

---

# ANATOMIE D'UNE BULLE

THOMAS+ANTOINE GERBAUD

---

# RÉSUMÉ

L'intelligence artificielle (IA) est un domaine cyclique.

Les bulles d'intérêt sont classiques. On les qualifie de "printemps" de l'IA, par opposition aux hivers.

Le troisième printemps de l'IA a commencé en 2016. Il concerne l'utilisation d'une classe d'algorithmes qui relève de l'apprentissage automatique profond (*deep learning*). C'est un sous-domaine du domaine plus général de l'intelligence artificielle.

Les algorithmes de *deep learning* permettent de construire des outils d'aide à la décision à partir de l'exploitation statistique de vastes bases de données. Ces bases de données ont typiquement des millions ou des milliards d'objets, voire plus.

Le *deep learning* est très utilisé par les BigTechs, GAFA en tête. Ils sont responsables de progrès spectaculaires en traitement d'images, en classification (produit, service, utilisateur) et en recommandation (article, texte).

Fin 2022, OpenAI a donné accès à un agent conversationnel nommé ChatGPT. Une partie de son fonctionnement reste secret, voire obscur. Il est construit sur un modèle de langage (LLM) qui exploite de très vastes bases de données textuelles (corpus). Les performances de ChatGPT sont remarquables.

Le *deep learning* est une avancée technologique intéressante qui permet de construire des outils d'aide à la décision aux performances excellentes - dans certains cas. Cette technologie n'est cependant ni révolutionnaire, ni incoutournable, ni universelle.

Le propos de cet document est d'inscrire ce troisième printemps, la bulle du *deep learning*, dans une perspective historique plus générale. Et de donner le point de vue informé, et parfois critique, de deux *data scientists* expérimentés sur la frénésie qui s'est emparée du monde économique.

Rédaction : hiver 2022 - printemps 2023



Essayons de converser sans nous exalter,  
puisque nous sommes incapables de nous  
taire.

Samuel Beckett (1948)

L'ensemble des acteurs privés et publics rêve d'adopter le modèle opérationnel des BigTechs : déléguer la prise de décision à des algorithmes opaques qui exploitent systématiquement de larges bases de données. Stratégie, gouvernance et innovation sont concernés. La décennie 2010-2020 voit l'économie succomber à cette mode et s'exprimer uniquement par mot-clés : tech, *data mining*, Big Data, *data science* et maintenant *deep learning*.

Cette expression percutante désigne une technique d'apprentissage automatique très spécifique : les réseaux de neurones profonds. Ces technologies largement incomprises accaparent les budgets consacrés à l'innovation. Sans prouver leur pertinence.

Cette situation est déraisonnable mais, heureusement, temporaire. En effet, l'histoire de l'intelligence artificielle alterne entre périodes de surexcitation et de dépression.

Dans ce domaine, les hivers succèdent aux printemps ; que retiendra-t-on du printemps de l'apprentissage profond ?

---

# INTRO DUCTION

L'intelligence artificielle est un domaine cyclique, dans lequel les hivers succèdent aux printemps. Depuis 2016, nous sommes dans le troisième printemps, celui de l'apprentissage profond (*deep learning*). Préparons-nous maintenant aux premières gelées.

Ce printemps, très médiatique, est une conséquence de l'ouverture du web au grand public, commencé en 1993-1994. Les principaux acteurs du web sont représentés par les GAFAM nord-américains et leurs pendants chinois (BATX). Dans leur sillage, de nombreuses sociétés se sont spécialisées dans le traitement des données en ligne. Ce modèle d'entreprise est celui des BigTechs. Il est très rentable pour les rares sociétés qui percent dans le domaine.

Ces BigTechs, GAFA en tête, ont dû résoudre de nouveaux problèmes opérationnels rendus complexes par la masse grandissante des données. Les techniques d'apprentissage automatique, dont font partie les réseaux de neurones, sont des outils adaptés. Ils ont engendré des progrès spectaculaires en traitement d'images, en classification (produit, service, utilisateur) et en recommandation (article, texte).

Tout s'accélère entre 2014 et 2016. Yann LeCun, un des principaux acteurs académiques des réseaux de neurones profonds, est recruté par Facebook. DeepMind est acheté par Google en 2014 ; son programme, AlphaGo, écrase un champion humain au go (AlphaGo, 2016), vingt ans après la victoire d'IBM aux échecs (DeepBlue, 1997). Les médias s'emballent, la bulle économique enfle, le printemps fleurit.

La sphère économique butine les premières fleurs. Entreprises, industries, pouvoirs publics se mettent à rêver de lendemains qui chantent. L'innovation en informatique décisionnelle et en analyse de données passe obligatoirement par ces fameux réseaux de neurones profonds.

Mais comment croire sérieusement que les solutions adaptées aux problématiques extrêmement spécifiques des GAFAM se déclinent à l'ensemble des acteurs économiques ?

Les voies dissonantes sont rares dans le concert de louanges permanent. Les fameux réseaux de neurones sont maintenant identifiés à l'intelligence artificielle pour le grand public. Cette situation déraisonnable dure depuis huit ans. Watson (IBM) a échoué à révolutionner la médecine, la voiture automatique est reportée d'année en année depuis 2014, ChatGPT d'OpenAI ne comprend rien des phrases qu'il assemble. Le roi est nu - il lui reste une plume, peut-être.

---

# PLAN

Ce texte est organisé en deux parties.

Nous présentons dans un premier temps une synthèse commentée de l'histoire de l'intelligence artificielle. « Quand le passé n'éclaire plus l'avenir, l'esprit marche dans les ténèbres », disait Tocqueville. L'éclairage de ce domaine, à mi-chemin entre l'ingénierie et la recherche, s'étend sur près d'un siècle, depuis ses débuts en 1930. Notre but est de donner au lecteur une vision générale des cycles qui ont façonné l'histoire chahutée de l'intelligence artificielle.

Le second temps est une lecture critique, technique et épistémologique de la bulle actuelle, conclue par quelques recommandations et perspectives.

---

<b>7</b>	<b>100 ANS</b>
<b>8</b>	<b>CALCULER RAPIDEMENT</b>
<b>11</b>	<b>LE PRINTEMPS SYMBOLIQUE</b>
<b>18</b>	<b>L'HIVER SYMBOLIQUE</b>
<b>22</b>	<b>LA CONNAISSANCE</b>
<b>26</b>	<b>INTÉGRATION PROGRESSIVE</b>
<b>30</b>	<b>DEUX RÉUSSITES</b>
<b>33</b>	<b>L'APPRENTISSAGE AUTOMATIQUE</b>
<b>36</b>	<b>LES RÉSEAUX DE NEURONES</b>
<b>45</b>	<b>3 HISTOIRES</b>
<b>47</b>	<b>DEEP BUBBLE</b>
<b>48</b>	<b>EN 2014</b>
<b>51</b>	<b>LIMITES</b>
<b>55</b>	<b>UN ÉCHEC</b>
<b>58</b>	<b>WINTER IS COMING</b>
<b>62</b>	<b>RÉFÉRENCES</b>
<b>63</b>	<b>QUI ?</b>

---

# 100 ANS



L'intelligence artificielle a bientôt 100 ans. Mouvement intellectuel et théorique dans un premier temps, l'IA est devenue une activité scientifique et technique avec la construction des premiers ordinateurs. La littérature abordait directement le sujet des machines pensantes depuis la fin du XIXème siècle. Ces nouvelles machines permettait de calculer rapidement, et peut-être de simuler le raisonnement, voire la pensée humaine : que pouvait en dire la science ?

---

# CALCULER RAPIDEMENT

ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO  
THE ENTSCHIEDUNGSPROBLEM

By A. M. TURING.

[Received 28 May, 1936.—Read 12 November, 1936.]

The “computable” numbers may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means. Although the subject of this paper is ostensibly the computable *numbers*, it is almost equally easy to define and investigate computable functions of an integral variable or a real or computable variable, computable predicates, and so forth. The fundamental problems involved are, however, the same in each case, and I have chosen the computable numbers for explicit treatment as involving the least cumbersome technique. I hope shortly to give an account of the relations of the computable numbers,

## LES MACHINES DE TURING

L'histoire aurait pu commencer au XIX<sup>ème</sup> avec Ada Lovelace et Charles Babbage, avec la créature du Dr Frankenstein de Mary Shelley, avec le Golem du folklore juif, avec les automates fabriqués par Héphaïstos.

Nous la faisons commencer à l'ère moderne, dans les années 1930, lorsque le mathématicien de génie Alan Turing invente une machine conceptuelle pour résoudre un problème mathématique. Est-il possible de décider si un énoncé logique, ou un algorithme, a un sens ? Il s'agit du problème de la décision. Turing reformule ce problème dans un cadre nouveau, en s'inspirant de nombreux scientifiques dont le logicien K. Gödel. Cette machine, dite de Turing, va pouvoir exécuter cet « algorithme ». Il prouve ainsi que ce problème est équivalent à imaginer une méthode universelle qui décidera si un algorithme exécuté par cette machine s'arrêtera, ou pas. C'est le problème de l'arrêt. La réponse est négative.

Alan Turing cristallise ces idées et façonne le concept d'un calculateur universel : l'ordinateur. Il montre dans le même mouvement que ce calculateur, si universel soit-il, ne peut pas résoudre tous les problèmes. L'ironie est savoureuse. Ces travaux fondateurs ne sortent toutefois pas d'un cadre très confidentiel. La guerre à venir contre l'Allemagne n'améliorera pas la situation.

## LES CALCULATEURS COMME OUTILS STRATÉGIQUES

Les débuts de l'informatique sont foisonnants. De nombreux ingénieurs et scientifiques inventent des machines et les assemblent en Europe et aux Etats-Unis. Avant la Seconde Guerre mondiale, quelques calculateurs automatiques existent. Ces machines réalisent un petit nombre d'opérations de base par seconde. Additions et multiplications sont en binaire ou en décimal.

Les usages sont principalement militaires. En Pologne, la Bomba (1938) de M. Rejewski est utilisée pour des calculs de cryptanalyse. En Allemagne, les Z2 et Z3 (1940) de K. Zuse sont utilisés pour des calculs d'aérodynamique. Ces calculateurs électromécaniques sont les ancêtres des ordinateurs que nous connaissons aujourd'hui. Au Bletchley Park en Angleterre, Alan Turing et Gordon Welchman conçoivent The Bomb en 1940, un calculateur capable de décrypter le code Enigma utilisé par les Nazi pour chiffrer leurs communications. Colossus prend la suite en 1944 : 5 tonnes, 2500 tubes à vides, 7 km de câblage, une pièce de 20m<sup>2</sup>. C'est le premier ordinateur programmable, à l'aide de connecteurs électriques et de relais physiques. ENIAC est ensuite construit aux Etats-Unis, en 1945.

Ces premiers ordinateurs sont câblés en dur. Leur fonctionnement est fixé physiquement. Les circuits logiques de calculs doivent donc être modifiés lorsqu'on veut modifier leur fonctionnement. Il faudra attendre le bond conceptuel majeur qu'apporte les travaux de von Neumann (EDVAC) et Turing (ACE) en 1945-46, pour parler de programmation. Les instructions utiles pour décrire le fonctionnement d'un programme sont traitées comme des informations, qui peuvent donc être stockées en mémoire. Le fonctionnement de l'ordinateur devient une donnée. La mise en œuvre intervient en 1948-49.

Ces calculateurs marquent symboliquement le début d'une ère où les machines viendront réaliser des calculs à des vitesses que les humains ne peuvent égaler.

Les hommes ont autrefois confié la pensée aux machines dans l'espoir de se libérer ainsi. Mais cela permet seulement à d'autres hommes de les réduire en esclavage, avec l'aide des machines.

