



HAL
open science

Qu'est ce que l'IA et qu'est ce que ce n'est pas ?

Aurelie Lagarrigue, Thierry Viéville

► **To cite this version:**

Aurelie Lagarrigue, Thierry Viéville. Qu'est ce que l'IA et qu'est ce que ce n'est pas?. 2021. hal-03494002

HAL Id: hal-03494002

<https://inria.hal.science/hal-03494002v1>

Submitted on 18 Dec 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.



Distributed under a Creative Commons Attribution 4.0 International License

Qu'est ce que l'IA et qu'est ce que ce n'est pas ?

Article d'Aurélie Lagarrigue et Thierry Viéville

Publié dans L'intelligence artificielle au service de la lecture LECTURE JEUNE 180 |
DECEMBRE 2021

<http://www.lecturejeunesse.org/livre/lintelligence-artificielle-au-service-de-la-lecture180-decembre-2021/>

Chapô : « L'intelligence artificielle correspond à ce que l'on peut faire faire à une machine et qui eût été intelligent si cela avait été fait par un humain ». En posant cette définition, Marvin Minsky, un des fondateurs de l'intelligence artificielle, positionne le débat exactement au bon endroit : comment nommer un certain nombre de tâches cognitives, c'est-à-dire liées aux grandes fonctions de l'esprit, lorsqu'elles peuvent être déléguées à une machine ?



Aurélie Lagarrigue, ingénieure pédagogique à l'INRIA Learning Lab

Docteure en neurosciences cognitives, Aurélie Lagarrigue s'intéresse aux processus d'apprentissage. Elle a codéveloppé le MOOC Class'Code/INRIA IAI et a participé à son animation.



Thierry Viéville, chercheur en neurosciences

Spécialisé en neurosciences computationnelles appliquées à la compréhension de l'apprentissage humain dans l'éducation, Thierry Viéville participe à des actions de médiation en sciences du numérique, en lien notamment avec l'intelligence artificielle. Il participe au MOOC Class'Code.



La formation citoyenne Class'Code IAI, <https://classcode.fr/iai>, réutilisable, permet à toutes et tous de mieux appréhender l'IA

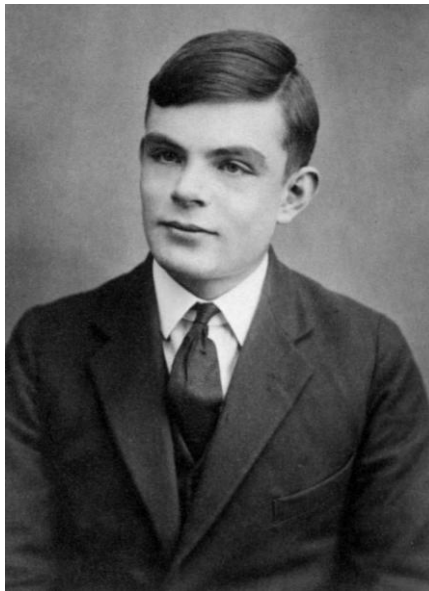
Aux origines de l'intelligence artificielle

Dès le début du XX^e siècle, les chercheurs réfléchissent à la manière de pouvoir automatiser des calculs symboliques et des raisonnements logiques avec deux préoccupations : le faire de la manière la plus rigoureuse et efficace possible. Ils découvrent cependant pendant les années 1930 que quel que soit le mécanisme, certaines questions resteraient indécidables.

Reconnu comme discipline scientifique à partir de 1956, l'IA va alors se développer jusqu'en 1974 sans prendre en compte ces limites, poussant les investisseurs, découragés par le manque de résultat, à se désintéresser de ce secteur jusque dans les années 1980, où sont développés des systèmes experts visant à reproduire des capacités cognitives, sans parvenir néanmoins à répondre aux espoirs placés dans cette nouvelle technologie.

Il faut attendre le milieu des années 1990 pour qu'une nouvelle manière d'appréhender l'intelligence artificielle en bouleverse les performances. On essaye en effet de reproduire ces fonctions cognitives de manière approximative par des calculs numériques massifs, dont les paramètres sont ajustés sur des ensembles de données gigantesques. Les performances atteignent enfin les attentes placées dans l'IA, que ce soit dans la reconnaissance d'images, l'analyse d'imagerie médicale, le traitement en langage naturel ou même les jeux de société : on pense ainsi à la victoire, très symbolique, remportée en 1997 par le programme Deep Blue d'IBM contre le champion d'échecs Garry Kasparov, même si la méthode n'est pas intelligente en soi.

