



HAL
open science

Intelligence artificielle et singularité

Nicolas P. Rougier

► **To cite this version:**

Nicolas P. Rougier. Intelligence artificielle et singularité. *Hypermondes #01 Robots*, , 2021, 978-2-36183-749-5. hal-03452067

HAL Id: hal-03452067

<https://inria.hal.science/hal-03452067>

Submitted on 26 Nov 2021

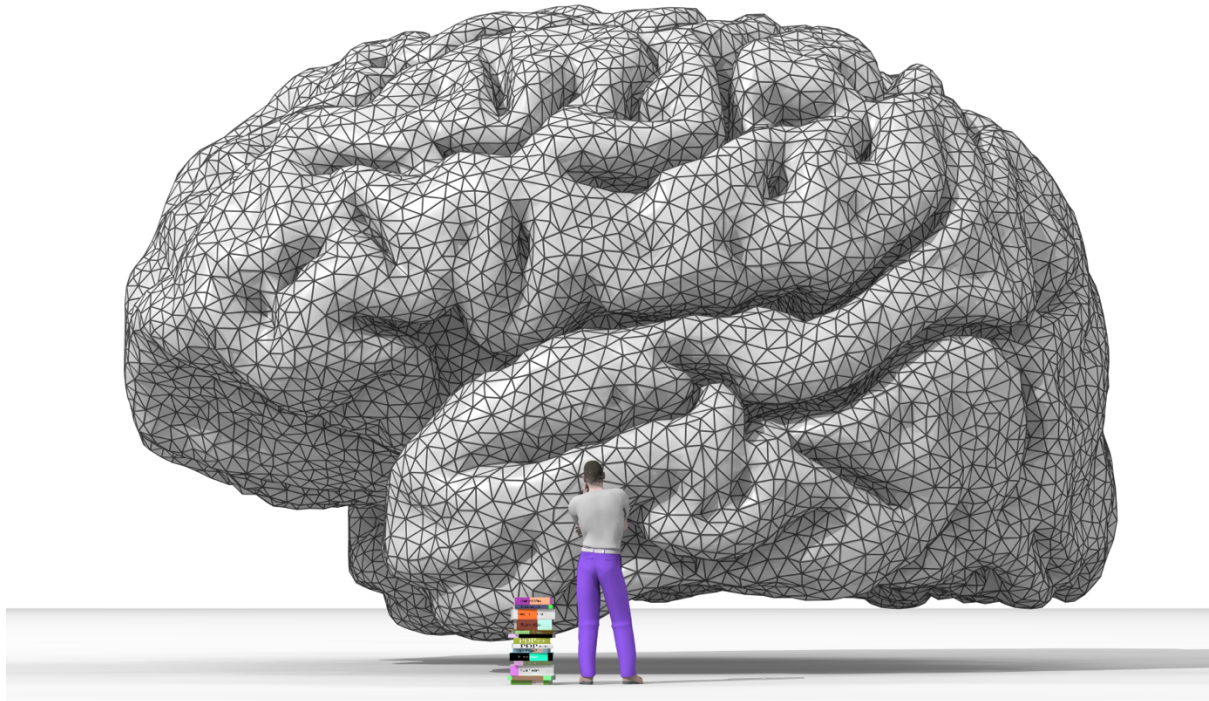
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intelligence artificielle et singularité

Nicolas P. Rougier

Cet article est une version étendue de textes initialement parus sur les sites Interstices et The Conversation.



Les origines. L'intelligence artificielle, ou IA, est officiellement née en 1956 lors d'une conférence à Dartmouth (USA). Même si les recherches ont de fait commencé dès l'après-guerre (par exemple le « Test de Turing » en 1950), cette conférence est historique en ce qu'elle a rassemblé en un même lieu presque toutes les figures emblématiques de l'IA, notamment John McCarthy, mathématicien, Herbert Simon, théoricien des organisations, Allen Newell, mathématicien, Claude Shannon, père de la théorie de l'information et Marvin Minsky, mathématicien. Ces pionniers rêvaient alors de construire des machines pouvant égaler voire surpasser l'intelligence humaine, sur la base des nouveaux moyens de l'informatique naissante. C'était un projet d'une ambition folle, puisque, à cette époque, tout restait à faire. Or, malgré les immenses difficultés qui se trouvaient devant eux, ces chercheurs, aidés d'autres scientifiques, vont rapidement obtenir des résultats qui firent date dans l'histoire de l'informatique.