## QUI A CONSTRUIT LE PREMIER ORDINATEUR ?

La question n'est pas tranchée. L'ensemble des briques se met en place, à différents endroits. L'invention de l'ordinateur est une histoire collective, qui se déroule en partie dans le brouillard et le secret de la Seconde Guerre. Les questions de paternité sont donc difficiles à établir - et peu intéressantes, en définitive.

Une sorte de légende se noue autour d'Alan Turing, et de son travail crucial dans le décryptage des messages nazis. Ses contemporains le décrivent comme un génie, et son apport à l'informatique en tant que science et technique est considérable. En plus de son travail sur the Bomb, il participe activement à l'invention de calculateurs (ACE, Manchester Mark 1) et pense les bases de l'intelligence artificielle (1950) et les biomathématiques (1952). Sa mort en 1954 est tragique et reste mystérieuse.

Avant et pendant la Seconde Guerre mondiale, tous les ordinateurs sont des machines destinées à un usage spécial. Elles sont conçues pour résoudre un problème particulier lié à la guerre. Après la guerre, les machines à usage général, dites Turing-complète, voient le jour.

Trois groupes de personnes ont conçu et réussi à faire fonctionner des ordinateurs dans les années 1940. Ils travaillent en relative indépendance. Le groupe anglais de Bletchley Park situé dans la campagne anglaise entre Oxford et Cambridge, auquel appartient Turing, invente Colossus. Aux États-Unis, le groupe auquel appartient von Neumann invente l'ENIAC et EDVAC (1949), qui font partie des premiers ordinateurs programmables à partir d'instructions situées en mémoire vive. Isolé et allemand, l'ingénieur Konrad Zuse finalise le Z3 en 1941. Détruit pendant la guerre, il est peut-être le premier ordinateur à architecture moderne. Mentionnons aussi IBM SSEC (USA, 1948), Manchester Mk1 (UK, 1949), EDVAC (UK, 1949).

Les progrès techniques suivants sont mieux connus. Les transistors sont inventés en 1947. Plus petits, moins chers et plus fiables, ils remplacent progressivement les lampes à vide à partir de 1955. Les circuits intégrés sont inventés en 1957 et commencent à équiper les ordinateurs dans les années 1960. Les microprocesseurs suivront en 1971 (Intel 4004).

# L'INTELLIGENCE ARTIFICIELLE



## CERVEAUX ÉLECTRONIQUES

Dans les années 1960, les bases de l'architecture de l'ordinateur moderne sont jetées. L'architecture générale pensée par von Neumann et Turing est restée inchangée.

Ces machines peuvent effectuer des opérations basiques : addition, mémorisation, branchement logique. Leur fonction majeure est de pouvoir réaliser ces opérations sans se tromper, sans se fatiguer et très vite. Scientifiques et ingénieurs amélioreront énormément leurs performances, en optimisant les composants matériels (Loi de Moore) et les techniques de programmation. Le champ de recherche mêle des contraintes très pratiques (câblage électrique, surchauffe et fragilité des composants, inventions des premiers processeurs, programmation), des sujets scientifiques (comment modéliser l'intelligence?) et des interrogations philosophiques (qu'est-ce que l'intelligence ? qu'est-ce que la conscience ?). Nous nous limiterons principalement aux aspects scientifiques.

# DE LA DÉLICATE DÉFINITION DE L'INTELLIGENCE

L'objectif premier de l'intelligence artificielle est de rendre les machines plus intelligentes. Présentés à l'époque comme des cerveaux électroniques, ces appareils n'ont cependant pas d'intelligence au sens animal du terme.

L'intelligence pose de nombreuses questions, non résolues. Qu'est-ce que l'intelligence ? Intelligence humaine ? Intelligence animale ? Intelligence générale ? Définir l'intelligence est une tâche qui occupe encore les philosophes, il n'y a pas vraiment de définition claire. Doit-on s'intéresser au cerveau, siège matériel de l'intelligence, ou à l'esprit ? Que faire de la conscience ? Est-elle nécessaire à la manifestation de l'intelligence ? Est-ce une conséquence ? Que faire de la relation au corps et au monde, si importante dans le cas des exemples que nous connaissons instinctivement : les bébés humains ?

Le scientifique cherchera à comprendre ce qu'est l'intelligence. L'ingénieur cherchera à rendre les machines utiles. Le positionnement de l'activité est différent si on cherche à comprendre ou à utiliser. D'un point de vue opérationnel, le domaine cherche à rendre les machines capables d'accomplir des tâches qui, à nos yeux d'humains et selon un consensus implicite, nécessitent de l'intelligence. Les acteurs se sont également retrouvés impliqués dans les questions fondamentales de l'informatique, telles que la programmation et l'architecture des ordinateurs.

Nous n'essaierons pas d'être plus précis. Définir le domaine précis de l'intelligence artificielle est un cauchemar conceptuel, que les auteurs de science-fiction arpentent depuis le début du XXème siècle.

## DARTMOUTH 1956

On parle souvent d'âge d'or pour décrire les deux décennies qui s'étendent entre l'école d'été de Dartmouth (USA) en 1956 et la publication du rapport Lighthill (UK) en 1974.

L'école d'été de Dartmouth est un des mythes fondateurs du domaine. Une vingtaine de scientifiques ont assisté à cette école d'été, dont John McCarthy, Ray Solomonoff, Marvin Minsky, Claude Shannon, Oliver Selfridge, Herbert Simon, Allen Newell et John Nash. Cette conférence permet de lancer officiellement le domaine - et de créer le terme d'intelligence artificielle !

Nous proposons qu'un groupe de dix personnes mène une réflexion autour de l'intelligence artificielle pendant deux mois lors de l'été 1956 à Dartmouth College [USA]. Cette réflexion doit être basée sur la conjecture selon laquelle chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence puisse, en principe, être précisément décrite pour qu'une machine puisse être conçue afin de la simuler.

Nous pensons que des progrès significatifs pourront être réalisés [sur les composantes de l'intelligence] si un groupe soigneusement sélectionné de scientifiques travaillent ensemble pendant un été.

John McCarthy, (1955)

Nous retiendrons la définition selon laquelle l'intelligence artificielle consiste à rendre une machine capable de comportements qui, s'ils étaient réalisés par un humain, seraient qualifiés d'intelligents.

John McCarthy, (1955)

Il ne s'agissait pas d'un projet de recherche dirigé. C'était plutôt comme inviter une bande de personnes brillantes à une conférence de huit semaines, où chacun débordait d'idées.

G Solomonoff (2011)

---

# DÉCOMPOSER L'INTELLIGENCE HUMAINE

Les ordinateurs sont des machines nouvelles. L'IA a pour but d'étudier la possibilité de les rendre intelligents et, si possible, de créer les outils et méthodes pour y arriver. Pour manier ces concepts abstraits, voire mouvants, les chercheurs font un pari qui paraît raisonnable : diviser cette « intelligence » en un ensemble de fonctions, construire une brique solution pour chacune, et ensuite assembler l'ensemble de ces briques. C'est la méthode classique appelée « diviser pour régner » (divide & conquer). Des thématiques principales émergent : observer l'environnement en le mesurant avec des capteurs ; interpréter les informations reçues ; mener un raisonnement logique ; établir une stratégie de résolution d'un problème ; planifier l'exécution d'une stratégie ; apprendre à partir des connaissances passées et de l'expérience accumulée ; utiliser et comprendre le langage humain ; interagir avec le monde physique et humain. La liste n'est pas exhaustive.

Les scientifiques se concentrent surtout sur le raisonnement logique et la modélisation de la pensée humaine. Le but est de créer une intelligence artificielle générale. Elle est illustrée par le HAL de 2001, *l'Odyssée de l'espace* (S. Kubrick, 1969), qui représente bien les attentes de l'époque : un ordinateur avec une intelligence surhumaine, capable de se substituer à un équipage entier. Cette approche sera appelée plus tard GOFAI, *Good Old Fashion Artificial Intelligence*.

L'ambition est forte, les attentes sont énormes. L'ambiance est à l'optimisme.

## DES PROGRÈS RAPIDES

Les progrès sont rapides, le champ est en défrichage. Scientifiques et ingénieurs s'attèlent à la réalisation concrète de leurs idées. Le but d'une intelligence artificielle générale (AGI) n'est toutefois pas atteint.

Le domaine fait de beaux progrès dans la thématique de la résolution de problèmes (problem solving). De nombreux systèmes dits intelligents sont inventés, principalement des programmes (SHRDLU, ELIZA, General Problem Solver) et certains robots (SHAKY). Les premiers langages de programmation sont inventés pour programmer plus efficacement les premiers ordinateurs qui existaient ; LISP en fait partie, et il est encore utilisé aujourd'hui.

---

La Logic Theory Machine (LTM) réussit à prouver 18 théorèmes logiques considérés comme clés par Russell, en proposant une démonstration jugée plus élégante. Le logicien, ravi, propose de publier un article sur cette performance, mais le journal refuse d'accepter le programme comme co-auteur.

Le *General Problem Solver* (GPS) généralise le LTM. Il n'est plus limité à la logique et peut être appliqué à n'importe quel problème. D'où son nom. Une fois que l'humain a décrit le problème à traiter, sous la forme d'objectifs hiérarchisés et de moyens d'actions, GPS s'occupe de la logique. GPS sait par exemple résoudre le problème des missionnaires et des cannibales, qui est difficile même pour un humain (essayez !).

*Pandemonium* est une assemblée de démons, sous la forme d'un programme de logique symbolique parallélisé. Chaque démon est spécialisé dans la reconnaissance d'un motif précis. Il se manifeste de façon plus ou moins bruyante, et de façon souvent contradictoire par rapport aux autres voisins. Dans cette cacophonie, un démon supérieur choisit le motif le plus plausible. *Pandemonium* a beaucoup influencé les approches connexionnistes et symboliques.

Enfin, SHRDLU. C'est un des programmes de traitement du langage les plus connus. Il évolue dans un monde de blocs cubiques (block world). Créé au début des années 1970, il peut construire des assemblages de blocs à partir d'instructions en langage naturel. Il peut aussi modifier un assemblage initial, et déterminer seul la liste des actions à mener. Son influence est importante et sort du domaine restreint des spécialistes de l'IA. Une avancée importante réside dans sa capacité à traiter la syntaxe de phrases parfois complexes - jusqu'à un certain point.

## JUSQU'À UN CERTAIN POINT

L'âge d'or prend fin au début des années 1970. L'optimisme communicatif porté par les pionniers est progressivement douché par la réalité, souvent frustrante. Le programme de recherche était bien entendu trop ambitieux, et les résultats trop limités au vu des investissements et de l'engouement public.

Les critiques sont nombreuses. Elles culminent avec le fameux *Alchemy and AI* du philosophe américain H. Dreyfus (*Alchemy and AI*, 1965, *What Computers Can't Do: A Critique of Artificial Reason*, 1972) et le rapport anglais de Pr. Lighthill (*Artificial Intelligence: A General Survey*, 1972). L'approche dite connexionniste, représentée par les premiers réseaux de neurones, se révèle décevante. Les financements se tarissent, avec notamment une réorientation en 1972 des budgets de recherche états-uniens vers des applications militaires dans le cadre de la guerre du Vietnam.

---

Mon but n'est pas de vous surprendre ou de vous choquer.  
Pour le dire simplement, nous disposons aujourd'hui de machines qui pensent, qui apprennent et qui créent. De plus, leurs capacités sur ces sujets vont augmenter rapidement jusqu'à ce que, à un horizon proche, l'éventail des problèmes qu'elles pourront traiter sera proche de ce que l'esprit humain peut faire.

Herbert A. Simon, (1958)

D'ici dix ans, un ordinateur sera champion du monde d'échecs, à moins que les règles ne l'empêchent de concourir.

