

# **Intelligence artificielle: Faux espoirs et réelles contraintes**

## **Résumé**

*Cet article poursuit l'étude critique de l'intelligence artificielle (IA) déjà amorcée par d'autres, dans le but de relancer la réflexion sur le sujet. L'IA peut-elle tenir ses promesses? Notre schème thématique comporte un court rappel de la discussion sur l'IA et la considération du sujet sous l'angle d'un ouvrage récent qui contribue au progrès du débat.*

## **Abstract**

*This paper, based on studies critical of Artificial Intelligence (AI), questions the assumed promises of AI. Our thematic scheme calls for a short reminder of the debate over AI, and the consideration of the subject through the point of view of a recent work that contributes to the advancement of the discussion.*

## ***Intelligence artificielle et éducation***

L'intelligence artificielle (IA) est certainement un sujet à la mode: dans les milieux de recherche, on multiplie depuis une décennie le nombre de publications sur ce thème passionnant. L'imagination populaire, attisée par les échos qu'en font les médias, ne doute pas un seul instant de la réalité de l'intelligence artificielle, dont les robots industriels ne seraient qu'un des signes précurseurs. Chercheurs ou simples citoyens, ce qui nous fascine et nous captive tous est l'impression d'y percevoir le reflet de notre propre intelligence, l'intelligence humaine. Dans cette perspective enivrante, tous les espoirs sont permis. . . , mais sont-ils tous pour autant fondés?

Nous prenons peu à peu du recul: à l'expression "systèmes experts", par exemple, les spécialistes préfèrent aujourd'hui "système à base de

connaissances". Pourquoi cette nouvelle humilité? La pratique concrète de la recherche en IA conduit les chercheurs à mieux apprécier les difficultés et les obstacles rencontrés et à mieux cerner l'écart entre des modèles théoriques et des réalisations tangibles, car les contraintes sont réelles.

Cet article continuera donc sur la lancée critique déjà amorcée par d'autres. Il s'agit d'une réflexion philosophique sur l'IA, plutôt qu'un discours savant sur les techniques de l'IA. Nous croyons qu'il est essentiel d'identifier les faux espoirs et les réelles contraintes que nous associons à l'IA. C'est à ce prix seulement que nous pouvons évaluer à leur juste mesure les promesses de l'IA.

Il nous semble important de traiter ce sujet du fait que les concepts véhiculés par l'IA (raisonnement, abstraction, jugement, intelligence) sont des concepts fondamentaux puisqu'ils touchent les capacités humaines qui, elles, sont à la base de l'éducation.<sup>1</sup> Cela, indépendamment des applications pratiques de l'IA. Soulignons que toute prétention de technicité en IA a été évitée.

### *Schème thématique de départ*

L'IA pourra-t-elle un jour atteindre le stade de la pensée humaine? Le débat se poursuit, la bataille menée autour de certains laboratoires d'informatique est virulente, et les écrits de ceux qui croient en l'IA et de ceux qui n'y croient pas sont nombreux. Pour notre part, nous allons considérer le sujet principalement sous l'angle d'un ouvrage récent intitulé, *The Emperor's New Mind* (Penrose, 1990). Se basant sur des lois physiques, ce dernier rejette la possibilité d'une intelligence artificielle et se dit convaincu que l'ordinateur ne pourra jamais reproduire l'esprit humain. Ses propos contribuent à une réflexion plus poussée. Mais pour mieux situer le débat, un court rappel s'impose.

### *Que disent les adeptes de l'IA?*

La pierre angulaire de la recherche en IA, y compris celle de Minsky(1986), l'un de ses fervents partisans, est que l'homme<sup>2</sup> est une machine pensante. En élargissant la notion d'intelligence, cet auteur soutient que la machine est intelligente à partir du moment où elle exécute des tâches qui, si elles étaient accomplies par des hommes, seraient considérées comme intelligentes.<sup>3</sup> Ce sont des termes empruntés à Turing (1950) qui devinrent le point de mire de beaucoup d'articles en philosophie et en IA.<sup>4</sup> Selon Minsky, le cerveau humain ne serait au fond qu'une machine complexe, issue d'une longue évolution, tandis que les machines savantes n'ont que quelques dizaines d'années d'existence. En d'autres termes, le cerveau humain ne serait qu'une somme d'agents qui, en interaction, forment ce qu'on appelle l'esprit. Pour l'instant, la complexité des relations entre ces agents nous échappe, dit ce chercheur. Un jour viendra pourtant, estime-t-il, où des instruments nous

permettront de “voir” dans le cerveau cette circulation des idées - la combinaison de courants électriques et de substances chimiques - qui constitue ce que nous nommons par convention l'intelligence. A ceux qui soulignent que ces machines ne comprennent pas le sens des mots, Minsky répond qu'elles finiront par le comprendre. Comment? A l'aide de réseaux d'ordinateurs opérant en parallèle qui rempliront l'équivalent des fonctions du cerveau et parviendront inéluctablement à des performances analogues à celles du cerveau. Selon lui, ce n'est qu'une question de temps.

