

**Les ordinateurs digitaux peuvent-ils penser ?** (Alan M. Turing)

Les ordinateurs digitaux ont souvent été décrits comme des cerveaux mécaniques. La plupart des scientifiques regardent probablement ces descriptions comme de simples slogans journalistiques, mais d'autres non. Un mathématicien m'a exposé le point de vue opposé plutôt violemment en ces termes "On dit communément que ces machines ne sont pas des cerveaux mais vous et moi savons que c'en sont". Dans cet exposé, j'essaierai d'expliquer les idées derrière les différents points de vue possibles, mais je ne le ferai pas de façon impartiale. J'accorderai davantage d'attention au point de vue qui est le mien, qui est qu'il n'est pas déraisonnable de décrire les ordinateurs digitaux comme des cerveaux. Un point de vue différent a déjà été défendu par le Professeur Hartree.

D'abord, nous pouvons considérer le point de vue naïf de l'homme de la rue. Il entend des compte-rendus surprenants à propos de ce que ces machines peuvent faire : la plupart semblent avoir des capacités intellectuelles qu'il ne possède lui-même pas. Il ne peut l'expliquer qu'en supposant que la machine est une sorte de cerveau, même s'il préfère plutôt ne pas croire ce qu'il a entendu.

La majorité des scientifiques méprisent ces attitudes quasiment superstitieuses. Ils savent quelque-chose des principes sur lesquels ces machines sont construites et de la manière dont on les utilise. Leur organisation générale a été résumée par Lady Lovelace il y a une centaine d'années environ, lorsqu'elle a décrit le moteur analytique de Babbage. Elle dit, comme Hartree l'a citée "Le moteur analytique n'a aucune prétention à initier quoi que ce soit. Il peut faire tout ce qu'on lui ordonne de faire.". Ceci décrit très bien la manière dont les ordinateurs digitaux sont utilisés au jour d'aujourd'hui, et la manière dont ils seront principalement utilisés dans de nombreuses années à venir. Pour le moindre calcul, la totalité de la procédure que la machine va utiliser est planifiée à l'avance par un mathématicien. Moins il y a de doute sur ce qui va se produire, plus le mathématicien est content. C'est comme planifier une opération militaire. Dans ces conditions, on peut dire que la machine n'initie rien.

Il y a pourtant un troisième point de vue, qui est le mien. Je suis d'accord dans la mesure du possible avec l'énoncé de Lady Lovelace, mais je crois que sa validité dépend du fait de considérer la manière dont les ordinateurs digitaux sont utilisés plutôt que celle dont ils pourraient être utilisés. En fait, je crois qu'ils pourraient être utilisés de telle manière qu'on les décrirait adéquatement par le terme cerveaux. Je dirais également que "si une machine peut être décrite de manière appropriée comme un cerveau, alors tout ordinateur digital peut également être décrit ainsi".

---

15 Mai 1951.

© P. N. Furbank, for the Turing estate.  
<http://www.turingarchive.org/browse.php/B/5>

Cette dernière phrase nécessite d'être expliquée. Cela peut sembler surprenant, mais avec quelques réserves, cela semble inévitable. On peut montrer que cela découle d'une propriété caractéristique des ordinateurs digitaux, que j'appellerai leur universalité. Un ordinateur digital est une machine universelle au sens où elle peut remplacer toute machine d'un ensemble très grand. Elle ne remplacera pas un bulldozer ou une machine à vapeur ou un télescope, mais elle remplacera toute autre machine à calculer, c'est-à-dire toute machine dans laquelle on entre des données et qui plus tard renvoie des résultats. Pour que notre ordinateur imite une machine donnée, il est seulement nécessaire de le programmer pour qu'il simule ce que la machine en question aurait fait dans telles circonstances, et en particulier quelles données en sortie elle aurait fournies. L'ordinateur peut ainsi être programmé pour fournir les mêmes réponses.