Ainsi, dès 1955, Newell et Simon conçoivent un programme (*Logic theorist* ou logiciel théoriste en français) qui permet de démontrer automatiquement 38 des 52 théorèmes du traité *Principia Mathematica* d'Alfred North Whitehead et Bertrand Russell (1913). C'est un résultat majeur et extrêmement impressionnant pour l'époque, puisque pour la première fois, une machine est capable de raisonnement. On considère légitimement ce programme comme la toute première intelligence artificielle de l'histoire. Quelques années plus tard, Newell et Simon généralisent cette approche et conçoivent le GPS — *General Problem Solver* (solveur universel) — qui permet de résoudre n'importe quel type de problème, pour peu que l'on puisse le spécifier formellement à la machine. À cette même époque, d'autres chercheurs défrichent les domaines de la traduction automatique, de la robotique, de la théorie des jeux, de la vision, etc. Jusqu'au milieu des années 1970, l'intelligence artificielle est extrêmement prolifique et engendre un nombre considérable de résultats. Mais elle va bientôt être victime de ses folles promesses originelles. En effet, si des avancées extraordinaires ont

été réalisées, les chercheurs sont encore très loin d'avoir résolu le problème général de l'intelligence. Ils se rendent compte qu'ils ont été un peu trop optimistes, à l'exemple de Marvin Minsky qui affirmait en 1970 que « Dans trois à huit ans nous aurons une machine avec l'intelligence générale d'un être humain ordinaire » (citation rapportée par Brad Darrach dans le magazine Life en 1970,). On parle de ce coup d'arrêt comme du premier hiver de l'IA. Mais avec l'arrivée des systèmes experts et la résolution du problème de l'apprentissage dans les réseaux de neurones artificiels, l'IA connaît un second souffle jusqu'à la fin des années 1990 où, de nouveau, tout s'arrête ou presque pour à peu près les mêmes raisons. Le lecteur intéressé pourra se référer au très joli texte de Dominique Cardon, Jean-Philippe Cointet et Antoine Mazières « La revanche des neurones » pour un historique un plus complet.

Deep Blue vs. Kasparov. Il aura fallu attendre 1997, pour enfin lire ces quelques lignes dans le journal « L'Humanité » du 13 mai 1997 :

Kasparov mis en échec par Deep Blue. [...] le super-ordinateur d'IBM a vaincu le champion du monde [...] en gagnant dimanche soir la sixième et dernière partie de leur choc. [...] Le grand maître électronique s'impose 3-1/2 à 2-1/2 face au meilleur des humains et devient surtout le premier ordinateur à l'emporter sur un champion du monde dans une rencontre classique.

Avec presque trente ans de retard, la prophétie de Newell et Simon s'est réalisée (« d'ici dix ans un ordinateur sera le champion du monde des échecs »), l'automate joueur d'échec d'Ambrose Bierce (Le maître de Moxon, 1899) était devenu réel. Les objectifs de l'intelligence artificielle auraient-ils été finalement atteints ? Serions-nous capables de concevoir des machines intelligentes et omnipotentes ?