Elève de Minsky et inspiré par lui, Winston (1984), désire rendre les machines plus intelligentes, 1) comme moyen pour mieux comprendre l'intelligence humaine et 2), dans la conviction que rendre ces machines plus intelligentes c'est leur donner plus d'utilité.

Il définit l'IA comme suit: “l'IA est l'étude des concepts qui permettent de rendre les machines (ordinateur, robot ou autre) intelligentes. Mais qu'est-ce que l'intelligence, continue Winston: est-ce la faculté de raisonner, celle d'apprendre et d'exploiter un savoir, ou celle de percevoir et de manipuler des objets du monde réel? L'intelligence intègre, sans nul doute, toutes ces facultés, mais ne saurait se réduire à cette simple énumération. Aucune définition exhaustive de l'intelligence ne semble possible, parce qu'elle se présente comme un amalgame d'innombrables talents nous permettant de représenter l'information et de la traiter” (p.1). En fait, Winston indique qu'il ne faut pas définir l'intelligence étroitement. Il fait valoir que le but de l'IA serait la compréhension de l'intelligence en tant que traitement plausible de l'information. En d'autres termes, intelligence est synonyme de calcul plausible. Il ignore la qualité de “bon sens” qui manque à la machine et la liaison entre l'intelligence et la conscience des choses qui sous-tend tout acte intelligent.

En tout état de cause, les chercheurs en IA se proposent de doter la machine de comportements proches de ceux de notre pensée, en particulier de la faculté que nous avons d'apprendre, de faire des déductions, de formuler des hypothèses et d'inventer de nouveaux concepts. Et en réaction aux opposants de l'IA, ils déclarent que personne n'accepte d'être comparé à une machine. La frustration dérive, selon eux, de ce que nous en avons une vision périmée. Selon ces militants, dire que notre cerveau est une machine n'est pas une vision réductionniste; c'est présager ce que seront les machines complexes du futur.

### *La robotique de Callatay*

Callatay (1986)<sup>5</sup> s'intéresse au cerveau/machine et ignore le côté esprit. Il soutient que “des notions d'évaluation, des facultés de jugement et l'auto-suffisance (qui manquent à l'ordinateur), ne sont pas définies en science de l'informatique”. Mais, rappelant la tortue de Gray (1953) qui est programmée pour dévoiler ses “sentiments” (dès que ses batteries s'affaiblissent - “dès

qu'elle a faim ou sent de la peine" - elle s'achemine vers un courant électrique, se branche, et dès que ses batteries sont chargées et qu'elle "sent le plaisir de ne plus être affamée", elle se débranche, pour renouveler ses aventures sur le plancher), Callatay affirme que ces "concepts" peuvent être appliqués à un robot. Cependant, dans sa conclusion, il admet que plus nous étudions l'intelligence naturelle, plus nous sommes impressionnés par sa supériorité sur les méthodes actuelles en IA, en particulier dans le domaine de la récupération de l'information, de l'apprentissage et de la découverte de nouveaux concepts. Rapportons aussi que Winograd (l'un des piliers du laboratoire en IA au M.I.T.) et son co-auteur Flores (1986) postulaient récemment "qu'il est impossible de construire des machines qui exhibent ou modélisent avec succès un comportement intelligent" (p. 36).<sup>6</sup>

### *Que répondent les adversaires de l'IA?*

C'est de l'euphémisme de la part des "mordus" des machines dites savantes, répliquent ceux qui pensent que l'IA ne pourra atteindre la capacité émotive de l'être humain. Par ailleurs, même les "systèmes experts" - le champ de l'IA où il y a eu le plus d'effort et de succès ces dernières années - même ces systèmes ne traitent pas le sens des entités qu'ils renferment, mais seulement des relations formelles entre elles, établies par des hommes qui ont accès au sens. A ceux qui déclarent que les machines intelligentes sont programmées pour écrire leur propre programme (qui devient un sous-programme), les adversaires répondent que l'ordinateur n'est qu'un simple exécutant. Qu'une donnée soit incomplète, qu'une instruction soit mal formulée et il se plaint lamentablement. Tout le cheminement des machines est **algorithmique**. Elles ne font que suivre les directives codées par le programmeur, même si le dote intentionnellement ou **accidentellement** la machine de qualités essentielles qu'il n'a pas anticipées. Par contre, l'homme agit par intuition, par la perception de l'ensemble, avant même de passer au détail, et non pas par **algorithme**, comme le font toutes les machines, soient-elles les plus rapides au monde, même celles dites de la cinquième génération.