D'ici dix ans, un ordinateur découvrira et prouvera un nouveau théorème mathématique important.

D'ici dix ans, la plupart des théories de psychologie prendront la forme de programmes informatiques ou de spécifications techniques précises de tels programmes.

Herbert A. Simon, (1958)

De mon point de vue, les progrès de l'intelligence artificielle dans les années 60 et 70 ont été rapides et très impressionnants, car tous ceux que je connaissais travaillaient sur le raisonnement, la représentation symbolique et ce genre de choses.

Marvin Minsky (2011)

Aujourd'hui, les machines résolvent les problèmes en fonction des principes que nous leur donnons. D'ici peu, nous pourrions apprendre à les faire travailler sur le problème très particulier de l'amélioration de leurs propres capacités à résoudre des problèmes. Une fois un certain seuil franchi, cela pourrait conduire à une spirale d'accélération et il pourrait être difficile de mettre au point des garde-fous fiables pour la freiner.

Marvin Minsky (1968)

Les découvertes faites jusqu'à présent n'ont pas produit l'impact majeur promis dans les années 1960.

J. Lighthill (1974)

# L'HIVER SYMBOLIQUE



## MODÉLISER LE RAISONNEMENT

L'approche dite symbolique se heurte à la difficile modélisation du raisonnement logique (humain). Ce problème n'est pas résolu aujourd'hui, et ne semble pas en passe de l'être.

Une mauvaise nouvelle n'arrivant jamais seule, ce premier printemps a mis en évidence le problème du cadre, ou du sens commun (frame problem). Un programme ne sait rien d'implicite ; or l'expérience humaine du monde est construite sur un empilement d'implicites. Il faut tout décrire à un programme. Cette constatation est conceptuellement problématique, et limite fortement les interactions entre humaines et machines ( robots).

Il faudrait s'atteler à la lourde et difficile tâche d'explication des chansons de Georges Brassens. Ces limites dures sont en réalité culturelles. Margareth Boden donne un exemple pertinent, et taquin, d'utilisation de bon sens qui a toutes les chances d'échapper à un programme.

## LE MUR DE L'EXPLOSION COMBINATOIRE

Les programmes de résolution de problèmes rencontrent aussi une limite mathématique : la complexité algorithmique.

La stratégie classique revient souvent à décomposer une tâche complexe en une suite de tâches suffisamment simples pour être exécutée par un ordinateur. On mise ensuite sur l'enchaînement rapide des calculs pour résoudre l'ensemble des petites tâches pour remonter à la tâche initiale.

Les ordinateurs utilisés pour construire ces premiers systèmes ont des capacités de calcul qui excèdent largement les capacités humaines. Ils sont capables de réaliser quelques milliers d'opérations (additions, multiplications) par seconde. En pratique, lorsque la tâche à résoudre devient suffisamment proche de cas réels, ceux qui nous intéressent, la réponse ne peut pas être calculée en un temps raisonnable. Les ordinateurs actuels, plus rapides, rencontrent cependant exactement la même limite, un peu plus tard. L'augmentation des performances des processeurs ne suffit pas à résoudre des problèmes proches des cas réels. La cause est mathématique.

Cette approche logique de décomposition en tâches simples multiplie les branchements. L'espace des possibles devient gigantesque. On parle d'explosion combinatoire. Ce phénomène est au cœur de nombreux problèmes qui intéressent les chercheurs en intelligence artificielle : optimisation logistique (problème du voyageur de commerce), allocation de ressources (problème du sac à dos), parcours de graphes (problème de la clique, coloration de graphe) et généralement des problèmes de puzzles logiques. C'est une découverte majeure de l'époque, qui intéresse les mathématiciens et les informaticiens théoriciens : si certaines classes de problèmes peuvent être résolues par un calcul mécanique en utilisant un algorithme adapté, le calcul en lui-même peut être extrêmement long. Curiosité mathématique ? Non, car de nombreux problèmes d'optimisation du monde réel se réduisent à ces cas.

L'intelligence ne se résume pas à un calcul. Même rapide.

## DÉSILLUSION

Les dommages infligés au milieu des années 1970 sont tels que de nombreux universitaires commencent à considérer l'IA comme une pseudoscience. La guérison sera longue.

Jusque dans les années 1970s, les chercheurs font des prédictions pour le moins téméraires sur les progrès futurs dans le domaine. Leur sincérité et leur enthousiasme sont compréhensibles. Cette époque est marquée, rétrospectivement, par une bonne dose de naïveté. Ces prédictions n'ont cessé de hanter l'IA depuis lors. Au milieu des années 1970, la période faste était terminée et un retour de bâton vicieux s'est amorcé - un cycle d'expansion et de ralentissement de l'IA amené à se répéter au cours des prochaines décennies.

Il n'existe pas d'algorithme pour  
l'intelligence artificielle générale [...]

Il est beaucoup plus probable qu'une percée scientifique majeure soit nécessaire, et personne ne sait à quoi elle ressemblerait et encore moins des détails pour y parvenir.

E. Larson (2021)

Il ne faut pas sous-estimer les dommages infligés à l'IA au milieu des années 1970.

De nombreux universitaires ont commencé à considérer l'IA comme une pseudo-science - cette mauvaise réputation s'est installée lors de l'hiver de l'IA et le domaine en a souffert jusqu'à très récemment."

M Wooldridge (2021)

Mon exemple préféré est le suivant. Si un homme de vingt ans peut cueillir cinq kilogrammes de fraises en une heure, et qu'une femme de dix-huit ans peut en cueillir quatre, combien en ramasseront-ils s'ils y vont ensemble ? Il est certain que "neuf" n'est pas une réponse plausible. Ce pourrait être beaucoup plus (parce qu'ils font les malins tous les deux) ou, plus probablement, beaucoup moins.

M. Boden (2006)

# LA CONNAISSANCE



## INJECTER LA CONNAISSANCE

L'âge d'or s'était trop concentré sur des approches générales comme la recherche et la résolution de problèmes, pensent certains chercheurs. Un ingrédient clé manquait, un élément crucial de toute activité intelligente : la connaissance. Les années 1980 voient l'émergence de systèmes de conception et d'analyse, ou systèmes experts. Ce sont des programmes qui répondent à des problèmes très précis, avec des applications commerciales possibles. L'objectif n'est pas de créer une intelligence générale, mais de se spécialiser sur des domaines particuliers, étroits ou profonds.

Les connaissances humaines sont assemblées en bases de données. Leur exploitation est articulée par des règles utilisant des relations causales. Le système apprend avec l'aide d'un expert.

# LES SYSTÈMES EXPERTS PROUVENT LEUR EFFICACITÉ

De nombreux systèmes experts sont développés dans les années 1960 et 1970, sur des thématiques spécifiques.

DENDRAL (Dynamic ENvironment for Deducing REasoning and Learning) est un système expert développé à l'université Stanford par E. Feigenbaum et B. Buchanan, dans les années 1960 et 1970. Il est conçu pour aider les scientifiques à déterminer la structure moléculaire de composés chimiques en utilisant des données spectroscopiques. C'était l'un des premiers systèmes experts à être développé et a été un jalon important dans l'histoire de l'intelligence artificielle (IA).

MYCIN est un système expert développé à l'université de Stanford dans les années 1970. Son nom vient du fait que de nombreux antibiotiques ont pour suffixe «-mycine». Le but de ce programme est de fournir une assistance au médecin dans le diagnostic et le traitement des infections du sang. MYCIN a la capacité de poser ses questions et peut expliquer son diagnostic en répondant à certaines questions de l'utilisateur. Le système peut aussi exprimer l'incertitude de sa décision... et être parfois plus efficaces que des humains.

R1/XCON, pour eXpert CONfigurer, est développé à l'université Carnegie Mellon dans les années 1980. Il est utilisé comme outil de configuration des systèmes informatiques vendus par DEC à partir des demandes des clients. XCON est l'un des premiers systèmes experts à être utilisés dans un contexte commercial et a été un véritable succès pour le domaine de l'IA. DEC estime qu'il a permis d'économiser 40 millions de dollars en six ans.

Nous pourrions citer aussi SOAR (*Success Oriented Achievement Realized*), qui a l'ambition d'être un modèle cognitif complet. Il intègre la perception, l'attention, la mémoire, les associations, la capacité à réaliser des inférences et des analogies, et enfin l'apprentissage. Il a été perfectionné et est utilisé aujourd'hui en médecine ou dans les problèmes de planification logistique.

Ces succès ont attiré l'intérêt des investisseurs, ce qui a créé un boom économique. Et c'est reparti pour une bulle d'intérêt, entre 1980 et 1987 environ.

## UN RAISONNEMENT ACCESSIBLE

Les règles logiques simples permettent de construire les schémas de représentation des connaissances utilisés dans les systèmes experts. La connaissance est toutefois difficile à représenter et à modéliser, surtout dans des environnements complexes. Les chercheurs ont travaillé sur ces sujets, afin

---

de continuer à construire des systèmes qui s'appuient sur une logique transparente, proche du raisonnement humain.

## **CYC LE MONSTRE**

Le plus célèbre des systèmes experts est le projet Cyc. Ce projet immense vise à créer une IA générale en lui injectant une quantité énorme de savoir. Cette base de connaissances de Cyc a besoin d'une description complète de la «réalité consensuelle» - le monde tel que nous le comprenons, tel que l'apprend un enfant. L'approche proposée par son fondateur, le chercheur Doug Lenat, est quelque peu extrême. L'estimation basse avant que Cyc puisse s'auto-éduquer monte à 200 personnes pendant un an. Projet époustouflant, qui trouve néanmoins un financement durable. En 2017, Cyc contiendrait 70 milliards de faits sur 500 millions d'entités. Ce projet monstrueux a, sans surprise, échoué à créer une IA générale.

## **UN SECOND HIVER ?**

Les systèmes experts sont des exemples de programmes utiles sur des domaines très précis. Ils sont encore utilisés aujourd'hui. Leurs limites sont nombreuses, notamment la gestion des incertitudes, voire des contradictions, de la vie quotidienne. De manière plus pratique, ces systèmes sont coûteux à concevoir et maintenir, difficiles à mettre à jour voire figés et surtout «fragiles». Un système expert peut tout à fait raconter n'importe quoi, donc faire des erreurs, lorsqu'on lui donne des données inhabituelles, pathologiques ou en-dehors de son champ d'expertise. Finalement, vers la fin des années 80, l'intérêt du public pour les systèmes d'expert a cessé, sans qu'ils aient forcément déçu cependant. Cette période illustre que la connaissance est difficile à représenter et à modéliser.

Il n'existe pas de raccourci pour l'intelligence, pas d'équations de Maxwell de la pensée à découvrir, ni de formalisme puissant qui nous dispensent de la nécessité d'accumuler beaucoup de connaissances.

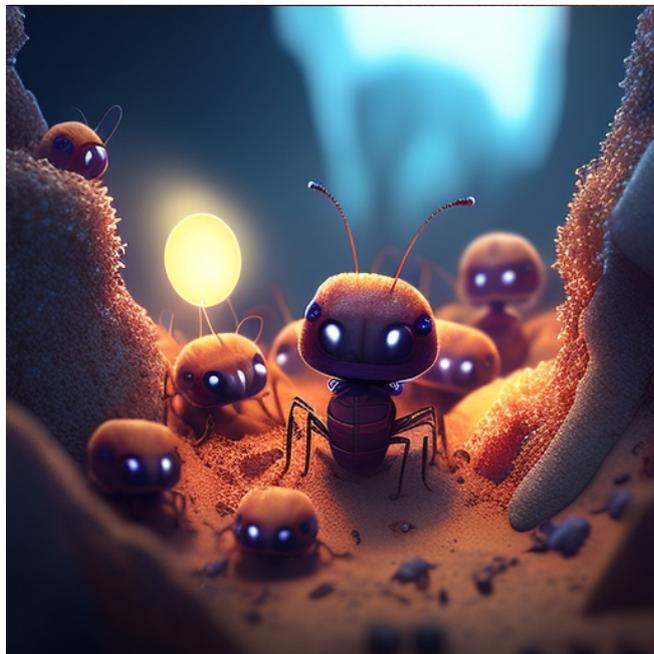
En fait, la plupart des choses que nous devons savoir pour nous débrouiller dans le monde réel sont trop évidentes pour être dans des livres : par exemple, on ne vit qu'une fois, rien ne peut être à deux endroits à la fois, les animaux n'aiment pas la douleur. Peut-être que la vérité la plus difficile à affronter, une vérité que le domaine essaie d'esquiver depuis 34 ans, est qu'il n'existe probablement pas de moyen élégant et facile d'obtenir cette immense base de connaissances.