Ces prouesses technologiques, loin d'être infaillibles, n'en bouleversent pas moins notre société. Dans les domaines de la justice, l'IA peut par exemple s'appuyer sur les données de la jurisprudence pour rendre une sentence plus fiable que celle d'un juge qui se fierait à sa seule mémoire, tandis que la conduite d'un train ou d'un geste chirurgical délicat peut être confié à un robot plutôt qu'à un humain. Les champs d'application de l'IA, toujours plus nombreux, nous obligent alors à adopter une attitude technocritique éclairée par une vision citoyenne de la science. En effet, l'extension du domaine de l'IA implique des changements en profondeur : or, il convient de débattre de ces choix de société afin qu'ils puissent être démocratiquement tranchés.



Alan Turing est l'un des fondateurs de la science informatique, qui comprend dès 1936, quels types de traitements de l'information peuvent être effectués par une machine, et quelles en sont les limites. Il imagina, avant les débuts de l'IA, que le fonctionnement d'une telle machine pouvait être confondu avec la cognition humaine. ©wikicommon

L'IA, c'est aussi un jargon à comprendre

Tout au long du développement de l'IA et de sa complexification, il a fallu trouver des termes techniques pour désigner ses modes de fonctionnement spécifiques.

Par exemple, on parle d'**apprentissage machine** pour signifier l'opération consistant à présenter des exemples à l'IA afin de laisser l'algorithme ajuster lui-même ses paramètres pour minimiser l'erreur entre sa prédiction et le résultat attendu. Cela fonctionne un peu comme pour **Le juste prix** ou **Qui est-ce** : une proposition est lancée et en fonction des erreurs, les valeurs sont ajustées pour aboutir au résultat recherché.

On désigne également par le terme d'**apprentissage profond** (ou **deep learning**) le fait d'empiler plusieurs couches de calculs pour travailler de manière plus efficace avec toutes sortes de données.

Enfin, on utilise le terme **big-data** (littéralement **grosses données**) pour désigner des masses énormes de données rendues plus facilement accessibles avec internet. Celles-ci ont cependant une grande valeur et peuvent concerner notamment des informations sensibles, ce qui n'est pas sans problèmes éthiques. Il est par ailleurs important de prendre du recul par rapport aux données car si celles-ci sont biaisées, par exemple parce qu'on aurait dit que tous les chats étaient gris, l'apprentissage le sera également et les résultats seront inutilisables (en l'occurrence par rapport aux chats d'autres couleurs).

Il est alors très étonnant de voir comment l'intelligence est réduite à un simple calcul : c'est un changement radical de la vision que nous avons de l'esprit humain.

Comment fonctionne l'IA ?

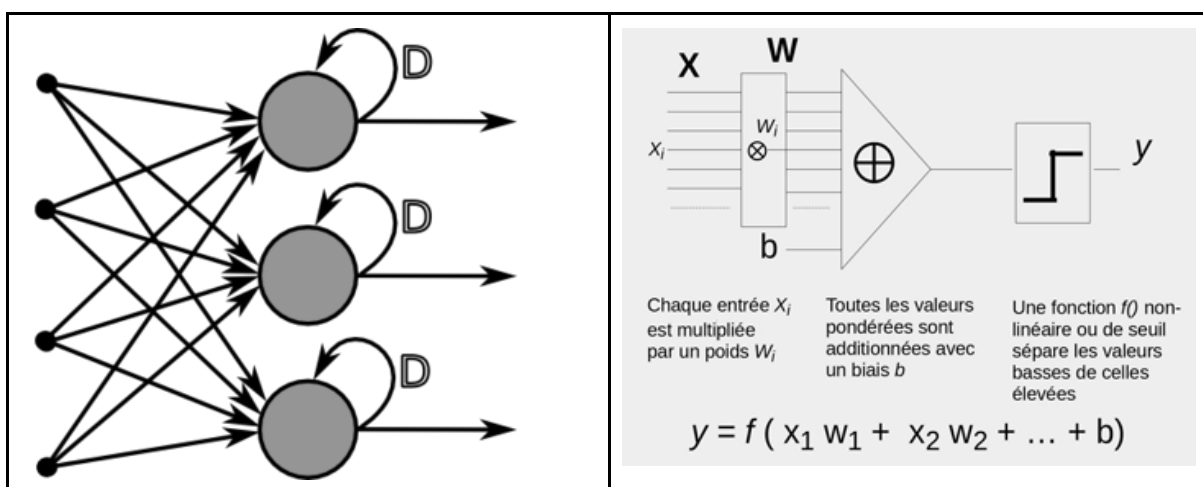
Jusque-là, ces lignes que vous avez lu n'ont pas été tapées au clavier mais... dictées par ordinateur ? Ce dernier comprend-il ce qui lui est dit ? Nullement. Alors comment fait-il ? ! Comprend-il ce que je dis ? Nullement. Alors comment fait-il ?