### *La notion de sens*

Pour sa part, Arzac (1987) affirme que les termes "machines pensantes" ou "systèmes intelligents" sont des abus de langage. L'ordinateur manipule des caractères, dit-il, (des symboles, dans les termes des chercheurs en IA), alors que par notre esprit nous réfléchissons. Mais admettons que l'homme intelligent serait une machine, continue Arzac, il n'est pas qu'intelligence: il est capable d'amour, d'émotions, de création artistique, en plus de l'intuition.

### *The Emperor's New Mind (Penrose)*

Nous passons maintenant à l'ouvrage que nous nous sommes proposé de traiter au début de notre travail.

Pour ce physicien-mathématicien, le fait même que notre esprit nous guide vers des vérités non-traitables par ordinateur devrait nous convaincre que ce dernier ne remplacera jamais notre esprit. Dans cet ouvrage, il s'élève à des sphères philosophiques en cherchant des explications sur l'univers et le monde qui nous entoure.

Des machines sont fortes au jeu d'échecs, dit-il dans son introduction, comprennent-elles pour cela mieux le jeu? Puisque l'électricité se transfère par des fils plus rapidement qu'une autre forme d'énergie (sauf la lumière), l'ordinateur manipule des symboles plus rapidement qu'un calculateur mécanique et remplit des fonctions complexes. Mais, "comprend"-t-il ce qu'il fait d'une façon supérieure à la compréhension d'un abaque? Même si on peut programmer des "sentiments" sur une échelle de 100 (le score -100 représentera la peine, tandis que le score +100 représentera le plaisir (par analogie à la tortue de Gray (1953), citée plus haut), de quel droit pouvons-nous soutenir que la machine sent de la peine ou du plaisir, demande Penrose? Les ordinateurs ne pourront jamais penser, dit-il, car les lois de la nature, la physique quantique incluse (nous reviendrons plus loin sur ce point précis), ne le permettent pas.

Dans le développement de sa thèse, l'auteur aborde des sujets tels que la théorie des nombres et la physique quantique. Il insiste en particulier sur certaines lois physiques qui affectent l'esprit du fait que notre cerveau, biologique de nature, est assujéti aux mêmes lois.

Tous les ordinateurs se servent d'algorithmes ou d'un ensemble de règles qui indiquent comment résoudre des problèmes. Or, il y a des problèmes qui ne peuvent pas être abordés par des méthodes formelles. Le meilleur exemple nous est offert par Kurt Gödel (1931, cité par Penrose, p. 102), dont le théorème démontre que dans tout système mathématique on trouve des propositions qui ne sont pas démontrables par les méthodes existantes, ce qui confirme les limites des systèmes formels.

### *La théorie des nombres*

La théorie des nombres à l'appui, Penrose continue la discussion sur les limites des algorithmes dans la résolution de problèmes complexes. D'ailleurs, bien des parties du domaine mathématique - les plus profondes et les plus intéressantes - possèdent un caractère non-algorithmique. Il se pose les questions suivantes: Se peut-il que le cerveau humain (que pour les besoins de la discussion, il considère en tant qu'"appareil physique", quoique compliqué, délicat et subtil en structure) suive la théorie quantique (en physique, quantités minimales d'énergie pouvant être émises, propagées ou absorbées)?<sup>7</sup> L'esprit est-il assujéti aux lois physiques (car selon l'auteur, le cerveau semble suivre le phénomène quantique)? Comprendons-nous les voies où les effets quantiques

peuvent être utilisés dans la résolution de problèmes et la formulation de jugements? Devons-nous nous avancer au-delà de l'actuelle théorie quantique pour expliquer le fonctionnement du cerveau? Nous devons comprendre, dit-il, la marche de la Nature pour évaluer comment la conscience (notion, sentiment) en fait partie.

