

APPRENDRE
DEMAIN

Magaly Alonzo
Alexia Audevert

APPRENDRE DEMAIN

Quand
intelligence artificielle
et neurosciences
revolutionnent
l'apprentissage

DUNOD

Couverture : Valérie Le Roux

Mise en pages : Nord Compo

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, 2019 pour la version française
11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-079880-3

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

À tous les data & brain heroes

Sommaire

Chapitre 1. Une histoire pas si récente	9
Quand la machine défie l'homme : go	9
Un peu d'histoire.....	10
Les neurosciences modernes.....	12
Les deux mouvements de l'intelligence artificielle	14
L'intelligence artificielle symbolique fondée sur la logique.....	16
L'intelligence artificielle connexionniste reposant sur les données	18
Qu'est-ce qu'apprendre ?	19
Qu'est-ce que l'intelligence ?	21
Les limites du bio inspiré	24
Chapitre 2. Comment apprend une machine ?.....	27
Des algorithmes à l'informatique	27
De l'informatique traditionnelle à l'apprentissage automatique	28
Les domaines d'application du <i>machine learning</i>	30
Les données	32
Les différents types d'apprentissage.....	33
Apprendre à partir de données	40
Les 5 courants du <i>machine learning</i>	46
Du neurone biologique au neurone artificiel.....	51
Quand les neurones se connectent	56
Apprendre en corrigeant ses erreurs.....	58
Apprentissage profond ou <i>deep learning</i>	63
Les réseaux de neurones à convolution	67
Les réseaux de neurones récurrents.....	71
Les réseaux de neurones génératifs	74

Chapitre 3. Comment apprend un humain ?	79
Mémoire non déclarative	80
Mémoire déclarative	88
La mémoire du futur	110
Facteurs d'influence.....	112
Chapitre 4. Des modèles d'apprentissage pas si différents	121
Des modèles pour apprendre.....	122
La configuration des modèles.....	125
L'ajustement des modèles	136
La minimisation des erreurs.....	140
Transfert de compétences.....	151
De l'information brute au concept.....	155
Apprendre, c'est généraliser	163
La résilience	165
Chapitre 5. Neurosciences fiction	171
Pourquoi les machines ne nous ont-elles pas encore égalés ?	171
Qu'en est-il de l'éthique dans le monde de demain ?	194
Quels changements sont induits par l'intelligence artificielle et les neurosciences ?	203
Vers une symbiose intelligence artificielle neurosciences ?	214
Conclusion	219

Chapitre 1

Une histoire pas si récente

Quand la machine défie l'homme : go

En mars 2016 avait lieu un événement majeur avec la confrontation d'un humain et d'une machine. Difficile de ne pas en avoir entendu parler, l'intelligence artificielle nommée AlphaGo¹, développée par la start-up DeepMind Technologies (devenue depuis filiale de Google et appelée Google DeepMind), affrontait le meilleur joueur du monde de go du milieu des années 2000, le Sud-Coréen Lee Sedol.

Les règles du jeu de go sont assez simples et peuvent se résumer en seulement 9 axiomes. En revanche, toute la complexité réside dans le nombre de combinaisons possibles qui excède le nombre d'atomes connus dans l'univers.

Les technologies classiques informatiques ne permettent pas de coder un algorithme formel pour modéliser une partie de go. AlphaGo a été conçue en utilisant deux types de techniques d'apprentissage automatique permettant à la machine d'acquérir des capacités cognitives : la première, inspirée de réseaux neuronaux, appelée *deep learning* et la seconde, se fondant sur des parcours de graphes. AlphaGo a été entraînée à la fois en jouant contre des êtres humains, contre d'autres ordinateurs et contre elle-même. Ainsi, cette intelligence a pu non seulement bénéficier de l'expérience cumulée de toutes les personnes

1. AlphaGo de DeepMind : <https://deepmind.com/research/alphago/>

rencontrées mais également de celle d'autres ordinateurs, jouant encore et encore des parties sans risquer l'épuisement.

La confrontation a eu lieu en 5 manches avec un score de 4 victoires à 1 pour AlphaGo.

Lee Sedol réussit cependant à remporter la 4^e manche en jouant un coup non envisagé par AlphaGo (probabilité inférieure à 1 chance sur 1 000). Cette attaque déstabilisa AlphaGo qui enchaîna des coups incohérents et mena à la victoire de Lee Sedol qui déclara : « Cette victoire n'a pas de prix ! ».