Si maintenant une machine particulière peut être décrite comme un cerveau, nous n'avons qu'à programmer notre ordinateur pour l'imiter et ce sera aussi un cerveau. Si l'on accepte que les cerveaux réels, comme ceux des animaux, et en particulier ceux des humains, sont des sortes de machines, il en découlera que nos ordinateurs digitaux convenablement programmés se comporteront comme des cerveaux.

Cet argument implique un certain nombre de suppositions qui peuvent être raisonnablement défiées. J'ai déjà expliqué que la machine à imiter doit ressembler plutôt à une calculatrice qu'à un bulldozer. C'est seulement une réflexion au sens où nous sommes en train de parler d'analogues mécaniques des cerveaux, plutôt que des jambes ou des mâchoires. Il est aussi nécessaire que cette machine soit d'un genre dont le comportement peut être prédit par le calcul. Nous ne savons certainement pas comment un tel calcul devrait être fait, et il a même été expliqué par Sir Arthur Eddington que du fait du principe d'incertitude de la mécanique quantique, une telle prédiction n'est même pas possible théoriquement.

Une autre supposition était que la capacité de la mémoire de l'ordinateur devrait être suffisante pour prédire le comportement de la machine à imiter. Il faudrait aussi disposer d'une vitesse de calcul suffisante. Nos ordinateurs actuels n'ont probablement pas assez d'espace mémoire, bien qu'il soit possible qu'ils aient la vitesse de traitement appropriée. Cela signifie en effet que si nous souhaitons imiter quelque-chose d'aussi compliqué que le cerveau humain, nous avons besoin d'une bien plus grande machine qu'aucun des ordinateurs dont nous disposons actuellement. Nous avons vraisemblablement besoin de quelque-chose qui soit au moins cent fois plus grand que l'ordinateur de Manchester. Alternativement bien sûr, une machine de taille égale ou plus petite pourrait aller si des progrès suffisants étaient faits en termes de stockage de l'information.

Il faudrait noter qu'il n'est pas nécessaire d'augmenter la complexité des ordinateurs utilisés. Si nous essayons d'imiter des machines encore plus compliquées ou des cerveaux,

nous devons utiliser des ordinateurs de plus en plus gros pour le faire. Nous n'avons pas besoin d'en utiliser qui soient de plus en plus compliqués. Cela peut sembler paradoxal, mais l'explication n'en est pas difficile. L'imitation d'une machine par un ordinateur nécessite non seulement que nous ayons fabriqué l'ordinateur, mais également que nous l'ayons programmé de la façon appropriée. Plus la machine à imiter est compliquée, plus le programme doit l'être. Peut-être cela pourra-t-il être rendu plus clair par une analogie. Supposons que deux hommes veuillent écrire leur auto-biographie, et que l'un ait eu une vie pleine d'événements mais que très peu d'événements se soient produits dans la vie de l'autre. Il y aurait deux difficultés troublant l'homme avec une vie pleine, et qui le gêneraient plus que l'autre. Il lui faudrait plus de papier et il aurait plus de difficulté à décider de ce qu'il va écrire. Le fait de le pourvoir en papier ne semble pas être une difficulté sérieuse, à moins par exemple qu'il soit sur une île déserte, et dans tous les cas, ce ne serait qu'un problème technique et financier. L'autre difficulté serait plus fondamentale et deviendrait plus sérieuse encore si plutôt que d'écrire sa vie, il s'agissait d'effectuer un travail sur un sujet auquel il ne connaît rien, disons sur la vie sur Mars. Notre problème pour programmer un ordinateur pour qu'il se comporte comme un cerveau est quelque-chose qui ressemble à l'écriture de ce traité, et dans tous les cas, nous ne savons pas ce que nous devrions écrire si nous l'avions. C'est un piètre état de choses, mais, pour poursuivre l'analogie, il faudrait savoir quoi écrire, et apprécier le fait que la plupart des connaissances peuvent être incarnées dans les livres.