### *La théorie quantique*

Il se peut que le phénomène de notre conscience des choses ne puisse être compris dans des termes de physique classique (qui inclut, entre autres, la mécanique de Newton et la relativité d'Einstein), dit Penrose. Il se peut, en un certain sens, que nous devions, en tant qu'êtres conscients, vivre dans un monde régi par la notion quantique, plutôt que dans un monde régi par les lois de la physique classique, malgré la richesse et les mystères qui y sont déjà présents.

Mais comment la notion d'état quantique (une réalité physique objective) se présente-t-elle? Il existe une équation précise qui nous fournit la notion d'évolution-temps déterministe pour cet état. Cependant, il y a quelque chose de curieux dans la relation entre l'évolution-temps de l'état quantique et le comportement de la réalité physique qui est censée se produire. De temps en temps, quand nous considérons qu'une "mesure" a eu lieu, nous nous apercevons qu'il y a eu un "saut" d'un état quantique à un autre, sans traverser l'espace entre eux. Penrose avance une théorie spéculative que ces "sauts" d'un état quantique à un autre des neurones du cerveau sont à la base de notre conscience et de la pensée créatrice de l'individu.

### *Cerveau réel et modèles de cerveau*

Un chapitre de Penrose porte sur le cerveau réel et les modèles de cerveau. L'auteur compare l'activité du cerveau humain et celle des ordinateurs actuels. La fréquence d'impulsions et de décharges neuronales semble être au désavantage du cerveau par comparaison avec l'ordinateur. Cependant le cerveau a une plasticité - la capacité de renouveler ses synapses, que l'ordinateur ne possède pas. Selon une théorie dominante, la mémorisation à long terme du cerveau se fait à l'aide des synapses (les zones d'interaction) qui permettent de stocker l'information nécessaire. Ainsi le phénomène de plasticité du cerveau devient un trait essentiel de l'activité cérébrale. (A cet égard, Callatay<sup>8</sup> qui entre ici en ligne de compte, s'interroge et répond: Comment se fait-il que notre organisme, dans lequel les protéines sont renouvelées chaque mois, peut enregistrer des milliards de "bouts de mémoire" durant une centaine d'années? Selon lui, la mémoire possède un mécanisme homéostatique qui n'est pas là simplement pour modifier les quantités de protéines).

### *La conscience des choses*

Mais comment expliquer la conscience et la compréhension des choses? Comment nous vient l'inspiration? Penrose pense qu'à leur état actuel, ni la mécanique classique, ni la mécanique quantique ne peuvent expliquer la façon dont nous pensons. Un ingrédient **non-algorithmique** logé, semble-t-il, dans notre cerveau paraît déterminer le processus conscient de notre pensée. Selon lui, la question de l'intelligence est subsidiaire à celle de la conscience. Il n'y a pas une véritable intelligence sans un état conscient. C'est le défaut des chercheurs en IA de parler d'intelligence, en ignorant la conscience qui la soutient.

Bien qu'il ne soit pas capable de le démontrer (mais rappelons Gödel, dont le théorème prouve que dans tout système mathématique, et par extension, dans le monde physique qu'il représente, on trouve des propositions qui ne sont pas démontrables), Penrose est persuadé que notre état conscient suit les lois de base de l'univers (les lois de la gravité et des quanta) et que notre pensée créatrice est la manifestation de neurones cérébraux passant par "sauts" d'un état quantique à un autre, sans traverser l'espace entre eux. Quant aux chercheurs en IA, ils sont capables de simuler des "états intelligents", mais ne seront jamais capables de créer des états conscients, qui sont inséparables des précédents.

### *Conclusion*

Dans notre schème thématique de départ, nous nous sommes posé la question: l'IA pourra-t-elle un jour atteindre le stade de la pensée humaine? Pour pouvoir y répondre, nous avons essayé de lever le voile sur les prétentions de certains chercheurs en IA qu'intelligence est synonyme de calcul plausible, leur hypothèse de base étant que nos processus mentaux sont algorithmiques. Pour poursuivre le pour et le contre du débat, nous avons aussi rapporté certains ouvrages consacrés à la notion de sens en informatique.

Nous avons suivi la thèse de Penrose qu'intelligence et conscience (notion, sentiment) vont de pair. Et puisque l'ordinateur ne possède pas des états conscients, il ne peut pas reproduire l'esprit humain. Pour ce chercheur, le phénomène de notre état conscient (l'esprit) suit les lois de base de la nature (les lois de gravité et des quanta).