La réponse est non. S'il est effectivement possible aujourd'hui de battre n'importe quel humain aux échecs, au Go ou au Poker, il reste impossible, par exemple, de faire jouer un robot au football contre un quelconque adversaire humain, quand bien même les scientifiques y travaillent en organisant des compétitions des matches de football entre robots (la Robocup) dont le but à terme est de faire jouer les robots contre des humains. Le football ne peut pourtant prétendre rivaliser en finesse et en intelligence avec le jeu des échecs. Mais comment peut-on battre le champion du monde des échecs et se trouver en même temps incapable de jouer au football ? Cette situation paradoxale s'explique par la diversité d'interprétation de la notion d'intelligence. En effet, l'intelligence que l'on veut prêter à Deep Blue est ici fondée sur le calcul, l'analyse et le raisonnement (courant symbolique de l'IA). Depuis le *Discours de la méthode* de Descartes, le courant de pensée des rationalistes, mené par Descartes, Spinoza et Leibniz, a voulu croire à cette intelligence, une intelligence fondée sur l'esprit et la raison, équivalente à l'intelligence humaine. À l'opposé, le courant de pensée des empiristes (courant connexionniste de l'IA), parmi lesquels Bacon, Locke, Berkeley et Hume, prôna la prise en compte de l'expérience sensible du monde et rejeta l'idée de la connaissance réduite à l'esprit et à la raison.

Ainsi, dès la naissance de l'IA, deux courants de pensée vont coexister plus ou moins pacifiquement. Un premier courant, symbolique, considère la machine comme un système de manipulation de symboles qui peut être utilisé pour manipuler des représentations formelles du monde. Il **est fondé** sur la logique, se faisant ainsi l'héritier des rationalistes. Sa philosophie peut se résumer à la volonté de construire un esprit (« *making a mind* »). Mené par Allen Newell et Herbert Simon, ce courant symbolique stipule que l'intelligence repose sur la notion de symbole. Le deuxième courant, connexionniste, considère en revanche la machine comme un support de la modélisation du cerveau, offrant les capacités nécessaires pour simuler les interactions entre les neurones. Il **s'appuie** sur le domaine des statistiques et sa philosophie peut se résumer à la volonté de modéliser le cerveau (« *modelling the brain* »). Ce courant connexionniste, mené entre autres par Frank Rosenblatt, propose une vision numérique du traitement de l'information et s'oppose en cela à l'hypothèse symboliste.

La victoire de Deep Blue sur Kasparov représente le paroxysme de l'approche symbolique. En effet, le jeu d'échecs est un très beau problème de nature symbolique, où il faut en substance déplacer des pièces sur des cases pour atteindre un but précis (échec et mat). La plus grande difficulté est l'explosion combinatoire, lorsque le programme souhaite prédire des séquences de coups. Cependant,

grâce à des algorithmes très ingénieux, à la force brute, à la rapidité de calcul et une bibliothèque d'ouvertures bien fournie, Deep Blue a pu prédire plus de coups à l'avance que ce qu'il est humainement possible de faire. Deep Blue est-il intelligent au sens commun du terme ? A priori non, c'est simplement un logiciel qui joue aux échecs et ne fait que cela.

Certains n'ont pas été réellement surpris par ce résultat. Ainsi Rodney Brooks écrivait dès 1990 un article au titre un peu curieux, « Les éléphants ne jouent pas aux échecs ». Ce qu'il voulait expliquer alors, c'était que les éléphants ont une certaine intelligence — ils doivent se nourrir, trouver des points d'eau, se reproduire, fuir les prédateurs, etc. — mais n'ont a priori pas besoin de manipuler des symboles abstraits pour le faire. C'est de fait une critique assez sévère du courant symbolique de l'IA, que l'on désigne aussi par le terme GOFAI (« *Good Old Fashion Artificial Intelligence* », la bonne vieille intelligence artificielle). Or, nous savons aujourd'hui que ce courant de l'IA ne pourra pas résoudre le problème de l'intelligence générale à cause du problème de l'ancrage du symbole notamment. En effet, comment mettre en relation les symboles avec les objets qu'ils désignent ? Comment faire le lien entre un verre, le mot, et un verre, l'objet physique que je vois posé sur la table ? Autant cela est évident pour nous humains, autant cela reste un défi pour une machine qui n'a jamais soif et qui n'a pas de corps pour manipuler l'objet. C'est pourquoi cette théorie de la cognition incarnée a émergé dès les années 1990, en mettant en avant le rôle du corps dans la cognition et son intelligence propre.