L'essentiel de l'effort doit plutôt consister, du moins au début, en une saisie manuelle de faits.

Doug Lenat (1990)

---

# UNE INTÉGRATION PROGRESSIVE



## REMISE EN QUESTION

À mesure que la compréhension scientifique progresse, les hypothèses évoluent sous la pression des échecs manifestes. À la fin des années 1980, l'époque faste des systèmes experts est révolue et une nouvelle crise de l'IA se profile. Une fois de plus, la communauté de l'IA est critiquée pour avoir trop promis. L'hypothèse remise en question n'est pas seulement la doctrine du «savoir, c'est le pouvoir», responsable du boom des systèmes experts, mais les hypothèses de base qui sous-tendent l'IA depuis les années 1950 - l'IA symbolique en particulier. Les critiques les plus féroces de l'IA viennent de l'intérieur du domaine.

## TRAVAILLER SUR LE COMPORTEMENT

La vision originelle du domaine est celle d'un système d'IA qui maintient un modèle central symbolique et logique de son environnement, autour duquel gravitent toutes les activités de l'intelligence. Ce modèle, poussé par le scientifique John McCarthy, doit être révisé. Certains le rejettent.

Les programmes doivent cesser d'être pensés pour un monde fermé. Ils doivent sortir des laboratoires et s'implanter dans le monde réel. L'IA comportementale devient le nouveau paradigme, à travers le développement d'une architecture dite réactive (*subsumption architecture*), qui permet le fonctionnement d'un robot en définissant une série de comportements possible et en les classant par ordre de priorité. Le robot-aspirateur Roomba (iRobot, 2002) a été créé sur ces principes.

L'un des principaux objectifs de l'IA comportementale est de créer des systèmes intelligents capables de s'adapter et d'apprendre de leur environnement, et d'adopter des comportements appropriés à une situation donnée.

## VERS LES AGENTS

Ces IA comportementales ont cependant deux profondes limitations : leurs compétences sont limitées, et elles n'évoluent pas. La communauté s'oriente vers la construction d'agents dans les années 1990.

Un agent est vu comme un système IA complet. C'est une entité autonome, située dans un environnement donné, qui exécute une tâche spécifique pour le compte d'un utilisateur. Un agent fournit un ensemble complet et fonctionnel, incluant une base de connaissance et une capacité de raisonnement logique.

Ces agents représentent une forme de synthèse. L'âge d'or de l'IA a mis l'accent sur le comportement proactif, comme la planification et la résolution de problèmes. L'IA comportementale a souligné l'importance d'être réactif, incarné et attentif à l'environnement. L'IA basée sur les agents exige les deux et ajoute une nouveauté : l'idée de devoir travailler avec d'autres agents et, pour cela, disposer de compétences «sociales», comme la capacité non seulement de communiquer, mais de coopérer, de coordonner et de négocier avec d'autres agents dans l'accomplissement de leurs tâches. Ces compétences sociales permettent de penser un nouveau paradigme pour l'IA basée sur les agents.

Avec le recul, il semble un peu étrange qu'il ait fallu si longtemps à l'IA pour commencer à réfléchir sérieusement à la manière dont les systèmes d'IA pourraient interagir entre eux et aux problèmes qui pourraient survenir lorsqu'ils le feraient.

## LE SOUS-MARIN HOMER

Cet agent est une simulation de robot sous-marin en deux dimensions, créée en 1990 (Vere & Bickmore).

Il intègre des capacités de compréhension et de génération du langage naturel, de planification et de raisonnement, la capacité à exécuter un plan, une perception symbolique simulée, une mémoire et une connaissance générale du monde. Il peut communiquer avec un vocabulaire d'environ 800 mots anglais courants, avec une grammaire basique. L'agent a une mémoire épisodique des événements de sa vie et a une capacité limitée à réfléchir sur ces événements. L'opérateur peut parler à l'agent, lui poser des questions et lui donner des ordres. En réponse, un planificateur de tâches temporelles intégré synthétise un plan d'action, qui est ensuite exécuté ou amendé en fonction des nouvelles informations fournies.

HOMER, c'est son nom, est capable de donner des réponses concises à des questions sur ses expériences passées, ses activités et perceptions présentes, ses intentions futures et sa connaissance générale.

## LES ASSISTANTS

Ce type d'IA est utilisé aujourd'hui pour créer des agents logiciels. Ces agents ont pour but de travailler de manière active avec nous ou pour nous, en traitant par exemple nos emails ou en nous agaçant lorsque nous nous servons de la suite MS Office (quelle horreur, ce trombone !). L'assistant d'Apple, SIRI, est un exemple d'agent conversationnel basé sur cette approche ; pour la petite histoire, il a été créé par l'institut SRI, qui avait créé SHAKEY à la fin des années 1960.

Avec l'apparition des assistants, les attentes évoluent. L'IA ne doit plus faire des choix identiques à ceux qu'aurait fait un humain ; l'IA doit faire de meilleurs choix. Cette théorie remonte aux années 1940 (von Neumann, Morgenstern). Elle se base sur les préférences de l'utilisateur et utilise les probabilités afin de déterminer les meilleures situations dans les cas les plus complexes. C'est encore cette méthode qu'on utilise dans les IA aujourd'hui.

Dans les années 1990, la construction d'agents intelligents capables d'agir de manière rationnelle est la nouvelle orthodoxie de l'IA et le reste aujourd'hui. Dans presque tous les systèmes d'IA actuels, il existe un calcul d'utilité incluant les préférences de l'utilisateur, afin de maximiser les résultats attendus, c'est-à-dire d'agir rationnellement au nom de l'utilisateur.

## TRAVAILLER EN TOUTE INCERTITUDE

Une voiture automatique obtient des flux de données de ses capteurs. Ceux-ci ne sont pas parfaits, il y a toujours une chance, faible si le système est bien conçu, qu'un télémètre se trompe lorsqu'il annonce «je ne vois pas d'obstacle». Ces informations ne sont pas inutiles, pour autant, mais nous ne pouvons pas simplement supposer qu'elles sont toujours correctes. Alors comment l'utiliser, en tenant compte de la possibilité d'erreur ?

La gestion de l'incertitude est un problème qui rôde depuis longtemps. Tout système d'IA réaliste doit faire face à une forte incertitude. Ce problème devient crucial dans les années 1990. L'inférence bayésienne s'impose alors.

Inventé par un mathématicien britannique du XVIIIe siècle, le révérend Thomas Bayes, le théorème de Bayes permet de traiter des données imparfaites. Ce fonctionnement est naturel pour tout être vivant qui évolue dans un monde incertain. Les informations de l'environnement sont utilisées pour mettre à jour notre compréhension de celui-ci. Le théorème de Bayes est intéressant parce qu'il met en évidence un processus de prise de décision probabiliste. Ce processus s'adapte bien aux programmes.

---

# DEUX RÉUSSITES



## UNE LENTE MATURATION

L'utilisation et l'acceptation de ces outils entraîne un changement important dans la perception de l'IA au sein même de la communauté. Pour la première fois, les chercheurs n'ont plus l'impression de tâtonner dans le noir, en s'accrochant désespérément à ce qu'ils trouvent. La communauté a le sentiment de disposer d'une base scientifique bien établie, issue de la théorie des probabilités et de la théorie de la prise de décision rationnelle.

## DEEPBLUE

La percée la plus importante est réalisée par IBM, en 1997. Un système d'IA, appelé DeepBlue, est capable de battre le grand maître russe Garry Kasparov aux échecs. Désormais, les échecs sont un jeu maîtrisé par les ordinateurs. Les systèmes d'IA deviennent donc capables de battre de manière fiable tous les joueurs, sauf les meilleurs. La domination des programmes s'applique sur l'ensemble des jeux dits logiques, où le hasard joue un rôle mineur : il n'y a plus de compétition possible aux échecs ou au go, comme

nous le verrons plus bas.

DeepBlue n'utilise pas de réseaux de neurones. Le terme deep vient de la capacité du programme à effectuer des recherches à grande profondeur dans les arbres de branchements

## SOLVEURS SAT

La seconde réalisation majeure de l'époque est restée plus confidentielle. Elle concerne la complexité des calculs et l'explosion combinatoire qui en découle. À la fin des années 1990, les programmes de résolution de certains de ces problèmes NP-complets (SAT, problème de satisfaisabilité booléenne) sont suffisamment puissants pour être utilisés sur des problèmes à l'échelle industrielle. On parle de solveur SAT. Cela ne signifie pas que les problèmes NP-complets peuvent désormais être résolus de façon nette : il y aura toujours des cas qui mettront à genoux même les meilleurs solveurs SAT. Mais nous n'avons plus peur de la NP-complétude, et c'est l'une des principales réussites méconnues de l'IA des 40 dernières années.

## ... ET L'INTELLIGENCE GÉNÉRALE?

Toujours pas de feuille de route vers l'IA générale en vue, cependant.

Ce sujet est délaissé.

1997 est un moment historique, bien qu'il s'agisse d'une machine utilisant la force brute, et non d'une machine avec un raisonnement "humain". Mais comme nous l'avons découvert, les échecs étaient vulnérables à des machines très puissantes avec de bons algorithmes, de grosses bases de données et des processeurs très rapides.

Gary Kasparov (2017)

Face à une machine, vous serez rapidement dépassé. C'est pourquoi tout système fermé – et les jeux sont des systèmes fermés – donne automatiquement l'avantage aux machines. Je le savais en 1997. Lorsque vous regardez la puissance de calcul des ordinateurs d'échecs, Deep Blue était relativement faible par rapport aux normes modernes. Aujourd'hui, les machines sont absolument monstrueuses.

Elles sont beaucoup plus fortes que Magnus Carlsen [champion du monde d'échecs depuis 2013], et une application d'échecs gratuite sur votre appareil mobile est probablement plus forte que Deep Blue.

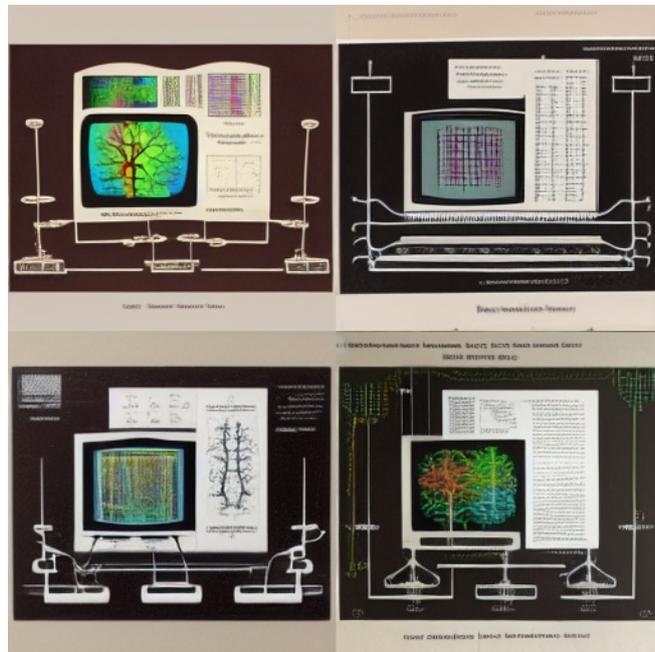
Gary Kasparov (2017)

Lorsque vous opposez les humains aux machines dans des jeux, les machines l'emporteront, que ce soit aux échecs ou au go, ou tout autre jeu. Pas parce qu'elles peuvent résoudre le jeu. Les échecs sont mathématiquement insolubles. Le nombre de coups légaux est énorme...