Au vu de cela, il semble que le plus sage étayage sur lequel critiquer la description des ordinateurs digitaux en tant que "cerveaux mécaniques" ou "cerveaux électroniques" est que, bien qu'ils puissent être programmés pour se comporter comme des cerveaux, nous ne savons pas à présent comment cela pourrait être fait. Avec cet argument, je suis en total accord. Il laisse ouverte la question de savoir si nous finirons par réussir ou pas à trouver un tel programme. Je pense par exemple probable qu'à la fin du siècle, il sera possible de programmer une machine qui répondra à des questions de telle manière qu'il sera extrêmement difficile de deviner si les réponses en sont données par un homme ou par une machine. J'imagine quelque-chose comme un examen de viva voce, mais avec des questions et réponses toutes tapées à la machine de manière à ce que nous n'ayons pas à considérer des éléments non pertinents comme la qualité avec laquelle la voix humaine peut être imitée. C'est seulement mon opinion ; il y a de la place pour de nombreuses autres opinions.

Il reste des difficultés. Se comporter comme un cerveau nécessite le libre-arbitre, mais le comportement d'un ordinateur digital, lorsqu'il a été programmé, est complètement déterministe. Ces deux faits devraient être réconciliés en quelque sorte, mais faire cela semble nous ramener à la controverse d'un autre âge du "libre-arbitre-et-déterminisme". On ne peut pas s'en sortir. Il est possible que l'impression de libre-arbitre que nous partageons tous ne soit qu'une illusion. Ou bien il est possible que nous ayons effecti-

vement un libre-arbitre, mais qu'il soit impossible de dire qu'il en est ainsi uniquement en observant notre comportement de l'extérieur. Dans ce dernier cas, aussi réaliste que soit la manière dont une machine pourra imiter le comportement humain, elle ne pourra être considérée que comme un simulacre. Je ne sais pas comment nous pourrions jamais décider entre ces deux alternatives mais quelle que soit l'alternative correcte, il est sûr qu'une machine sensée imiter un cerveau doit sembler se comporter comme si elle avait un libre-arbitre, et c'est aussi bien de se demander comment on pourrait faire ça. Une des possibilités est que son comportement dépende de quelque-chose comme une roulette ou un atome de radium. Le comportement de ces dispositifs pourrait sembler pouvoir peut-être être prédit, mais si tel est le cas, nous ne savons pas comment faire cette prédiction.

Il n'est, cependant, pas vraiment nécessaire de faire ça. Il n'est pas difficile de concevoir des machines dont le comportement semble assez aléatoire à quiconque ne connaît aucun détail de leur construction. Naturellement, l'inclusion de cet élément aléatoire, quelle que soit la technique utilisée, ne résoud pas notre problème principal, qui est de savoir comment programmer une machine pour imiter un cerveau, ou, comme nous pourrions le dire plus brièvement, même si c'est moins précis, pour *penser*. Mais cela nous donne une indication sur ce à quoi pourrait ressembler le processus. Nous ne devons pas nous attendre à toujours savoir ce que l'ordinateur va faire. Nous devrions être content si la machine nous surprend, de la même façon que nous sommes content lorsqu'un élève fait quelque chose que nous ne lui avons pas explicitement demandé de faire.

Reconsidérons maintenant l'énoncé de Lady Lovelace : "La machine peut faire tout ce dont on sait comment lui apprendre à le faire". Le sens du reste du passage est que l'on est tenté de dire que la machine ne peut faire que ce dont on sait lui expliquer comment le faire. Mais je pense que cela peut ne pas être vrai. Certainement que la machine ne peut faire que ce que nous lui ordonnons, si elle faisait autre chose, cela proviendrait d'un problème mécanique. Mais il n'est pas nécessaire de supposer cela, quand nous lui donnons ses ordres, nous savons ce que nous faisons, ce que les conséquences de ces ordres vont être. On n'a pas besoin de comprendre comment ces ordres vont amener la machine à avoir le comportement associé, ni de comprendre le mécanisme de germination quand on plante des graines dans le sol. La plante poussera qu'on le comprenne ou pas. Si nous donnons à la machine un programme qui résultera dans le fait qu'elle fera une chose intéressante que nous n'avions pas anticipée, je serais enclin à dire que la machine a initié quelque-chose, plutôt que de dire que son comportement était implicite dans le programme, et qu'ainsi l'originalité n'est que de notre fait.