Un cerveau + un corps. Si vous possédez un animal domestique, par exemple un chien ou un chat, regardez-le attentivement et vous aurez alors un bon aperçu de tout ce qu'on ne sait pas faire en intelligence artificielle... « Mais mon chat ne fait rien de la journée à part dormir, manger et se laver », pourriez-vous me répondre. Et pourtant votre chat sait marcher, courir, sauter (et retomber sur ses pattes), entendre, voir, guetter, apprendre, se cacher, être heureux, être triste, avoir peur, rêver, chasser, se nourrir, se battre, s'enfuir, se reproduire, éduquer ses chatons, et la liste est encore très longue. Chacune de ces actions met en œuvre des processus qui ne sont pas directement de l'intelligence au sens le plus commun mais relève de la cognition et de l'intelligence animale. Tous les animaux ont une cognition qui leur est propre, de l'araignée qui tisse sa toile jusqu'aux chiens guides qui viennent en aide aux personnes. Pour certains, ils peuvent même communiquer avec nous. Pas par la parole, bien entendu, mais en utilisant le langage du corps ainsi que la vocalisation (par exemple des miaulements ou des aboiements). En ce qui concerne votre chat, lorsqu'il vient négligemment se frotter contre vous ou bien qu'il reste assis devant sa gamelle ou devant une porte, le message est assez clair. Il ou elle veut une caresse, a faim ou veut sortir, puis rentrer, puis sortir, puis rentrer... Il a appris à interagir avec vous pour arriver à ses fins. Parmi toutes ces aptitudes cognitives, il n'y en a aujourd'hui qu'une toute petite poignée que l'on commence un peu à savoir reproduire artificiellement. Par exemple la marche bipède. Ça n'a l'air rien de rien et c'est pourtant quelque chose d'extrêmement compliqué à réaliser pour la robotique et il aura fallu de nombreuses décennies de recherche avant de savoir construire et programmer un robot qui marche convenablement sur deux jambes. C'est-à-dire sans tomber à cause d'un petit caillou sous son pied ou lorsqu'une personne l'a simplement effleuré d'un peu trop près. Mais cette complexité existe aussi chez l'homme puisque si vous vous rappelez bien, il nous faut en moyenne une année pour apprendre à marcher. C'est dire la complexité du problème.

Qu'en est-il de la reconnaissance des objets ? S'il est vrai que l'on a vu apparaître ces dernières années des algorithmes capables de nommer le contenu de pratiquement n'importe quelle image, on ne parle pas pour autant d'intelligence ou de cognition. Pour le comprendre, il faut regarder comment ces algorithmes fonctionnent. L'apprentissage supervisé, qui reste aujourd'hui la méthode la plus populaire, consiste à présenter au programme des images ainsi qu'un mot décrivant le contenu de l'image. Le nombre total d'images est généralement bien supérieur au nombre de mots utilisés car pour un même mot, on va associer un très grand nombre d'images représentant l'objet dans différentes situations, sous différents angles de vues, sous différentes lumières, etc. Par exemple, pour reconnaître les chats, on peut présenter jusqu'à un million d'images. En faisant cela, le programme va se constituer une représentation visuelle interne de ce qu'est cet objet, en calculant une sorte de moyenne de l'ensemble des images. Mais cette représentation n'est *in fine* qu'une simple description qui n'est pas ancrée dans la réalité du monde.

En attendant Terminator. Aujourd'hui, en 2020, l'Intelligence Artificielle connaît un second souffle, notamment avec l'avènement du *big data* et une puissance de calcul inégalée jusqu'alors. Ainsi, la redécouverte récente de l'apprentissage profond (*deep learning*), qui date des années 1980, a permis de nouvelles avancées en traitement de données et reconnaissance d'images. Cet algorithme d'apprentissage est un réseau de neurones artificiel qui permet d'apprendre à partir d'exemples. De fait, la différence fondamentale entre les années 1980 et les années 2010 se situe au niveau de la puissance de calcul mise en jeu et de la masse de données qui a atteint des proportions gigantesques. Imaginez un peu que plus d'un milliard de personnes postent leur photo sur Facebook alors qu'il y a trente ans, les chercheurs se démenaient pour constituer des échantillons de quelques dizaines de portraits. De nos jours, ils peuvent donc avoir accès à des centaines de millions de visages. C'est ainsi que Facebook possède un algorithme de reconnaissance faciale qui égale statistiquement les performances humaines. Est-il besoin de préciser qu'on ne parle pas ici d'intelligence, mais bien d'algorithmes et d'apprentissage machine sur des gros volumes de données ?