Mais en fin de compte, la machine n'a pas besoin de résoudre le jeu. La machine doit gagner la partie. Et pour gagner la partie, elle doit simplement faire moins d'erreurs que les humains. Ce n'est pas si difficile, car les humains ne sont qu'humains et sont donc vulnérables; nous n'avons pas le calme et la sérénité d'un ordinateur.

Gary Kasparov (2017)

# L'APPRENTISSAGE AUTOMATIQUE



## DISCIPLINE FILLE DE L'IA

L'apprentissage automatique est une discipline de l'intelligence artificielle, aussi ancienne. L'histoire chahutée du domaine peut expliquer pourquoi, aujourd'hui encore, de nombreux spécialistes de l'apprentissage automatique ne se sentent pas acteurs du domaine de l'IA. Ce sujet pourrait mériter un débat à part entière, à la frontière entre technique et culture.

Le but de l'apprentissage statistique est de construire des programmes capables de calculer une sortie souhaitée à partir d'une entrée donnée, sans leur fournir de méthode explicite. On parle souvent d'apprentissage automatique, ou machine learning. L'idée est de construire un algorithme qui est capable d'exploiter les motifs et régularités présents dans un jeu de données afin de déduire des règles descriptives, sans intervention humaine. La reconnaissance de texte manuscrits ou de d'image en sont des exemples classiques. Pour fonctionner, ces programmes sont entraînés sur de vastes bases de données.

## QUEL APPRENTISSAGE ?

Il n'y a pas vraiment d'apprentissage au sens humain du terme. Le programme va détecter des relations entre les données d'entrées (=description d'un monde) et les données de sorties (=résultats). Ces relations sont sauvegardées par le programme (=base de connaissances) et sont utilisées pour prédire le résultat le plus probable pour des données d'entrées qu'il ne connaît pas. Le souci majeur vient de l'étroitesse de l'apprentissage. Ce qui est appris sur un jeu de données n'est utile que sur un jeu de données dont la structure et le contenu informationnel sont proches. Un agent conversationnel entraîné sur un corpus en anglais sera incapable de dire quelque chose d'utile sur du français. Un outil d'aide au diagnostic médical pour l'analyse d'IRM cardiaque ne sera pas lire une radiographie par rayons X. On parle d'apprentissage étroit. Les algorithmes d'apprentissage automatique sont donc extrêmement spécialisés.

La construction d'une IA générale à partir de techniques d'apprentissage automatique impliquerait, peut-être, de réussir à casser ces silos de spécialisations, pour réaliser du transfert d'apprentissage.

## LES TROIS COQUILLAGES

Il existe trois types principaux d'apprentissage automatique. L'apprentissage supervisé, où le programme est entraîné avec des exemples d'entrée et de sortie. C'est le type d'apprentissage le plus utilisé actuellement. L'apprentissage non supervisé, où le programme extrait seul des classes ou groupes de points présentant des caractéristiques communes. L'apprentissage par renforcement, où le programme est un agent qui reçoit des récompenses ou des punitions en fonction de ses décisions. L'agent apprend à maximiser sa récompense en prenant des décisions en fonction de ses expériences passées.

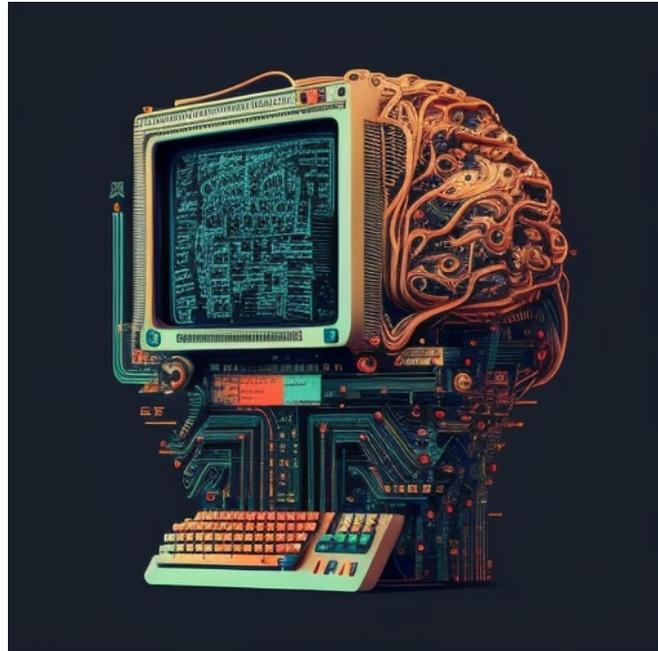
De nombreux algorithmes d'apprentissage automatique ont été inventés et perfectionnés au cours des soixante-dix dernières années. Les réseaux de neurones, dont le nom est bien connu du grand public, et bien d'autres : les différents types de régressions (Lasso, Ridge, logistique), arbre de décision, k plus proches voisins, modèles linéaires généralisés (des régressions, donc), machine à vecteurs de support, forêts aléatoires, et réseaux de neurones. Chaque technique s'appuie sur un algorithme et un modèle mathématique différent, avec ses particularités.

Programmer les ordinateurs pour qu'ils apprennent par l'expérience pour, à la fin, être dispensé d'une grande partie de l'effort requis par une programmation détaillée.



---

# LES RÉSEAUX DE NEURONES



## UNE VIEILLE IDÉE

Ces réseaux de neurones s'appuient sur une modélisation mathématique qui remonte aux années 1940 (McCulloch & Pitts). L'idée des chercheurs consiste à connecter des neurones artificiels très simplifiés pour concevoir un programme informatique. La structure artificielle ainsi créée pourrait favoriser l'émergence de fonctions d'apprentissage et de mémorisation, par analogie avec le fonctionnement d'un cerveau animal. Cette approche est datée, et probablement fautive.

La modélisation est simple. Les neurones sont des cellules qui peuvent communiquer entre elles via des connexions synaptiques. Un neurone reçoit des signaux électrochimiques et, selon l'importance relative des signaux reçus, va engendrer un signal de sortie qui sera à son tour reçu par d'autres neurones. Toute l'intelligence est dans le processus de pondération des signaux d'entrée et la structure des communications entre les neurones artificiels. Concernant les neurones biologiques, ce n'est pas qu'une histoire de pondération.

Dans les systèmes nerveux des animaux, les réseaux de neurones sont massivement interconnectés : le cerveau humain est composé d'environ cent milliards de neurones, et les neurones du cerveau humain ont généralement des milliers de connexions. Les réseaux de neurones observés dans la nature sont organisés selon des schémas (extrêmement) complexes. Pour des raisons pratiques, les réseaux de neurones artificiels sont généralement organisés en couches : chaque neurone de chaque couche successive reçoit toutes les sorties de la couche qui le précède.

Les poids associés à chaque connexion sont cruciaux pour le fonctionnement d'un réseau de neurones. L'entraînement d'un réseau de neurones consiste donc à trouver, d'une manière ou d'une autre, des poids numériques appropriés. L'approche habituelle consiste à ajuster le poids après chaque épisode d'entraînement, avec une descente de gradient par exemple, en essayant de faire en sorte que le réseau relie correctement les entrées aux sorties.

## LES PERCEPTRONS, PREMIERS RÉSEAUX DE NEURONES

Cette méthode est l'une des plus anciennes : les réseaux de neurones font partie de la proposition originale de John McCarthy pour son école d'été sur l'IA en 1956. Comme l'histoire de l'IA elle-même, l'histoire des réseaux de neurones n'est pas linéaire, avec ses «hivers» spécifiques. A la fin du XXème siècle, de nombreux spécialistes ne tenaient pas ce domaine en haute estime. Essayons de raconter leur histoire.

Les premiers réseaux de neurones sont appelés perceptrons (Rosenblatt, 1957). Cette démonstration est un programme exécuté sur un IBM 704 d'une valeur de \$2000000 à l'époque. Suivront d'autres implémentations physiques.

## UN COUP D'ARRÊT BRUTAL

Depuis le lancement en fanfare du Perceptron de Rosenblatt, les réseaux de neurones sont en vue. La question des financements est importante, tout comme les limites théoriques de ces techniques. L'histoire est un peu longue, et a passablement énervé les ayatollahs du connexionnisme. Nous renvoyons à Margareth Boden (*Mind as a Machine*, Chap 12.iii *Attack Without Apology*) pour plus de détails.

À l'époque, personne ne sait entraîner un réseau à couches multiples : il s'agit d'une possibilité théorique plutôt que d'une réalité opérationnelle. Tout le monde se concentre sur les réseaux à une couche.

Les réseaux de neurones artificiels (ANN) sont constitués de nombreuses unités interconnectées, chacune d'entre elles étant capable de calculer une seule chose. Décrits de cette manière, ils peuvent sembler ennuyeux. Mais ils peuvent sembler presque magiques. Ils ont en tout cas ensorcelé les journalistes. [Ces ANN] ont appris à reconnaître des lettres sans qu'on leur ait explicitement appris à le faire et ont été présentés avec enthousiasme dans les journaux des années 1960.

Les ANN ont fait beaucoup de bruit au milieu des années 1980 et sont encore régulièrement salués par les médias.

L'engouement récent pour les ANN concerne  
l'apprentissage profond

Margaret Boden (2016)

La Navy a dévoilé aujourd'hui l'embryon d'un ordinateur électronique qui pourra marcher, parler, voir, écrire, se reproduire et être conscient de son existence... Dans le futur, on s'attend à ce qu'ils puissent reconnaître les gens et les appeler par leur nom, et traduire instantanément d'une langue à l'autre [...]

Dr Rosenblatt précise qu'il ne manque plus qu'une étape de développement, difficile.

New York Times (1958)

Minsky et Papert montrent que cette structure basique à une couche a des limites fondamentales (Perceptrons: An Introduction to Computational Geometry, 1969). Différentes versions du livre circulent. Les critiques sont dures. La discipline ne s'en remettra pas.

La recherche sur les réseaux neuronaux a fortement diminué à la fin des années 1960 au profit des approches de l'IA symbolique défendues par McCarthy, Minsky, Newell et Simon. On parle volontiers d'âge noir du connexionisme. Il faudra attendre deux décennies, et des avancées techniques et scientifiques significatives, pour que l'entraînement de réseaux multicouches soit possible.

À la fin des années 1950, après les travaux de Rosenblatt, il y a eu une grande vague d'activités de recherche sur les réseaux neuronaux. Il y avait peut-être des milliers de projets au début des années 60, après les travaux de Rosenblatt. Par exemple, le Stanford Research Institute avait un bon projet. Mais rien ne s'est produit. Les machines étaient très limitées. Donc je dirais qu'en 1965 les gens commençaient à s'inquiéter. Ils essayaient d'obtenir de l'argent pour construire des machines plus grandes, mais cela ne semblait mener nulle part. C'est alors que [Seymour] Papert et moi avons essayé d'élaborer la théorie de ce qui était possible pour les machines sans boucles [perceptrons à anticipation].”

Marvin Minsk (1989)

Il nous semble que l'effet de Perceptrons n'a pas été simplement d'interrompre une saine ligne de recherche. Cette réorientation des préoccupations n'était pas une diversion arbitraire ; c'était un interlude nécessaire. Pour continuer à progresser, les connexionnistes devaient prendre un temps d'arrêt et développer des idées adéquates sur la représentation des connaissances.

