



Les machines pensent-elles?



JEAN-PAUL DELAHAYE

Faut-il comprendre pour imiter? La question est posée par les programmes qui simulent (et stimulent) l'intelligence humaine.

Définir l'intelligence n'est pas chose facile et il n'est pas certain que l'on y parvienne un jour de façon satisfaisante. Cette difficulté poussa le mathématicien britannique Alan Turing, il y a un peu plus de 50 ans, à proposer «le jeu de l'imitation» qui fixait un objectif précis à la science naissante des ordinateurs que l'on n'appelait pas encore informatique.

Le jeu de l'imitation consiste à mettre au point une machine qu'il serait impossible de distinguer d'un être humain. 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, désignés par *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 nous réussissions un jour à mettre au point des machines pour lesquelles l'identification correcte est impossible (c'est-à-dire que le juge *J* se trompe dans 50 pour cent des cas, pourcentage identique à ce que donnerait une réponse au hasard), alors nous pourrions affirmer que nous avons construit une machine intelligente ou – ce que nous considérerons comme équivalent – une machine qui pense. Sa procédure aujourd'hui dénommée *test de Turing* a donné lieu à de nombreuses discussions souvent intéressantes et a suscité une série de réalisations informatiques concrètes mises en compétition annuellement lors d'un prix organisé par Hugh Loebner. Les logiciels participant au jeu de l'imitation progressent chaque année.

Alan Turing (1912-1954)

Face aux prétentions parfois déraisonnables de certains chercheurs en Intelligence artificielle et aux doutes non moins étonnants de ceux qui affirment péremptoirement qu'on ne réussira jamais à créer des programmes intelligents, nous proposons un bilan de 50 ans de réflexion et de travail autour de ce fameux test de Turing.

LE TEST DE TURING ET L'INTELLIGENCE

Avant tout examen, mesurons combien la réalisation d'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 : «Combien vaut 429 à la puissance 3?» Si *A* répond 78 953 589 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. Évidemment, les spécialistes qui conçoivent les programmes pour passer le test de Turing ont prévu cette ruse grossière. Leur programme, bien que capable de calculer 429^3 en une fraction de seconde, refusera de répondre, ou demandera dix minutes avant de fournir le 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 100 chiffres aléatoires est un autre exemple) sera traité de la même façon : l'ordinateur fera semblant de ne pas réussir. Pour reconnaître l'être humain, le 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 possible : racontez une histoire drôle à vos interlocuteurs *A* et *B* et demandez-leur

d'expliquer où il faut rire et pourquoi? Les questions d'actualité 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 explicatif du titre de la première page du dernier numéro du *Canard Enchaîné*. Un grand nombre d'autres stratégies envisageables devraient, sans trop de mal, vous conduire à identifier qui est l'homme *H* et qui est la machine *M*. 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 et l'idée de Turing semble donc bonne : si l'on réussit vraiment à tromper un juge avec un programme, cela signifie indubitablement que l'on a mis dans l'ordinateur quelque chose ressemblant à ce que nous appelons de l'intelligence.

Disons-le tout de suite, nous sommes loin d'avoir atteint le but et ce qui est obtenu avec les meilleurs programmes actuels est un simulacre de conversation qui, malgré des progrès récents, ne réussit pas à entretenir l'illusion plus de quelques minutes.

QUE PROUVERAIT LA RÉUSSITE?

Malgré l'évidente difficulté du problème et notre échec relatif, certains philosophes, dont John Searle, pensent que même si nous réalisons un programme passant le test de Turing cela ne prouvera pas que nous avons mis de l'intelligence dans notre ordinateur. Le raisonnement de J. Searle est le suivant :

- les programmes informatiques sont syntaxiques (ils sont formels, 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, sens

que l'on ne peut ramener à de la syntaxe formelle. Ce sens provient de propriétés biologiques particulières de nos neurones qu'un programme, de par sa nature, ne peut pas posséder) ;

– donc les programmes ne penseront jamais et ne seront jamais intelligents.

Cette façon de voir décourage toute tentative concrète. Pour J. Searle, la réussite au test de Turing n'est pas une condition suffisante pour qu'on puisse qualifier les ordinateurs d'intelligents.

Nous n'entrerons pas dans les controverses qu'engendre ce type d'analyses. L'article original où, en 1980, J. Searle exposait son point de vue, était accompagné de 26 commentaires critiques, et depuis une multitude de discussions se sont déroulées. Remarquons cependant qu'il paraît un peu étrange d'affirmer qu'un système, que rien ne distinguerait dans un dialogue d'un être humain, ne possède pas – l'intelligence – qu'on attribue 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. J. Searle manque de fair-play : si un jour les ordinateurs réussissent à nous imiter, alors nous devrons alors avoir l'élégance de les qualifier d'intelligents.