Toute l'Asie a suivi très attentivement ces rencontres diffusées en direct sur Internet et était stupéfaite de voir qu'une machine pouvait battre un tel champion. L'homme était-il en train de se faire dominer par une machine ? La réponse est oui, mais seulement au jeu de go.

Lors de ces échanges, Lee Sedol a montré une palette d'émotions et d'expressions dont seuls peuvent être dotés certains êtres vivants telles que la confiance, l'anxiété, la peur, la joie, la déception, etc. La machine, quant à elle, est restée « concentrée » sur la tâche qu'elle devait effectuer à savoir jouer et battre son adversaire au jeu de go.

Mais comment ces machines apprennent-elles ? Comment ces algorithmes fonctionnent-ils ? Qu'ont-ils à voir avec les neurones qui constituent notre cerveau ? Comment comparer ces deux types de réseaux ? La machine est-elle réellement une imitation de ce qui existe dans la nature ?

Un peu d'histoire

L'histoire des neurosciences est indissociable à la fois de l'histoire de la médecine et de celle de la philosophie. Aux prémices de la première discipline, le cerveau était considéré comme inutile. Pour preuve, il était ignoré, voire retiré, au cours des autopsies et autres crémations. On estimait alors que le siège de l'intelligence se situait au niveau du cœur. Les Égyptiens furent les premiers à se pencher sur la question du cerveau, à partir du xvii^e siècle av. J.-C.

Il faudra ensuite attendre la révolution scientifique et Hippocrate notamment, au iv^e siècle av. J.-C., qui fait du cerveau le centre des

sensations et le lieu de l'intelligence. Toutefois, dans le même temps, Aristote met le cœur au centre du corps et ne reconnaît pratiquement aucune fonction au cerveau dans sa conception de la sensation et de l'intelligence et ne lui octroie qu'un rôle de refroidissement du système sanguin. C'est paradoxalement par Aristote que neurosciences et intelligence artificielle sont liées pour la première fois. En effet, la logique aristotélicienne, et plus précisément la syllogistique, est à la base du raisonnement appliqué aux machines. Cette logique est illustrée par la célèbre maxime : « Tous les hommes sont mortels, Socrate est un homme, donc Socrate est mortel ».

Un siècle plus tard, les philosophes, avec Platon comme figure de proue, séparent l'âme en trois parties dont l'une, immortelle, se situe dans la tête.

Au ^{xiv}^e siècle, l'anatomiste flamand André Vésale réalise un remarquable travail sur le corps humain. Il étudie les nerfs, les dissèque. Il vient ainsi à la conclusion que ces derniers constituent le liant entre l'esprit, les émotions et le cerveau et les définit comme moteur de la transmission des émotions et des sensations.

Au ^{xvii}^e siècle, René Descartes aborde la notion de machine (et d'intelligence artificielle) dans la cinquième partie du très célèbre *Discours de la méthode* :

« Bien qu'elles fissent plusieurs choses aussi bien ou peut-être mieux qu'aucun de nous, elles manqueraient infailliblement en quelques autres en lesquelles on découvrirait qu'elles n'agissent pas par connaissance, mais seulement par la disposition de leurs organes. »

Également très intéressé par la question du dualisme du corps et de l'esprit, Descartes lie ce dernier à des structures du cerveau.

Enfin, les neurosciences ont émergé comme une branche de la biologie et de la médecine. Philosophiquement inspirée par le scientisme du ^{xix}^e siècle, cette nouvelle discipline postule l'absence de toute cause endogène (auto-générée) du comportement humain. Elle est alors portée par les progrès technologiques qui lui sont contemporains, à l'instar de l'électricité, qui permet au physicien et médecin Luigi Galvani d'étudier la conduction nerveuse.

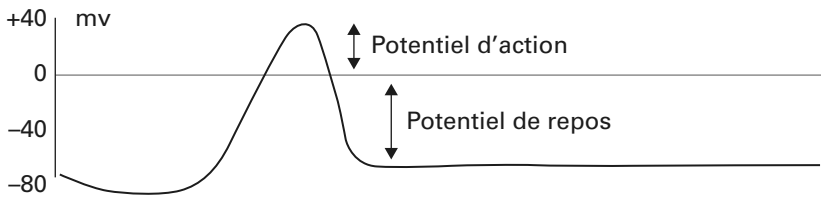


FIGURE 1.1 : POTENTIEL D'ACTION ET POTENTIEL DE REPOS

Lorsque le courant atteint une certaine valeur au sein d'un neurone, un mécanisme nommé potentiel d'action va provoquer son activation. Le neurone va, à son tour, pouvoir stimuler ou inhiber un ou plusieurs neurones.