Il y avait une certaine hostilité dans notre motivation à travailler sur Perceptrons ... Une partie de notre motivation venait, comme nous l'avons reconnu de manière très claire dans notre livre, du fait que les financements et l'énergie de la recherche étaient dissipés sur ... des tentatives trompeuses d'utiliser les méthodes connexionnistes dans des applications pratiques

Marvin Minsky et Seymour Papert (1988)

Comment les scientifiques concernés ont-ils réagi à [notre analyse] ? La version populaire est que la publication de notre livre a tellement découragé la recherche sur l'apprentissage dans les réseaux de neurones qu'une ligne de recherche prometteuse a été interrompue.

Notre version est que les progrès s'étaient déjà virtuellement arrêtés en raison du manque d'une théorie de base adéquate, et que les leçons de ce livre ont donné au domaine [l'IA en général], un nouvel élan, en réorientant un peu paradoxalement ses préoccupations immédiates.

Marvin Minsky et Seymour Papert (1988)

En effet, Minsky a admis plus tard que  
"en toute modestie, nous avons été excessivement  
influents"  
en dissuadant les scientifiques travaillant sur GOFAI de  
travailler sur l'apprentissage [sic].

Margaret Boden (2016)

Ce que j'aimerais souligner ici, c'est que les principaux  
points des arguments de Minsky et Papert contre le  
perceptron étaient bien connus au milieu des années 1960,  
et que ces arguments avaient déjà eu un effet critique sur la  
recherche sur les réseaux neuronaux.

M. Olazaran (1993)

## LE MOMENT CONNEXIONNISME

Le domaine se relance en 1986 avec la présentation d'une technique appelée rétro-propagation, ou *backprop*, (*Parallel Distributed Processing*, Rumelhart & McClelland) : elle permet d'entraîner enfin des réseaux de neurones multicouches. Un réseau typique de cette époque compte une centaine de neurones.

La bulle PDP est intense et de courte durée. Au milieu des années 1990, la recherche sur les réseaux neuronaux stagne, pour être à nouveau délaissée. Cette fois-ci, les limites concernent l'informatique : les ordinateurs de l'époque ne sont pas assez puissants pour entraîner les modèles. Et la concurrence est rude dans le domaine de l'apprentissage automatique, car les machines à vecteurs de support (SVM) et les forêts aléatoires montrent de bons résultats.

## LES RÉSEAUX DE NEURONES, À NOUVEAU INCONTOURNABLES

Le cycle d'intérêt actuel des réseaux de neurones commence en 2006 avec une nouvelle technique d'entraînement de certains réseaux de neurones proposée par G. Hinton (*deep belief network*). C'est le troisième cycle, le troisième printemps. Les réseaux sont plus profonds : on parle désormais de *deep neural networks*, ou l'apprentissage profond (*deep learning*) pour le grand public qui découvre le domaine. Ces réseaux profonds ont plus de neurones qui sont mieux connectés et organisés en plus de couches. En 2016, on compte typiquement un million de neurones connectés à 10 000 autres neurones, et organisés en une dizaine de couches.

Et des données, beaucoup de données. Traitées par des ordinateurs puissants, elles sont les clés du succès. L'apprentissage profond va profiter pleinement de l'accumulation de données des années 2010 (*Big Data*). L'utilisation massive du web et la démocratisation des objets connectés, donc potentiels capteurs de données, sont responsables de ce nouveau paradigme technique - et économique.

Une démonstration d'efficacité marque les esprits. Une compétition de classement d'image propose à des équipes de chercheurs de classer des images issues de la base ImageNet, réparties en 1000 catégories. Lors de l'édition de 2012, dont le jeu de données est constitué de 1.2 million d'images d'entraînement et 150 000 images de test, un réseau de neurones arrive en tête avec des performances remarquables.

L'excitation suscitée par l'arrivée du PDP était due en grande partie au fait que les [réseaux neuronaux] (i.e. le connexionnisme) avaient été considérés comme une impasse vingt ans plus tôt. [Ce] jugement avait été rendu dans les années 1960 par une critique virulente de Marvin Minsky et Seymour Papert, qui jouissaient tous deux d'une excellente réputation dans la communauté de l'IA. Dans les années 1980, [les réseaux neuronaux] semblaient être non seulement une impasse, mais réellement morts.

Presque tous les fonds de recherche avaient été consacrés à l'IA symbolique.

Margaret Boden (2016)

L'apprentissage profond est passionnant car il permet enfin aux ANN [réseaux de neurones] de traiter la hiérarchie. Depuis le début des années 1980, les connexionnistes se sont efforcés de représenter la hiérarchie en combinant la représentation locale/distribuée ou en définissant des réseaux récurrents. [...]

Leur succès est limité, et les ANN ne sont toujours pas adaptés à la représentation précise de hiérarchies avec ou au raisonnement déductif.

L'apprentissage profond a été initié, en fait, dans les années 1980 (par Jürgen Schmidhuber). Mais le domaine n'a explosé que plus récemment, lorsque Hinton a fourni une méthode efficace permettant aux réseaux multicouches de découvrir des relations sur de nombreux niveaux.

Margaret Boden (2016)

Nos résultats montrent qu'un grand réseau de neurones convolutif profond est capable d'obtenir des résultats records sur un ensemble de données complexe en utilisant un apprentissage purement supervisé.

Geoffrey Hinton (2012)

---

# UNE TRIPLE LECTURE

L'intelligence artificielle est un domaine complexe avec une histoire chahutée. Il y a trois histoires, pour trois mondes. Peut-être plus !

Le premier monde de l'IA est le monde de la recherche, du futur. Les équipes scientifiques travaillent à comprendre ce qu'est l'intelligence pour la reproduire dans le silicium des processeurs et la logique de haut niveau des programmes. Le deuxième monde de l'IA est le monde de l'application, du présent. Les ingénieurs utilisent les méthodes inventées par le premier monde pour construire des outils d'aide à la décision efficaces, rapides et pertinents. Le troisième monde de l'IA est le monde du mythe, du récit. La sphère médiatique s'empare de sujets qu'elle ne maîtrise pas et construit l'imaginaire collectif, à grand renforts d'exagération, de délires, de fantasmes.

Ces trois mondes sont découplés. Les outils créés par le premier monde, celui de la science, sont réinterprétés librement par le troisième monde, celui du récit. Les scientifiques disent un possible, un potentiel, un pensable, qui est transformé par les communicants en récits parfois plausibles, souvent exagérés - voire faux. Promettre des lendemains qui chantent ou décrire d'horribles cauchemars : ces leviers sont particulièrement efficaces pour le domaine de l'intelligence artificielle.

Voici le drame de l'intelligence artificielle, depuis ses débuts dans les années 1950. Les enjeux symboliques sont puissants, si bien que les scientifiques eux-même peinent à garder la tête froide. Les succès ont un écho médiatique important : les victoires de DeepBlue, et bientôt AlphaGo, contre les humains marquent les esprits. Tout comme les annonces fracassantes des transhumanistes kurzweilliens ou d'autres beaux-parleurs concernant les risques de superintelligence, l'augmentation des capacités neuronales humaines via l'ajout de puces, les interfaces neuroniques. Les annonces concernant les véhicules automatiques, d'apparence plus prudentes, entretiennent aussi les attentes du grand public. L'intelligence artificielle est coutumière des annonces fracassantes suivies d'attentes forcément déçues. Les succès durables sont discrets, à bas bruit. L'IA doit être comprise comme une limite sans cesse repoussée.

“Il y a ce mythe stupide selon lequel l'IA a échoué, mais l'IA est partout autour de vous à chaque seconde de la journée. Les gens ne le remarquent tout simplement pas. Il y a des systèmes d'IA dans les voitures, qui règlent les paramètres des systèmes d'injection de carburant. Quand vous atterrissez dans un avion, votre porte d'embarquement est choisie par un système de programmation d'IA. Chaque fois que vous utilisez un logiciel Microsoft, vous avez un système d'IA qui essaie de comprendre ce que vous faites, comme écrire une lettre, et il fait un sacré bon travail. Chaque fois que vous voyez un film avec des personnages générés par ordinateur, ce sont tous des petits personnages d'IA qui se comportent comme un groupe. Chaque fois que vous jouez à un jeu vidéo, vous jouez contre un système d'IA.”

Rodney Brooks (2022)



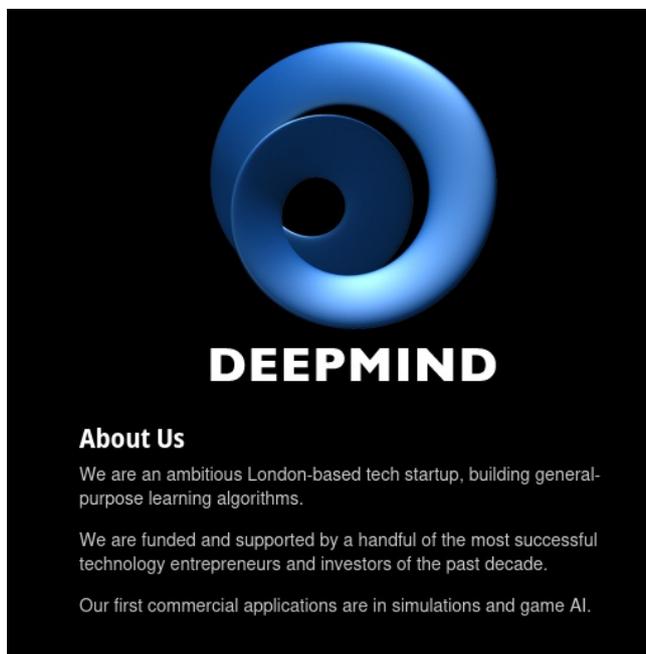
# DEEP BUBBLE

**L**a décennie 2010-2020 voit les acteurs économiques entrer dans une transformation profonde de leurs usages numériques.

L'envergure de cette transformation rappelle l'utilisation massive des outils numériques, initiée dans les années 1960s-1970s. A l'époque, l'introduction des mini-ordinateurs donne accès à une puissance de calcul et une capacité de traitement à des coûts plus abordables que les *mainframes* IBM. La transformation actuelle concerne l'utilisation systématique des outils décentralisés (*SaaS*), qui n'est pas sans rappeler l'architecture client/serveur du début de l'informatique d'entreprise, et l'exploitation massive des données opérationnelles.

Ces deux sujets sont une conséquence directe de l'ouverture du web pour le grand public et de la massification des usages, commencé dans les années 2000. Les survivants de la bulle internet (*dot.com*) sont devenus les nouveaux acteurs dominants. Google (Alphabet), Facebook (Meta), Amazon en sont les symboles. Microsoft, IBM et Apple se sont réinventés. La Chine a créé ses propres géants. Leur point commun ? Un modèle économique basé sur la manipulation de symboles et l'utilisation massive et systématique des décisions algorithmiques. Leur forte rentabilité vient d'une position commerciale hégémonique. Leur image de marque très flatteuse comme principaux acteurs de l'innovation informatique, et maintenant en intelligence artificielle, finalise le tableau. Les BigTechs sont des modèles.

La situation actuelle s'inscrit donc dans un mouvement continu. A lire la presse économique des dernières années, les uniques relais de croissance pour les acteurs économiques (entreprises, états) sont liés à l'innovation numérique, cette fameuse tech, et supposent la constitution et l'exploitation de vastes bases de données. Les mots-clés sont connus : BigData, puis data science, et maintenant *deep learning*.



