

Sur l'intelligence artificielle comme modélisation de l'humain

La recherche autour du thème de l'intelligence artificielle, assimilée à l'informatique, a été menée depuis ses débuts par des mathématiciens et informaticiens, par des praticiens de sciences dures. Depuis ses débuts ce nouveau domaine de recherche a connu plusieurs théoriciens, beaucoup de définitions, ce qui le rend peu cernable. Effectivement il serait difficile de définir en quelques lignes un sujet aussi plein de tensions et de désaccords. Il n'y a pas une vérité absolue encore dans cette science encore bien jeune. L'étude de l'intelligence artificielle (par les sciences dures) est l'héritage d'une longue, très longue, antique réflexion sur la logique, la conscience, l'être humain et du développement des sciences bien sûr. L'intelligence artificielle et la philosophie portent sur la connaissance, l'une et l'autre se posent des questions générales sur la structure des connaissances. Bien des philosophes se sont penchés sur des questions comme la raison, l'intelligence, les mécanismes de la connaissance, la machine qu'est l'humain, la conscience... Descartes, par exemple, dégage la notion moderne de l'esprit, Kant aussi reprend certaines de ces thématiques dans *Critique de la raison pure*. L'idée de machine pensante ou de mécanisme de l'être humain avait toujours été utilisée comme métaphore pour illustrer des thèses, elle est désormais une éventualité, une réalité potentielle. L'intelligence artificielle est-elle seulement héritière de la philosophie, ou soulève-t-elle aussi des nouveaux questionnements, de nouvelles approches de pensée ?

On peut voir les recherches dans le champ de l'intelligence artificielle comme une branche de l'informatique, comme la recherche d'une parfaite simulation de l'intelligence et du savoir. Une science où on part de théories sur ces sujets, à la recherche de réponses pratiques, programmes, logiciels, qui ne seront finalement que reflet de ces théories sur l'intelligence. Ou alors considérer l'intelligence artificielle comme une science qui cherche à comprendre les mécanismes de la connaissance, qui est en permanence en questionnement et où chaque nouvelle réponse apporte une nouvelle question.

Le philosophe Hubert L. Dreyfus, qui critique fortement les théoriciens de l'intelligence artificielle dans son texte *Alchimie et Intelligence Artificielle* où il énumère les échecs et déceptions des résultats de la recherche en IA, tient à prouver dans son texte le lien étroit entre l'IA et l'étude de la métaphysique occidentale. Il pose l'intelligence artificielle à une place centrale dans la réflexion qui vise à saisir l'homme, Dreyfus présente l'intelligence artificielle comme l'aboutissement mais

aussi la fin de 2500 ans de métaphysique. En supposant que la réussite de l'un entraînerait l'inutilité de l'autre. Dans ces textes Dreyfus parle de l'intelligence artificielle à la fois comme conséquence d'une étude de l'homme et de l'ontologie occidentale mais aussi comme propulseur de nouvelles idées. Il établit une relation de cause à effet entre le mode de pensée de certains philosophes comme Platon ou Leibniz à l'approche de Marvin Minsky dans le fait de supposer que le monde et les hommes doivent être représentés sous forme d'un système de descriptions fortement structuré, dans lequel les descriptions elles-mêmes sont bâties à partir d'éléments premiers.

« Si réellement nous sommes à la veille de créer une intelligence de synthèse, nous sommes près d'assister au triomphe d'une conception très particulière de la raison. De fait, si l'on peut effectivement douer de raison un ordinateur, voilà qui viendra confirmer cette conception de l'homme que depuis deux mille ans les penseurs de l'Occident cherchent vainement à étayer, mais sans avoir pu, jusqu'à présent, la formuler ou la soumettre à expérimentation, faute de l'outil dont ils disposent désormais: l'homme perçu comme objet. Mais si, d'un autre côté, l'intelligence artificielle se révèle une chimère impossible atteindre, alors nous devons chercher à distinguer la manière dont raisonnent les machines de celle dont nous raisonnons, nous. Et cela modifiera radicalement notre perception de nous-mêmes. Par conséquent, l'heure a donc sonné ou bien d'accorder à la tradition philosophique le bien-fondé de son intuition centrale, ou bien d'abandonner cette conception de la nature humaine comme une mécanique, attitude qui s'est progressivement répandue en Occident au cours des vingt derniers siècles »