Diverses interventions dans les médias soulignent les avancées extraordinaires en intelligence artificielle et ses dangers potentiels. S'il existe certainement des dangers liés à ce domaine comme à tout progrès scientifique, les vrais risques de l'IA sont peut-être masqués par la crainte que des machines prennent le pouvoir. Faut-il s'effrayer de l'avènement futur d'une super IA issue de la singularité ? Si cette idée est assez populaire en littérature et au cinéma (« Terminator », « 2001 A Space Odyssey », « Ex-Machina »), elle l'est beaucoup moins dans le domaine scientifique. Ce concept de singularité a été de fait introduit par Vernon Vinge (professeur de mathématiques et auteur de science-fiction) lors du symposium « *Vision 21: Interdisciplinary Science and Engineering in the Era of Cyberspace* » organisé par la NASA en 1993 :

Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended.

Dans trente ans, nous aurons les moyens de créer une intelligence sur-humaine. Peu de temps après, l'ère de l'humanité prendra fin.

Si l'on s'en tient à cette prédiction, nous pouvons en déduire que nous allons bientôt disparaître puisqu'une super IA doit d'ores et déjà exister dans quelques labos secrets. Or, force est de constater que nous sommes encore très loin de ces conclusions dramatiques car il n'y a pas aujourd'hui de machine « super-intelligente ». Pourtant, ce concept de singularité peut séduire car il repose sur une idée somme toute assez simple : si l'on construisait des ordinateurs de plus en plus « intelligents », alors ceux-ci deviendraient plus intelligents que leurs concepteurs et pourraient à leur tour concevoir de nouveaux ordinateurs plus intelligents qu'eux-mêmes. Le moment précis où l'Homme construirait cet ordinateur plus intelligent que lui-même est précisément ce que Vernor Vinge désigne par singularité : puisque les ordinateurs sont beaucoup plus rapides que nous, l'accroissement de l'intelligence se ferait alors de façon exponentielle. Avec les nouveaux résultats époustouflants de ces dernières années, ce concept a été remis au goût du jour, bien qu'il soit bancal (lire à ce sujet le livre de Jean-Gabriel Ganascia « Le Mythe de la singularité »). Pour interpréter les prédictions de Vernor Vinge sur la singularité, il conviendrait de lui faire préciser sa pensée, afin de savoir ce qu'il considère être une intelligence supra-humaine. Par exemple, cette intelligence devrait-elle parler ? avoir des buts propres ? avoir un corps ? avoir des opinions ? des amis ? des émotions ? des rêves ? etc. Et d'ailleurs, comment pourrions-nous reconnaître la singularité, c'est-à-dire, comment pourrions-nous savoir que la machine que nous venons de créer est intelligente ?

Au final, l'intelligence artificielle est un champ de recherche extrêmement fécond, qui a enfanté à son tour de très nombreux champs de recherche, que ce soit en reconnaissance de la parole, en algorithmes pour la génétique, en fouille de données, en vision par ordinateur, en réseaux de neurones artificiels, en apprentissage machine, etc. L'IA « canal historique » a permis de concevoir de très nombreux algorithmes que l'on retrouve aujourd'hui dans un très grand nombre d'applications grand public. Et si aujourd'hui on ne les appelle plus IA, ils en sont pourtant les héritiers directs. C'est

d'ailleurs là une des malédictions de l'IA, car dès qu'un algorithme fonctionne, on a tendance à lui retirer sa filiation. En effet, si une machine sait le faire, c'est qu'il ne s'agit plus d'intelligence !