## DEEPMIND, UN SYMBOLE

DeepMind est une petite société anglaise fondée en 2010 par des chercheurs en IA. Son équipe de 25 personnes est spécialisée dans une activité assez particulière : elle apprend à des programmes à jouer à des jeux vidéo. On peut sourire dans un premier temps, surtout si on sait que les jeux sont des antiquités des années 1980 comme Pong, Breakout ou Space Invaders. La publication scientifique date de 2015. On ne sourit plus quand on apprend que cette société est achetée par Google en 2014 pour 650 millions de dollars.

DeepMind est un symbole. Les jeux vidéo auxquels ses ingénieurs s'intéressent sont des mondes fermés où les règles sont limitées. Ce sont des terrains de jeu parfait pour les réseaux de neurones. La performance remarquable de DeepMind est d'avoir mis au point des techniques d'apprentissage qui s'adaptent au jeu rencontré. Le programme découvre le jeu et apprend à le maîtriser en observant les images affichées à l'écran et le score. Aucune autre information n'est fournie, et c'est un point crucial. Le programme découvre Pong, et au bout de quelques dizaines d'heures de jeu, atteint un niveau humain.

Ces techniques sont de véritables tours de force en ingénierie et en recherche appliquée. Un éventail de raffinements sont ajoutés à l'arsenal technologique de l'apprentissage profond au fil du temps, par les nombreux acteurs du domaine souvent proches des GAFA.

## LA BULLE GONFLE

Les principaux acteurs académiques des réseaux de neurones profonds rejoignent les GAFA. Geoffrey Hinton travaille pour Google depuis 2013. La même année, son ancien thésard Yann LeCun, est recruté par Facebook. Les deux recevront le prix Turing avec Yoshua Bengio en 2018. Nous plaçons le début de la bulle entre 2014 et 2016, année où AlphaGo fait les gros titres. En 2023, le buzz est toujours assourdissant.

## DES SUCCÈS IMPRESSIONNANTS

Les applications des réseaux de neurones sont impressionnantes dans de nombreux domaines. Les réseaux de neurones sont meilleurs que les humains pour la classification d'images. La traduction de texte en ligne progresse à un niveau correct pour les langues les plus représentées dans les corpus électroniques. La génération d'image et le text-to-image est une avancée majeure, avec Dall-E (2021) et MidJourney (2022). Des agents conversationnels intéressants sont disponibles, comme GPT-3 (2020) et ChatGPT (2022).

Les cas d'usage sont nombreux, probablement plus nombreux que ceux immédiatement visibles - mais qui marquent les esprits. De manière générale, les réseaux de neurones profonds sont des outils très utiles dans la détection et l'exploitation de tendances et motifs dans les bases de données massives, notamment les images. On s'attend à des progrès à long terme sur les sujets liés à l'industrie et, naturellement, les services numériques basés sur le profilage et la recommandation.

Aux échecs, AlphaZero de DeepMind montre de très belles performances face aux programmes spécialisés comme Stockfish, qui intègre maintenant des réseaux de neurones dans ses heuristiques de calcul. En 2016, DeepMind bat bruyamment et brillamment des champions humains au go. En 2019, DeepMind s'attaque à StarCraft 2 et propose un programme, DeepStar, qui égale les meilleurs joueurs humains.

Notons que pendant ce temps, les autres algorithmes d'apprentissage automatique (dont font partie les réseaux de neurones) sont tout à fait absents de la scène médiatique.

Notre méthode est plus performante qu'un expert humain sur **BreakOut**, **Enduro** et **Pong** et il atteint des performances proches de l'humain sur **Beam Rider**.

[Nous sommes] loin des performances humaines sur des jeux [plus difficiles] comme **Q\*bert**, **Seaquest**, **Space Invaders**, [qui] demandent au réseau de trouver une stratégie à long terme.

DeepMind (2013)

---

# LES LIMITES



## UN PAS VERS L'IA GÉNÉRALE ?

**E**n 2020, DeepMind publie Agent57, un agent qui surpasse les performances humaines sur 57 jeux disponibles sur Atari 2600. Cet agent prolonge et étend les performances obtenues en 2015.

Les réseaux de neurones ne sont pas adaptés à une IA générale. Si l'augmentation de la puissance des ordinateurs actuelle permet aux programmes de traiter beaucoup plus de neurones qu'auparavant, les limites dures ne bougent pas. Les réseaux de neurones sont différents des cerveaux sur d'innombrables points importants - et certains ne sont même pas encore connus !

## ÉTROITESSE, SPÉCIALISATION, CONSOMMATION

Les réseaux de neurones ne sont pas parfaits, loin de là. Les réseaux de neurones sont opaques. L'extrême complexité de l'architecture des systèmes IA construits par DeepMind, Meta, Google et d'autres, est un obstacle majeur à la compréhension des processus d'apprentissage. On empile des réseaux de neurones différents, chacun composé de multiples couches interconnectées (ou pas). Les architectures actuelles des agents conversationnels basés sur GPT-3 comptent jusqu'à 175 milliards de paramètres. On ne sait pas comment est encodée l'information dans ces outils. On ne sait pas comment ils calculent leurs résultats. On ne sait pas entrer dans la boîte noire qu'on entraîne, ou plutôt la base de données qu'on engendre. L'ingénieur IA travaille en aveugle, à partir d'exemples, et tente de construire une machine en empilant des légos dans le noir. Seules les performances comptent. La fin justifie-t-elle les moyens ?

Les réseaux de neurones sont spécialisés. On parle d'IA étroite, en opposition avec l'IA générale décrite plus haut. Un système construit par apprentissage automatique sait généralement effectuer une seule tâche. S'il sait jouer à Pacman, il ne saura pas jouer à Pong. S'il sait catégoriser des images d'animaux, il ne saura rien dire d'utile des voitures. Il faudra dans le meilleur cas ré-entraîner le système, et dans le pire des cas construire un nouveau système IA avec des traitements spécifiques ; on comprend encore mieux la puissance de AlphaZero, qui peut apprendre différents jeux sans être modifié. Même s'il maîtrise 1000 jeux, on doute qu'il soit bon sur le 1000 et unième. Les réseaux de neurones profonds sont des outils particulièrement fragiles. Jetables, oserons-nous dire.

Les réseaux de neurones sont gourmands. Gourmands en données, car les jeux de données d'entraînement nécessaires sont souvent énormes, avec des millions d'exemples. La constitution de ces bases implique fréquemment un travail humain d'étiquetage fastidieux, sinon abrutissant. Gourmands en énergie, car la puissance de calcul requise pour évaluer les milliards de paramètres lors du processus d'apprentissage est très importante. Cette puissance de calcul est fournie par des processeurs spécialisés (graphics processing unit, tensor processing unit), rares et coûteux. On parle de 190 000 kWh pour entraîner le modèle GPT-3 (large language model), soit un aller-retour pour la Lune en voiture. La question de l'utilité doit être posée.

[Ce programme] Atari a suscité l'enthousiasme – et mérité une publication dans Nature – en partie parce qu'il semblait être un pas vers l'intelligence artificielle. Un seul algorithme, n'utilisant pas une représentation spécifique de la connaissance, a développé un large éventail de compétences sur des tâches très variées impliquant une entrée en dimension assez grande. Aucun programme n'avait fait cela auparavant.

Cependant, une AGI complète ferait beaucoup plus. S'il est difficile de construire des systèmes IA spécialistes très performant, il est beaucoup plus difficile de construire un système d'IA généraliste.

L'apprentissage profond n'est pas la solution : ses aficionados admettent que “de nouveaux paradigmes sont nécessaires” pour le combiner avec un raisonnement complexe. C'est une formule classique en recherche, pour dire “nous n'avons pas la moindre idée de la suite”.

pourquoi la plupart des chercheurs en IA ont abandonné cet espoir initial, se tournant plutôt vers la résolution de tâches très spécialisées – souvent avec un succès spectaculaire

## ET LES AUTRES ?

Ces caractéristiques ne sont pas propres aux réseaux de neurones. Elles peuvent s'appliquer à d'autres algorithmes d'apprentissage automatique. La spécialisation est commune.

L'opacité est fréquente. Comprendre le processus de décision dans un arbre de décision est possible, mais rien ne garantit que les règles fassent sens pour un non spécialiste. Certains algorithmes (dits frugaux) ont des approches qui minimisent le nombre de paramètres utilisés en entrée, permettant ainsi de réduire le nombre de variables explicatives.

La gourmandise énergétique n'est pas une fatalité en soi. Elle est souvent liée à la quantité de données à intégrer. Il existe des méthodes qui peuvent fournir des résultats utiles à partir de petites bases d'entraînement.

“Par rapport au cerveau, les ANN sont trop propres, trop simples, trop limités et trop secs.

Trop propres, parce que les réseaux construits par l'homme privilégient élégance et puissance mathématiques, ce qu'on ne retrouve pas dans les cerveaux biologiques.

Trop simples, parce qu'un seul neurone est aussi complexe sur le plan informatique [...] qu'un petit ordinateur. Et il existe une trentaine de types de neurones différents.

Trop limités, car même les ANN comportant des millions de neurones sont minuscules comparés aux cerveaux humains. Et trop sec, parce que les chercheurs ignorent généralement [beaucoup de comportements biologiques].

Margaret Boden (2016)

---

# LE DEEP LEARNING EST UN ECHEC



## UN ÉCHEC SCIENTIFIQUE

**L**e *deep learning* est un échec scientifique, car la discipline de l'IA n'a pas progressé vers l'IA générale. La culture qui prévaut chez les spécialistes des réseaux de neurones est une culture de l'efficacité, doublée d'une tendance à construire de gros systèmes. On ne retient, pour l'instant, que la capacité à construire des outils d'aide à la décision qu'on sait rendre performants dans certaines conditions, et dans certains cas. C'est appréciable. Cependant, peut-on mieux qualifier l'intelligence en général ? Sait-on modéliser efficacement un processus d'apprentissage ? Comprend-on mieux le cerveau humain ? A-t-on la moindre idée de ce qui se passe quand un réseau profond apprend à différencier un chat d'un pneu ? Les réponses, si elles sont positives, n'ont pas émergé des laboratoires.

## UN ÉCHEC INTELLECTUEL

**L**e *deep learning* est un échec intellectuel. Certains auteurs parlent de Big Data AI, pour marquer le lien avec la mode technique précédente. Au-delà de l'approche purement descriptive et réactionnaire, cette réticence à faire appel à une modélisation mathématique, sinon scientifique, s'apparente à une démission intellectuelle. Le calcul des caractéristiques (*features*) relève plus de la cuisine que d'une réelle démarche scientifique. Utiliser un modèle permet de réduire la dimensionnalité des problèmes et, ainsi, contribue habituellement à des systèmes moins gourmands en ressources (données, calcul, mémoire). Les données ne parlent que si on les leur pose les bonnes questions. De plus, l'hypothèse d'un apprentissage sur des données issues du passé pour une utilisation dans le futur : c'est restrictif et simplement faux quand les résultats concernent les humains. Couplé à l'opacité intrinsèque des réseaux de neurones, le système se retrouve à calquer aveuglément des comportements passés. Cette situation est très certainement problématique, dans un cas général.

## UN ÉCHEC SOCIAL

**L**e *deep learning* est un échec social. DeepBubble a commencé depuis huit ans : quels sont les bénéfices concrets apportés par cette technologie pour l'humanité ? Pour les travailleurs ou employés, col blancs ou cols bleus ? La question semble peut-être exagérée, voire ridicule. Mais à reprendre les innombrables communications, dont celles de la puissance publique, les budgets engagés et les financements publics consentis, il faut s'y intéresser. A ce jour, force est de constater que les retours pour le public sont maigres. On pourrait même y voir un moyen de légitimer la puissance des acteurs majeurs du numériques, principalement états-uniens et chinois, qui bénéficie de la manne publique pour innover à peu de frais. Il faut, en passant, mentionner l'esclavage indirect des populations pauvres, à qui est confié les très pénibles tâches d'annotation des bases de données nécessaires aux algorithmes d'apprentissage automatique. Voire les micro-tâches réalisées par des humains pour les services faisant prétendument appel à l'IA, mais n'étant en réalité qu'un travail humain déguisé - on parle d'ailleurs d'intelligence artificielle artificielle. *Amazon Mechanical Turk* était honnête.