La parole humaine est constituée d'environ une cinquantaine de phonèmes. Les phonèmes sont des sons élémentaires (entre une voyelle et une syllabe) qui se combinent pour former

des mots. Nous utilisons entre 1 000 et 3 000 mots différents par jour, et en connaissons de l'ordre de 1 0000 en tout, selon notre usage de lecture. Pour un calcul statistique, ce n'est guère élevé : le son de la voix est donc simplement découpé en une séquence de petits éléments qui sont plus ou moins associés à des phonèmes, pour ensuite être associés à des mots. Le calcul statistique cherche, parmi tous les bouts de séquences de mots, celles qui semblent correspondre au son ainsi découpé et qui sont les plus probables. Il ne reste plus qu'à sortir le résultat sous forme d'une chaîne de lettres ou de caractères pour obtenir une phrase à l'oral.

Bien entendu, le mécanisme ne comprend rien à rien. Ces mots ne font pas du tout sens par la machine, puisqu'il ne s'agit que d'une mise en correspondance entre des sons et des symboles. Ce procédé n'est donc exploitable que parce que la voix humaine est moins complexe qu'elle n'y paraît, et surtout car il a été possible de se baser sur une quantité énorme de données (des milliers et des milliers de paroles mises en correspondance avec des milliers et des milliers de bouts de séquences de mots) pour procéder à une restitution fiable du propos dicté. Ces calculs sont à la fois numériques, puisque chaque son est représenté par une valeur numérique manipulée par calcul, et symboliques, chaque phonème ou mot étant un symbole manipulé par un calcul algorithmique.

Suivant un autre exemple concernant la reconnaissance d'images, l'IA va procéder à des calculs élémentaires comme si c'était un **réseau de neurones artificiels** pour distinguer le ciel de la mer. La couleur moyenne ne serait ici pas un critère très discriminant car le ciel est bleu clair et la mer d'un bleu plus profond. Par contre, la position verticale dans l'image est caractéristique, avec le ciel au-dessus et la mer au-dessous. C'est pourquoi l'IA combine deux entrées, l'une plus que l'autre.



Un réseau de neurones n'est qu'une accumulation de calculs faits par des unités organisées entre une couche d'entrée, des couches intermédiaires et des sorties, connectées entre elles.

Là où notre cerveau procède par analogie, un algorithme "raisonne" plus efficacement en énumérant un très grand nombre de cas ou en échantillonnant au hasard pour explorer les possibilités. L'IA ne calcule pas ailleurs pas comme un humain. Elle ne multiplie pas comme nous suivant un système décimal, mais en recourant au langage binaire qui correspond à des câblages électroniques dont la sortie correspond au résultat attendu. Ainsi, les algorithmes qui différencient, par exemple, un chat d'un chien dans une image, ne s'appuient pas sur des caractéristiques physiologiques, mais effectuent des comparaisons statistiques avec des images de références, à la suite de transformations numériques qui sont la plupart du temps peu interprétables.

Il en résulte qu'aucun de ces algorithmes ne peut fonctionner sans erreur compte tenu de son mode même de fonctionnement, et ce quelle que soit l'ampleur de l'apprentissage. L'IA sera toujours confrontée à des cas faisant exception et à des configurations inattendues qui le mettront en échec.

Plus la requête est spécifique à résoudre, plus l'algorithme pourra être efficace car il pourra ajuster son calcul en fonction d'une tâche très précise. A contrario, un résolveur de problèmes qui serait généraliste ne pourrait pas, en théorie, être performant. Super efficaces mais sans aucune intelligence, uniquement dédiés à des tâches spécifiques, et toujours faillibles : voilà ce qui caractérise les mécanismes dits d'intelligence artificielle, loin des images fantasmées par notre imaginaire (v. art. et itw Jérémie Dres).



an airplane is parked on the tarmac at an airport

a group of people standing on top of a beach

Figure reprise du travail de Lake et al 2017 (<https://arxiv.org/abs/1604.00289>) qui étudie dans quelle mesure l'apprentissage machine peut se comparer à l'apprentissage humain : ici on voit des interprétations tout à fait "plausibles" mais complètement absurdes, car il manque au mécanisme à la fois une compréhension générale du fonctionnement du monde et des informations de contexte.