Je n'essaierai pas d'en dire beaucoup sur la manière dont ce processus consistant à "programmer une machine à penser" doit être fait. Le fait est que nous n'en savons que très peu à ce sujet, et que très peu de recherche a déjà été faite. Il y a de nombreuses idées,

mais nous ne savons pas encore lesquelles sont importantes. Comme dans une histoire de détective, au début de l'enquête, tout détail peut être important pour l'enquêteur. Quand le problème a été résolu, seuls les faits essentiels doivent être racontés au jury. Mais à présent, nous n'avons rien à montrer devant un jury. Je dirai seulement cela, je crois que le processus sera très lié à celui de l'enseignement.

J'ai essayé d'expliquer quels sont les arguments rationnels principaux pour et contre la théorie selon laquelle on pourrait fabriquer des machines pensantes, mais il faut dire quelque chose à propos des arguments irrationnels. Beaucoup de personnes sont extrêmement opposées à l'idée qu'une machine puisse penser, mais je ne crois pas que ce soit pour aucune des raisons que j'ai mentionnées, ou pour une quelconque autre raison rationnelle, mais simplement parce qu'ils n'aiment pas cette idée. On peut voir de nombreuses caractéristiques qui rendent cette idée désagréable. Si une machine pense, elle pourrait penser plus intelligemment que nous ne le faisons, et alors où serions-nous ? Même si nous pouvons maintenir les machines dans une position d'esclaves, par exemple, en coupant leur alimentation à des moments stratégiques, nous pourrions nous sentir grandement humiliés en tant qu'espèce. Un danger et une humiliation similaire nous menacent lorsque nous envisageons la possibilité que nous puissions être remplacés par des cochons ou des rats. C'est une possibilité théorique qui est très controversée, mais nous avons vécu avec des cochons et des rats depuis si longtemps sans que leur intelligence n'augmente beaucoup, que nous ne sommes pas davantage troublés par cette possibilité. Nous ressentons que si ça devait arriver un jour, cela n'advierait pas avant quelques millions d'années. Alors que le nouveau danger semble plus proche. S'il advient, ce sera vraisemblablement au prochain millénaire. C'est dans un avenir lointain mais pas dans un avenir astronomiquement lointain, et c'est certainement quelque-chose qui peut nous rendre anxieux.

Il est coutumier, dans un exposé sur ce sujet, de fournir un peu de réconfort, en disant que quelques caractéristiques humaines particulières ne seront jamais imitées par une machine. On pourrait par exemple dire qu'aucune machine n'écrira jamais très bien l'anglais, ou qu'elle ne sera jamais attirée par le sexe ou ne fumera la pipe. Je ne peux pas fournir un tel argument réconfortant, car je crois qu'aucune telle limite ne peut être fixée. Mais j'espère certainement et je crois qu'aucun grand effort ne sera fait pour mettre en machine les caractéristiques les plus distinctives des humains, mais des caractéristiques non intellectuelles comme la forme du corps humain. Cela me semble assez stupide de faire de telles tentatives et leurs résultats ont quelque chose d'aussi déplaisant que lorsqu'on pense aux fleurs artificielles. Les tentatives de faire penser les machines me semblent d'un tout autre ordre. Le processus complet de la pensée humaine est encore plutôt mystérieux pour moi, mais je crois que la tentative de créer une machine pensante nous aidera grandement à trouver comment nous pensons nous-mêmes.