Les neurosciences modernes

Au-delà des progrès techniques et de ceux de la psychologie, le xx^e siècle est également celui de l'avènement des sciences cognitives. Alliant le corps et l'esprit, ces disciplines ont pour but de comprendre les fondements biologiques qui sous-tendent les fonctions du cerveau. Elles s'intéressent donc à des sujets aussi variés que l'attention, la prise de décision, l'apprentissage, la mémoire, le langage, les émotions et la perception.

Dans ce contexte, le physiologiste russe Ivan Pavlov étudie une forme archaïque d'apprentissage qu'est le conditionnement. De son côté, le neurochirurgien canadien Wilder Penfield fait considérablement évoluer les connaissances sur les localisations cérébrales et synthétise ses expériences sous la forme d'une sorte de cartographie du cortex moteur et sensoriel, l'homonculus.

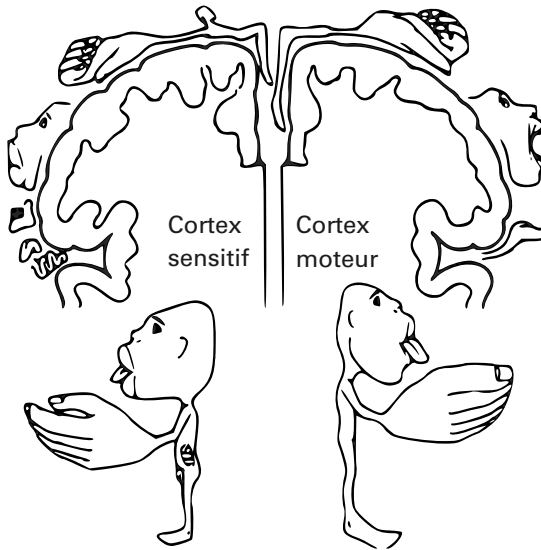


FIGURE 1.2 : HOMONCULUS DE PENFIELD

De son côté, le neuroanatomiste américain James W. Papez développe sa théorie viscérale des émotions et définit le circuit de Papez comme étant le siège du contrôle des émotions et en lien étroit avec des structures cérébrales liées à l'apprentissage.

La seconde partie du xx^e siècle a vu des progrès extraordinaires grâce à l'étude de personnes victimes de lésions cérébrales. Des patients sont ainsi devenus célèbres dans le monde des neurosciences. Le patient nommé « H.M. », étudié par la neuropsychologue canadienne Brenda Milner, est l'un d'entre eux. H.M. est devenu incapable de former de nouveaux souvenirs à la suite de l'ablation d'une partie de son cerveau à l'origine de crises d'épilepsie sévères. L'étude des lésions provoquées dans son cerveau a montré que cette nouvelle infirmité était le fait de l'ablation d'une structure nommée hippocampe. Cette structure est aujourd'hui au cœur de l'attention des chercheurs.

L'étude des mécanismes sous-tendant la mémoire a mené à l'émergence de plusieurs courants de pensée. En effet, s'il y a bien une question pour laquelle il est difficile de trouver une réponse qui satisfasse la communauté neuroscientifique, c'est bien celle des théories de la mémoire ou des mémoires. Neurons dits grand-mère, neurones miroirs ou

boucles de rétroaction sont autant de théories qui ont émergé pour expliquer son fonctionnement et qui n'ont toujours pas réussi à rassembler la communauté scientifique.

Ce raisonnement nous amène à entrevoir que l'intelligence artificielle n'est pas une imitation du cerveau humain, bien qu'inspirée de celui-ci, pour la simple raison que nous ne savons pas exactement comment fonctionne notre cerveau. En effet, si les neurosciences modernes ont fait d'énormes progrès et en font encore chaque jour, nous n'avons pas une compréhension complète des mécanismes qui expliquent le fonctionnement de cette masse logée dans notre boîte crânienne.

Les techniques modernes nous ont certes permis de mieux comprendre le fonctionnement de nos neurones, à l'échelle cellulaire et moléculaire, lors de la transmission de l'information. Des études en IRM (Imagerie par Résonance Magnétique) fonctionnelle notamment nous permettent d'appréhender toujours mieux le rôle des différentes aires cérébrales autant dans des conditions normales que pathologiques. Enfin, l'enregistrement de l'activité électrique du cerveau par des techniques dites électroencéphalographiques permet de faire le lien entre les connaissances cellulaires et macroscopiques ainsi accumulées. Malgré tout, la carte dressée par les chercheurs est criblée de trous. Comme souvent, des résultats expérimentaux peuvent se contredire, provoquant l'émergence de théories difficiles à vérifier par les moyens techniques actuels. Ces verrous seront levés à plus ou moins longue échéance grâce aux progrès techniques et à l'effort déployé par les chercheurs. Les équipes de recherche actuelles mettent un point d'honneur à favoriser la pluridisciplinarité incluant statisticiens et *data scientists*, qui travaillent de concert avec les neuroscientifiques afin de réaliser des modèles mathématiques du cerveau.