Cette critique de la position de J. Searle est qualifiée de *béhavioriste* : on propose de fonder un jugement uniquement sur les comportements sans entrer dans l'analyse même des processus internes (dans le cerveau et dans l'ordinateur) qui produisent les comportements. Cependant comme le font remarquer A. Saygin, I. Cicekli et V. Akman, cette accusation de béhaviorisme méritera d'être prise au sérieux le jour où, nous humains, disposerons d'autres moyens que l'analyse du comportement pour évaluer l'intelligence des êtres humains !

LE CRI DE LA MOUETTE

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

R. French imagine un peuple qui ne connaîtrait qu'une seule espèce d'oiseaux : les mouettes. Ce peuple se pose le problème de réaliser une machine volante et pour savoir s'il a réussi utilise le *test de la mouette* : une machine est 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

1. LE TEST DE TURING

Dans le jeu de l'imitation proposé par Alan Turing en 1950, le juge essaie de reconnaître, en dialoguant par écrit avec un interlocuteur, si celui-ci est une machine ou un être humain. Le juge ne voit pas son interlocuteur à qui il pose des questions et dont il obtient des réponses. La machine essaie de se faire passer pour un être humain et l'être humain essaie de se faire reconnaître comme tel. Pour Turing, si le juge se trompe dans 50 pour cent des cas, alors on doit considérer que la machine est intelligente et qu'elle pense. La photographie représente les juges dialoguant par ordinateur avec des hommes ou des programmes lors du prix Hugh Loebner.

Si la majorité des gens s'accordent pour dire qu'une machine passant le test de Turing doit être considérée comme intelligente et qu'elle dispose de la capacité de penser, Robert French, lui, considère que le test de Turing est trop difficile et ne caractérise pas l'intelligence en général, car il est trop lié au langage utilisé, à la façon dont fonctionne le cerveau humain au niveau le plus bas (sub-cognitif) et aux traits culturels particuliers de celui qui interroge.

R. French imagine ainsi un peuple qui, ne connaissant qu'un seul animal volant, la mouette, croit définir un critère absolu pour la propriété de voler en formulant le *test de la mouette* : on considérera qu'un objet est capable de voler, si, lorsqu'on observe son image à l'aide d'un radar, il est indiscernable d'une mouette. Il est clair que ce test manque son but, car si, incontestablement un objet passant le test de la mouette doit être considéré comme volant, en revanche aucun avion, ni aucune autre espèce d'oiseaux ou d'insectes ne réussiront à passer le test de la mouette qui est trop difficile. Il en va de même, pour French, de la réussite au test de Turing : c'est une condition suffisante d'intelligence, mais pas une condition nécessaire.

limite la précision de la demande d'imitation comme le dialogue, par le biais d'échanges dactylographiés, limite la portée de l'imitation dans le test de Turing. Bien sûr les avions, les hélicoptères, les montgolfières et même les oiseaux autres que les 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 R. 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. R. French poursuit son analyse et 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 qualifie ces processus de sub-cognitifs). Imposer une imitation servile du comportement humain c'est se protéger à trop bon compte du risque d'admettre que des machines intelligentes sont devenues nos concurrentes. Pour illustrer son idée, R. French imagine un exercice d'évaluation des néologismes. Il propose qu'avant d'appliquer le test de Turing, on fasse évaluer la qualité d'un ensemble de néologismes par un groupe de sujets humains de référence. Par exemple, on envisage les néologismes pour dési-

gner de très longues voitures. Les sujets humains notent de 0 à 10 les néologismes : auto-fusée, fusauto, auto-bas-set, longivoiture, finauto, auto-aiguille, voiture-serpent.

Les notes attribuées constituent des données types et en comparant ensuite, lors du test de Turing, les notes proposées par A et B à ces données, on distinguera l'humain de la machine, car – c'est l'idée de R. French – la manière dont nous notons les néologismes est liée à la façon particulière dont notre cerveau traite le langage, manière qui a peu de chances d'être identique à celle qu'utiliserait l'ordinateur (de même : la façon dont un humain joue aux échecs est très différente de celle utilisée par un programme). L'évaluation des néologismes démasquerait l'ordinateur. Pour R. French, de même qu'il n'est pas nécessaire d'être indiscernable d'une mouette pour voler, il n'est pas nécessaire d'être indiscernable d'un humain pour penser ou avoir un comportement intelligent.

Cependant pour R. French et la majorité des gens qui ont réfléchi à ces questions, il est certain que même si la difficulté du test de Turing est très grande (trop sans doute), la réussite à ce test ne laisserait aucun doute sur le fait que nous avons construit des machines intelligentes. Il semble donc intéressant d'essayer. Qu'avons-nous obtenu ?

