comme la réussite d'un test de Turing partiel mais, cette fois, en s'approchant d'un peu plus près des conditions imaginées par Turing pour son jeu de l'imitation.

mouettes ne réussiront pas le *test de la mouette* et ne seront donc pas considérés comme capables de voler. Est-ce bien raisonnable ? Certes, non.

Pour Robert French, le test de Turing est une condition suffisante d'intelligence, mais d'*intelligence humaine*. De plus, ce test est lié à la langue utilisée pour les dialogues, ce qui empêche de le considérer comme universel. French affirme qu'il se pourrait bien que nombre de nos comportements dépendent fortement de la façon particulière dont notre cerveau traite l'information au niveau le plus profond – il parle de *processus subcognitifs*. Imposer une imitation servile du comportement humain, c'est se protéger à trop bon compte du risque de devoir admettre que des machines intelligentes sont devenues nos concurrentes.

Aujourd'hui, nous ne savons pas fabriquer d'oiseaux imitant une mouette, mais nous savons fabriquer des objets volants aux qualités remarquables, puisque certains avions sont plus rapides que les plus rapides des oiseaux et volent plus longtemps que les plus endurants des oiseaux. De même, nous ne savons pas concevoir des systèmes informatiques imitant vraiment les humains, mais nous savons en faire ayant des capacités de mémoire et de calcul bien supérieures à nous. N'avons-nous pas déjà fabriqué une forme d'intelligence informatique ?

Programmes de conversation

C'est seulement dans la décennie 1960 que des programmes capables de dialogues en langues naturelles sont apparus (on les nomme *chatbots* ou *chatterbots* pour robots de conversation). Un de ces programmes est le célèbre ELIZA de Joseph Weizenbaum mis au point en 1964 au MIT. Le mécanisme interne du programme est assez simple. Les phrases de l'interlocuteur sont analysées et, à partir de leur structure grammaticale, une réponse est construite en utilisant certains des mots de l'interlocuteur. Des phrases passe-partout, pour détourner la conversation ou pour esquiver les questions, sont aussi utilisées lorsque le programme se trouve embarrassé. Le but d'ELIZA était de jouer le rôle d'un psychanalyste attentif à ce que dit son interlocuteur et l'incitant à se confier. Bien qu'il fût conçu comme un amusement sans prétention, certaines personnes utilisant ELIZA et informées de sa véritable nature se prirent au jeu et se confièrent à lui comme s'il s'agissait d'un être humain. Des utilisateurs passionnés prétendirent même s'y être attachés.

Le prix Hugh Loebner

Depuis 1990, un concours annuel (le *Prix Hugh Loebner*) se déroule et détermine, sur la base d'une procédure inspirée du Test de Turing, le meilleur programme de conversation générale. Aujourd'hui, aucun candidat ne s'est révélé très intéressant et il semble qu'il manque quelque chose pour réaliser de bons systèmes informatiques de dialogue réellement susceptibles de tromper un juge.

Certains chercheurs ont jugé que l'organisation du prix Loebner était inutile et même nuisible au progrès des recherches en *intelligence artificielle*. En effet, disent-ils, en détournant les chercheurs d'objectifs plus raisonnables et en les incitant à accumuler des astuces dans leurs programmes, qui ne sont finalement que de simples bases de données conçues pour faire semblant d'être intelligent sans l'être vraiment, on ne fera pas avancer la véritable compréhension de l'intelligence.

Même si certains chatbots sont conçus pour imiter – naïvement – des personnes humaines ayant existé (John Lennon, Elvis Presley, Jésus, etc.), aujourd'hui, les spécialistes les plus optimistes ne se risquent pas à parier pour un succès définitif avant 2030. Puisqu'il semble que les concurrents du prix Loebner en restent à faire des programmes qui, comme ELIZA en 1964, sont sans véritable ambition, la date réelle d'un succès définitif pour le test de Turing pourrait bien être plus lointaine encore.

En fait, c'est sans doute ailleurs qu'il faut porter notre attention. Le programme Watson de la firme IBM est capable d'accéder à un très grand nombre de connaissances (historiques, géographiques, politiques, etc.) stockées dans ses énormes mémoires et il dispose de multiples outils de traitement du langage naturel. Watson l'a emporté sur les champions humains du jeu Jeopardy qui est du même type que le jeu « Questions pour un champion ». Ce jeu exige pourtant une sorte de sens commun et d'esprit d'analyse très fin pour résoudre les astuces et surmonter les pièges des énigmes soumises aux concurrents. Nous ne savons pas faire de machines passant le test de Turing, mais nous savons en faire disposant de compétences de plus en plus variées et aptes maintenant à manipuler de l'information et à prendre des décisions : il devient donc parfois difficile de ne pas les qualifier d'intelligentes. ■

Bibliographie

- Robert Epstein, *Parsing the Turing Test : Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Springer, 2008.
- Robert French, *Subcognition and The Limits of the Turing Test*, Mind, 1990, 99, pp. 53-66.
- John Searle, *Minds, Brains and Programs. Behavioral and Brain Sciences*, 1980, 3, pp. 417-57.
- Alan Turing, *Computing Machinery and Intelligence*, Mind, 59, 236, pp. 433-460, 1950.

Pages du Prix Loebner. Documents sur le concours : <http://www.loebner.net/Prizef/loebner-prize.html>