John Searle, philosophe américain, définit deux types d'intelligence artificielle, et tout en respectant l'intelligence artificielle faible qui consiste à dire que des machines pourraient agir comme si elles étaient intelligentes, il qualifie d'impossible l'intelligence artificielle forte dit que les machines se comportant ainsi possèderaient vraiment une intelligence et non pas seulement une simulation de celle-ci. Searle repose son argumentation sur le fait que les ordinateurs n'accomplissent que des opérations formellement explicitées alors que l'esprit humain ne se résume pas à des processus syntaxiques. Il appuie ses propos sur la différence entre syntaxe et sémantique. Il maintient que l'esprit humain de par son intentionnalité a en plus d'une syntaxe une sémantique et réduit l'intelligence artificielle à des systèmes formel, donc à des manipulations syntaxiques. Cela voudrait dire qu'en aucun cas l'intelligence artificielle pourrait prétendre décrire l'esprit effectivement en faisant cette distinction le philosophe nie tout intérêt ontologique à l'intelligence artificielle

tout en reconnaissant sa technicité et sa pertinence scientifique. Searle illustre ces propos par une expérience de pensée devenue célèbre aujourd'hui appelée la chambre chinoise.

Beaucoup de philosophes se sont positionnées aussi pour ou contre l'intelligence artificielle en s'appuyant sur des thèses philosophiques pour critiquer la possibilité même de l'existence d'une machine pensante. Cela a soulevé des questions concrètes sur la conscience et l'esprit. La conscience est-elle physique ou immatérielle ? C'est à dire est-elle le résultat de courants électriques et n'existerait pas sans un cerveau et donc pourrait être reproduite artificiellement ? Cette idée voudrait que les états mentaux dépendent d'états physiques. Ou est-elle comme l'a écrit Descartes un processus sans étendue spatiale ni propriétés matérielles ? Bien sûr il n'y a pas de réponse, mais beaucoup de discussions, surtout sur les mécanismes de la pensée, de la conscience et du fonctionnement du cerveau. Descartes considère l'intelligence, et la conscience de celle-ci comme étant propre à l'Homme. Cette question sur la conscience s'étend à des questions sur l'intuition et les sentiments, est-ce que l'homme est double ? Une partie matérielle et une partie immatérielle ? Et si au lieu de se poser la question de savoir si une intelligence artificielle est possible et d'essayer d'y répondre avec des arguments empruntés à la philosophie on regardait les progrès dans cette science en espérant y trouver des réponses à des questions posées depuis des siècles sur l'étude de l'esprit, de l'intelligence et de la conscience humaine ? Les résultats de la recherche en intelligence artificielle ne pourraient-elles pas, dans le futur, ouvrir les portes à une connaissance plus précise de la complexité humaine ? Il se pourrait que les contraintes techniques aient engendré un mode d'approche original qui repris sur le plan philosophique serait à la source de nouvelles idées.

Le débat de savoir si la philosophie et l'intelligence artificielle étaient liées ou si il était pertinent de parler d'une philosophie de l'intelligence artificielle est né aux Etats-Unis. Est-ce que les philosophes ne prêteraient pas aux théoriciens de l'intelligence artificielle des desseins métaphysiques qui ne sont pas les leurs ? Je ne cite ici que deux philosophes et deux approches de l'intelligence artificielle par ce domaine, mais il en existe bien d'autres plus complexes, plus ou moins valides.

Bibliographie :

Jean-Gabriel Ganascia, *L'âme-machine*, les enjeux de l'intelligence artificielle, édition du Seuil, janvier 1990.

Jean-Gabriel Ganascia, *L'intelligence artificielle*, Flammarion, 1993.

Stuart Russel et Peter Norvig, *Intelligence Artificielle*, Pearson éducation 3eme édition, décembre 2010.

H.L. Dreyfus, *Intelligence artificielle : mythe et limites*, Flammarion, 1984.

John R. Searle, *Du cerveau au savoir*, Hermann, 1985.

Marvin Minsky, *La société de l'esprit*, InterÉditions, 1988.

[www.http://aitopics.org/](http://aitopics.org/)