Les deux mouvements de l'intelligence artificielle

Le terme d'« intelligence artificielle » est apparu il y a bientôt 70 ans. La notion a connu différentes périodes d'évolution et a vu émerger deux types d'approches : l'approche symboliste, fondée sur la logique, et l'approche connexionniste, fondée sur les données.

Durant l'été 1956, quelques scientifiques venant de domaines différents (mathématiques, psychologie, ingénierie, économie, science politique, etc.) se sont retrouvés lors du congrès de Darmouth avec un objectif commun : « re-crée l'intelligence humaine dans une machine ». Pour Marvin Minsky, John McCarthy, Claude Shannon, Nathaniel Rochester et d'autres scientifiques, l'avènement de cerveaux égalant l'homme était imminent. L'intelligence artificielle a ainsi été reconnue comme discipline académique, soit une véritable science, en 1956².

Les vingt années qui vont suivre constituent le premier âge d'or de l'intelligence artificielle. Des millions de dollars sont investis. L'ARPA, qui deviendra par la suite la DARPA³, (l'agence du département de la Défense des États-Unis chargée de la recherche et du développement des nouvelles technologies) y consacre 3 millions de dollars par an. De nombreuses prévisions très optimistes quant à cette nouvelle science sont énoncées, comme celle de Herbert Simon en 1965, « Des machines seront capables, d'ici vingt ans, de faire tout travail que l'homme peut faire » ou encore celle de Minsky en 1967, « Dans une génération [...], le problème de la création d'une "intelligence artificielle" [sera] en grande partie résolu ».

Dans les années 1970, l'intelligence artificielle subit de nombreuses critiques et des revers budgétaires. L'optimisme des chercheurs a en effet engendré une attente excessive et le constat que les promesses semblent irréalisables – les ordinateurs étaient alors trop primitifs, avec des puissances trop limitées pour réaliser des calculs complexes – fait que les subventions se tarissent.

S'ensuivent une phase de renaissance avec l'approche symbolique puis un nouvel hiver aux alentours des années 1990, et enfin un nouvel essor au début du XXI^e siècle avec l'approche connexionniste.

2. « A proposal for the Darmouth summer research project on artificial intelligence », en ligne : <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>

3. DARPA : *Defense Advanced Research Projects Agency*.

L'intelligence artificielle symbolique fondée sur la logique

L'intelligence artificielle symbolique consiste à automatiser l'expertise humaine dans une machine en utilisant des techniques fondées sur un raisonnement formel. L'implémentation la plus reconnue de cette approche est le système expert qui est composé de deux parties : une base de connaissances et un moteur d'inférence.

La première partie consiste à modéliser l'expertise humaine, ou plus précisément la connaissance d'un domaine communiquée préalablement et consciencieusement par des experts. Cette expertise est stockée dans une *base de connaissances* constituée d'une base de faits et d'une base de règles.

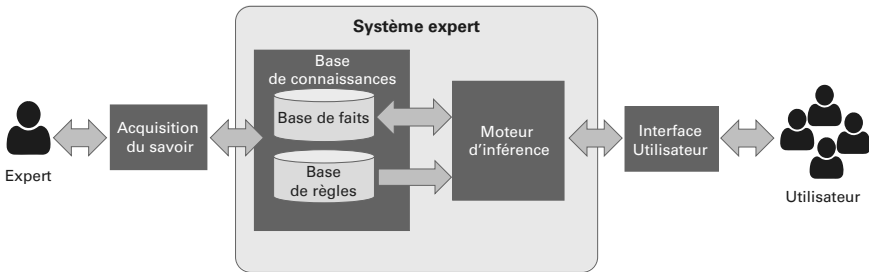


FIGURE 1.3 : FONCTIONNEMENT D'UN SYSTÈME EXPERT

La *base de faits* est un ensemble de connaissances considéré comme vrai qui a été initialement exprimé par l'expert ou déduit par le système. La connaissance est représentée par des objets et des symboles formels s'associant entre eux (est, appartient à, fait partie de, équivaut à, etc.).