Quant à nous, pour répondre à notre question de départ, nous faisons nôtre l'interprétation de Penrose: puisque l'ordinateur ne possède pas des états conscients, l'IA n'atteindra pas le stade de la pensée humaine.

Toutefois, on ne peut nier que suite à d'ingénieuses programmations, des succès individuels ont été atteints, chacun d'eux constituant un petit

triomphe d'ingénierie, une solution particulière adaptée à un problème spécifique, mais non généralisable.

Il faut, cependant, noter que les chercheurs en IA sont restés sensibles à la critique. Est-ce à cause de leurs attentes non comblées et de leur perception d'une réalité plus sobre?

Par ailleurs, nul ne peut contester l'affirmation selon laquelle l'informatique a apporté un flot d'idées dans le domaine de la pensée sur la pensée. Par ailleurs, la métaphore cerveau/ordinateur et, mutatis mutandis, la recherche en IA ont encouragé des études sur le fonctionnement du cerveau.

Enfin, nous savons que le débat n'est pas terminé et nous espérons que le présent article contribuera à relancer la réflexion sur l'IA.

## REFERENCES

- Arsac, J. (1987). *Les machines à penser*. Paris: Seuil.  
 Callatay, A. (1986). *Natural and artificial intelligence*. Amsterdam: North-Holland.  
 Dreyfus, H. (1979). *Intelligence artificielle*. Paris: Flammarion.  
 Minsky, M. (1986). *The society of mind*. New York: Simon & Schuster.  
 Penrose, R. (1990). *The emperor's new mind*. Oxford, U.K.: O.U.P.  
 Winograd, T. et Flores, F. (1986). *L'intelligence artificielle en question*. Paris: P.U.F.  
 Winston, P.H. (1984). *Intelligence artificielle*. Paris: InterEditions.

## NOTES ET REFERENCES NUMEROTES

- Pour les besoins de la discussion, nous nous satisferons simplement des définitions du Petit Robert:
  - Conscience: faculté qu'a l'homme de connaître sa propre réalité et de la juger.
  - Intelligence: a) Faculté de connaître, de comprendre (âme, esprit, pensée, raison); b) L'ensemble des fonctions mentales ayant pour objet la connaissance conceptuelle et rationnelle (abstraction, conception, entendement, discernement, jugement, perspicacité, réflexion); c) Aptitude d'un être vivant de s'adapter à des situations nouvelles.
- L'emploi du genre masculin pour désigner des personnes inclut tout autant le féminin que le masculin.
- Minsky, M. (1986). *The Society of Mind*. New York: Simon & Schuster, pp. 163, 186-7, 308, 329. Cet ouvrage résume en quelque sorte ses travaux précédents.
- Turing, A. (1950). Computing Machinery and Intelligence. *Mind LIX*, Vol. 236. Repris in Hofstadter, D.R. et Dennett, D.C. (éd.) (1981). *The Mind's I*. New York: Basic Books, pp. 53-67. Le "Test de Turing" qui a fait couler tant de mots, a été le plus récemment contesté par Penrose, 1990, pp. 34-42.
- Callatay, A. (1986). *Natural and Artificial Intelligence*. Amsterdam: North-Holland, pp. 74, 246, 455.

6. Winograd, T. et Flores, F. (1986). *L'intelligence artificielle en question*. Paris: P.U.F. (Traduction française de 1989). Le chapitre quatre présente les travaux de Maturana, connu pour ses recherches en biologie. Suite à une réflexion critique sur l'informatique et l'IA, ces auteurs désirent aboutir à l'interdiscipline de la philosophie, de la linguistique et de la biologie qui devraient sous-tendre toute conception d'ordinateurs.
7. A l'égard des neurones, Callatay, op. cit., p. 463, définit quantum comme la perturbation produite à l'intérieur du neurone, dans le soma, par une impulsion d'intrant de l'axone. Les quanta peuvent avoir différentes énergies. Agissant sur la même cellule, les quanta sont supposés être indépendants pour certains échanges d'information. Cependant, dans le cas d'un jaillissement d'impulsions ou quand ces jaillissements provenant des dendrites sont produits entre deux "points chauds", ils ont un effet cumulatif non-linéaire.
8. Callatay, op. cit., pp. 34-35.

*Avigdor Farine* est professeur au Département d'études en éducation et d'administration de l'éducation, Faculté des sciences de l'éducation, Université de Montréal.

*Avigdor Farine* is a professor in the Department of Studies in Education and Educational Administration, Faculty of Education, University of Montreal.