## LE CONSTAT EST SÉVÈRE

Il n'est pas lieu de condamner individuellement les ingénieurs ou scientifiques qui travaillent sur ces techniques. Le propos est systémique, il concerne la communauté, les financiers, les décideurs politiques et économiques. C'est essentiellement un problème d'inculture scientifique et de méconnaissance technique.

Investir des sommes importantes sur des technologies très complexes, coûteuses, dispendieuses en ressources et assez probablement inadaptées dans 90% des cas réels, c'est une erreur stratégique. Doublé d'une erreur technique flagrante. En faire l'alpha et l'omega des sujets d'innovations publics et privés est un non-sens.

---

# WINTER IS COMING



## INÉLUCTABILITÉ

**T**ous les éléments sont réunis pour que la DeepBubble laisse place à un nouvel hiver de l'intelligence artificielle. L'attention médiatique est forte, les attentes économiques sont fortes, l'intérêt du politique est fort. La question de la rentabilité économique se pose. Ces outils apportent-ils de la valeur ? Sont-ils utiles ?

Pour les BigTechs, la réponse est oui. Ils ont construit leur succès autour de l'exploitation des données, à des fins principalement publicitaires pour Google et Meta. Amazon utilise massivement les algorithmes de profilage et de recommandation pour sa place de marché. Les organisations qui disposent de données massives et qui ont un intérêt à les traiter, que ce soit pour des contraintes opérationnelles ou pour proposer de nouveaux services, peuvent tirer profit de ces outils. Si elles arrivent à embaucher et garder de bons techniciens.

Les professionnels sont rares, surtout lorsqu'il s'agit de techniciens expérimentés. Construire un réseau de neurones profond capable de traiter de manière satisfaisante des téraoctets de données est extrêmement difficile.

Les équipes sont très peu nombreuses. Pour ne rien arranger, les BigTechs accaparent les ingénieurs IA, profitant de leur image de marque et de salaires attractifs. Nos échanges avec les entreprises et industriels français vont tous dans ce sens : les membres du CAC40 recrutent difficilement. Il y aurait sûrement des parallèles historiques à faire, dans d'autres domaines.

## QUEL SERA LE DÉCLENCHEUR ?

Les prédictions sont difficiles, surtout quand elles concernent le futur, qu'on soit Niels Bohr ou Yogi Berra. Voici notre sentiment.

Le coup d'arrêt ne viendra pas de l'intérieur du domaine. Les spécialistes de l'IA, dont nous ne faisons pas partie, qui portent un message critique sur les réseaux de neurones ne sont pas visibles médiatiquement, sauf peut-être Gary Marcus (New York University). Le mode de financement de la recherche occidentale rend un tel revirement très improbable : trop risqué, trop extrême.

Le coup d'arrêt viendra de l'extérieur, sous la forme d'un scandale, d'un événement avec une charge symbolique forte. Plus forte que Cambridge Analytica, par exemple... Un scandale mondial sur un service qui délègue ses décisions à un algorithme ? Aux Etats-Unis ou peut-être en Chine ? Des gens doivent mourir, en masse, pour que l'opinion publique réagisse. La chute d'un des magnats de l'innovation comme Elon Musk ? La fin d'un des GAFA, Meta semblant mal embarqué dans son pari de Metaverse ? Difficile à dire.

Et si le coup d'arrêt ne venait pas ? Le désintérêt du public pourrait venir d'une forme de lassitude : l'homme de la rue a compris que les réseaux de neurones sont des outils puissants, c'est acté. Le désintérêt du public pourrait aussi venir d'une déconnexion des avancées avec son expérience quotidienne : ce qui paraissait auparavant époustouflant est maintenant classique, comme "converser" avec un perroquet stochastique (=un chatbot). Les entreprises, lassées de payer (cher) des ingénieurs sans générer suffisamment de retours, pourraient stopper les frais.

Sans vouloir jouer les oiseaux de mauvais augure, nous ne sommes pas très optimistes sur le futur du domaine.

## QUE RETENONS-NOUS ?

Il est tentant de proposer une lecture psychologique de la bulle.

Marvin Minsky est mort en 2016, au tout début de la DeepBubble. De nombreuses personnes le tiennent responsable du long coma du mouvement connexionnisme dans les années 1970 et 1980. Les réseaux de neurones ont eu leur vengeance... et leur véritable moment de gloire, après les perceptrons et le moment PDP. Au prix d'une consommation énergétique délirante, d'une opacité totale de la décision algorithmique et d'une démission intellectuelle. Pour quelle avancée théorique majeure ? Aucune.

La DeepBubble a fourni les outils permettant d'analyser les données du web. C'est un mouvement d'ingénierie des données extrêmement complexe, qui a pris le parti de ne pas se préoccuper du sens et de confier ces tâches à des algorithmes. La décision algorithmique n'est acceptable que si elle a une base humaine. Les difficultés d'applicabilité des réseaux de neurones, et plus généralement des algorithmes d'apprentissage automatique qui ont un fonctionnement opaque, sont réhivitoires à nos yeux pour un usage à long terme qui soit pérenne et serein. Ce constat est direct, et devrait être partagé par les pouvoirs publics.

« Ne travaillons surtout pas avec les gens qui connaissent le problème, cela pourrait introduire des biais dans l'apprentissage » : cette recommandation étonnante, nous l'avons rencontrée plus d'une fois au sein d'équipe de data scientists.

Le paradigme de la naïveté ? Non. Plutôt une erreur fondamentale et historique.

## ET APRÈS ?

De nombreux sujets buzzent autour de l'IA. Citons les inévitables préoccupations autour de l'éthique des algorithmes, des biais dans les données et la très vague inclusion. Ces sujets sont légitimes, mais relèvent-ils vraiment de la technique ? L'inclusion et la diversité sociale des développeurs est un sujet social, voire éducatif. Les biais dans les données d'entraînement sont naturels, ils sont documentés par les études statistiques que doivent mener les data scientists avant de les injecter dans un modèle, quel qu'il soit. Quant à l'éthique relative à l'impact des décisions des algorithmes, n'aurait-on pas oublié que les programmes ne s'écrivent pas tous seuls ?

Ce sont des problématiques importantes qui doivent être adressées. Nous notons que la mauvaise communication autour de l'IA entretient la confusion et le flou conceptuel, et permet ainsi de continuer à discuter en termes vagues et imprécis.. Nous y voyons surtout une bulle qui s'auto-alimente en enfilant les déclarations de bonne volonté teintée d'eau tiède et n'aboutissant finalement qu'à des impasses. Comment discuter d'éthique de la déci-

sion algorithmique si les algorithmes sont des boîtes noires dans lesquelles on injecte n'importe quoi?

Comme à chaque cycle, les technologies émergentes vont s'industrialiser et se normaliser. Les bonimenteurs iront vers d'autres marchés (coucou la #VR) et trouveront de nouveaux pigeons. Ces technologies iront compléter la boîte à outils des ingénieurs et scientifiques spécialistes de l'analyse des données. Le monde économique partira sur d'autres voies, peut-être vertueuses, comme la réduction de l'impact écologique de nos activités.

Concernant l'intelligence artificielle, la seule question intéressante à nos yeux est la suivante : quelle avancée théorique permettra d'utiliser efficacement la puissance de calcul disponible et les données accumulées pour, enfin, progresser sur le chemin de l'intelligence générale ?

**FIN**

---

# RÉFÉRENCES PRINCIPALES

**Mind as a machine**, Margaret Boden (Oxford, 2006). Une référence absolue sur l'histoire des sciences cognitives, le contre-exemple parfait du livre de plage. A la fois illisible et indispensable. Tous les domaines sont abordés, des premiers pas dans le calcul automatique à la vie artificielle, en passant par l'apprentissage automatique, les réseaux de neurones et le traitement du langage. 1500 pages de texte, 130 pages de références, le travail d'une vie.

Deux autres historiques de l'intelligence artificielle sont à conseiller : **AI : Its Nature and Future** (Oxford, 2016) de la même Margaret Boden, qui résume le domaine à grands traits dans un style direct et simple à lire, et le plus classique **The Quest for Artificial Intelligence** (2010) de Nils Nilsson.

**A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going** (Oxford, 2021) de Michael Wooldridge se concentre sur l'apprentissage automatique des dernières années, en donnant les éléments de compréhension nécessaire. L'auteur dépasse le cadre technique en élargissant sa réflexion. Très agréable à lire.

**The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do**, Erik J. Larson (2021) tient plus de l'essai que du panorama historique, mais amène une réflexion critique sur les développements actuels en apprentissage automatique.

**Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence** (Yale, 2021) de Kate Crawford propose un tour d'horizon géographique sur l'industrie de l'apprentissage automatique actuel, dans sa version réseaux de neurones essentiellement. Matériaux, énergie, données, applications : l'ensemble des sujets sont abordés, notamment la provenance des jeux de données utilisés habituellement.

**Computer: A History of the Information Machine**, de Martin Campbell-Kelly, William Aspray, Nathan Ensmenger et Jeffrey R. Yost. (Routledge; 4th ed, 2023)

**A New History of Modern Computing**, de Thomas Haigh et Paul E. Ceruzzi (MIT Press, 2021)

---

# QUI SOMMES-NOUS ?



**AltGR** est une entreprise créée en 2018 à Aix-en-Provence par quatre associés. Deux des associés ont un parcours académique qu'on retrouve rarement chez les fondateurs : Thomas Gerbaud est ingénieur centralien de Lyon et docteur en physique, et Antoine Gerbaud est ancien élève de l'ENS de Cachan et mathématicien. Ils ont tous deux une expérience de 10 ans en recherche, Thomas dans l'étude des plasmas de fusion qu'on observe dans les grandes installations expérimentales de confinement magnétique (tokamaks), Antoine dans la théorie des graphes probabilistes. Cette carrière scientifique leur a permis de développer une expérience solide dans le traitement scientifique des données, qu'ils appliquent depuis aux problématiques des entreprises.

**AltGR** a travaillé pour la banque, la finance, le transport, l'aéronautique et les laboratoires de R&D publics et privés. AltGR construit aussi des solutions informatiques pour l'Etat. Depuis leurs débuts dans le monde de la R&D externalisée en 2014, l'équipe technique a réalisé près de 80 projets pour plusieurs dizaines de clients différents, de la startup au grand groupe.

Nous avons donc une expérience de l'ensemble des techniques de traitement de données, dont la réalisation d'études statistiques, la conception de systèmes experts, l'utilisation d'algorithmes d'apprentissage automatique et la création d'heuristiques de calcul en recherche opérationnelle. Nous avons accompagné l'émergence du mouvement BigData et l'irruption fracassante de l'intelligence artificielle dans les problématiques d'innovation numérique des entreprises. Nous sommes donc des acteurs de cette "révolution", et des témoins privilégiés de ce moment puisque nous construisons les outils intelligents qui font l'actualité. Ce texte est issu de notre observation du réel.

Site web : <http://alt-gr.tech>

Thomas > Merci aux data scientists et dresseurs de données de l'équipe, Dr Antoine Gerbaud, Dr Sabine Nasr et Dr Olivier Kauffmann pour leur aide. Les erreurs et imprécisions qui subsisteraient sont de moi. Merci à Mathieu Jacq d'avoir initié ce texte lors de son stage de fin d'étude.

Contact : @tgerbaud (twitter) ou thomas(arobase)alt-gr.tech