Modélisons des relations familiales dans une base de faits avec les faits suivants :

- « Théo est le frère de Lucas »
- « Léane est la sœur de Théo »
- « Magali est la maman de Théo ».

La *base de règles* définit la connaissance sous forme de règles logiques. Les règles suivent cette topologie : « Si X, alors Y », où Y est un nouveau fait pouvant être une conclusion ou une déduction.

Reprenons notre exemple précédent et ajoutons les règles suivantes :

- « Si X est frère de Y, alors la maman de X est aussi celle de Y »
- « Si X est sœur de Y, alors la maman de X est aussi celle de Y ».

La seconde partie d'un système expert est le *moteur d'inférence* qui permet de simuler un mécanisme de raisonnements pour obtenir des déductions et des conclusions.

En s'appuyant sur la base de connaissances précédente, le système expert pourra en déduire que :

- « Magali est la maman de Lucas »
- « Magali est la maman de Léane ».

À partir des faits, le moteur d'inférence applique les règles afin d'en déduire d'autres faits pour résoudre un problème de logique. Ces faits déduits peuvent également enrichir la base de faits initiale.

L'un des systèmes experts les plus connus parmi les systèmes grand public est Akinator⁴. Akinator est un jeu gratuit sur Internet, capable de deviner à qui ou à quoi vous pensez en vous posant une série de questions.

Dans les années 1980, cette approche par les systèmes experts a connu son apogée notamment avec l'utilisation du langage Prolog créé par deux Français (Alain Colmerauer et Philippe Roussel) et du langage LISP. La victoire d'IBM Deep Blue contre le champion d'échecs russe Garry Kasparov en 1997 ou la victoire d'IBM Watson dans le jeu télévisé Jeopardy en 2011 ont été permises par des techniques d'intelligence artificielle symbolique.

Les deux principaux inconvénients de cette approche sont toutefois une modélisation de la connaissance en grande partie manuelle et son manque de flexibilité : le système a ainsi une capacité très limitée à réagir, dès lors qu'un événement ou une situation n'ont pas été prévus.

Les systèmes experts peuvent être vus comme des programmes informatiques traditionnels avec des algorithmes qui se programment manuellement. Bien que cette technique soit de nos jours beaucoup moins médiatisée que l'approche connexionniste, cette approche fondée sur de la logique formelle et non pas sur les probabilités reste fondamentale pour un grand nombre de systèmes.

4. Système expert Akinator : <https://fr.akinator.com/>

L'intelligence artificielle connexionniste reposant sur les données

En opposition à l'approche symbolique, l'approche connexionniste est incarnée par les réseaux de neurones et le *deep learning* (apprentissage profond). Elle s'inspire d'une modélisation mathématique partielle du cerveau humain et des neurones biologiques afin de reproduire certaines capacités, comme la mémoire, l'apprentissage ou le traitement d'informations incomplètes.

Bien que la recherche fondamentale ait commencé il y a plus d'une soixantaine d'années avec Franck Rosenblatt et ses perceptrons (base des réseaux de neurones toujours utilisées de nos jours), les limitations techniques n'ont pas permis à cette technique de se démocratiser avant le début du *xxi*^e siècle. En effet, depuis une décennie, l'augmentation des capacités de calcul des ordinateurs et la possibilité de paralléliser plus simplement les traitements, couplées à une importante volumétrie de données qualifiées et d'algorithmes performants, permettent de réaliser différentes tâches comme l'émulation des fonctions de la vision, le traitement naturel du langage, l'analyse de données avancées, etc.

Cette approche utilise le principe de biomimétisme des neurones biologiques et se fonde sur des méthodes d'apprentissage empiriques basées sur l'observation (données) et des méthodes statistiques. Par opposition à l'approche symbolique, très formelle et donc facilement interprétable, documentable et prouvable, l'approche connexionniste est empirique et probabiliste. L'explicabilité des algorithmes peut se révéler difficile. Le *deep learning* est ainsi souvent assimilé à une boîte noire. Mais de nombreux travaux sur ce sujet sont en cours, comme le projet XAI⁵ mené par la DARPA qui en a fait une de ses priorités.

En résumé, les systèmes fondés sur l'approche connexionniste sont capables d'apprendre par eux-mêmes, et les programmes s'éduquent plus qu'ils ne se programment.

5. Explainable AI (XAI) mené par la DARPA : <https://www.darpa.mil/program/explainable-artificial-intelligence>