



ROYAUME DU MAROC  
ACADÉMIE HASSAN II DES SCIENCES ET TECHNIQUES

# Renaissance et promesses actuelles de l'intelligence artificielle

**Jean-Gabriel GANASCIA**

Professeur à l'Université Pierre et Marie Curie - Paris  
Président du comité d'éthique du CNRS - France

Rabat, le 26 mars 2018

Dépôt légal : 2018MO2876  
ISBN : 978-9954-520-14-7

Réalisation : **AGRI-BYS S.A.R.L.**

Achévé d'imprimer : juillet 2018  
Imprimerie Lawne : 11, rue Dakar, Océan, 10040-Rabat, Maroc



**Sa Majesté Le Roi Mohammed VI - que Dieu Le garde -  
Protecteur de l'Académie Hassan II  
des Sciences et Techniques**





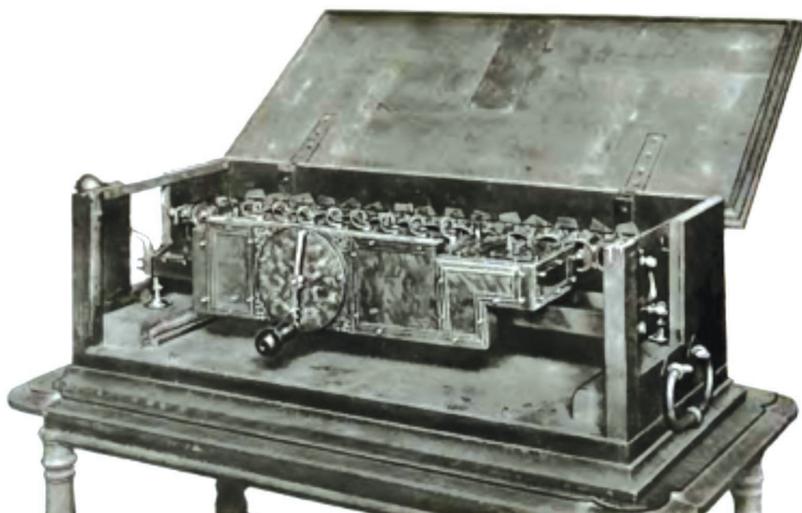
Merci de m'avoir invité à parler d'intelligence artificielle. Je vais commencer tout de suite mon exposé sur «*la renaissance et les promesses actuelles de l'intelligence artificielle*» et puis je parlerai ensuite des questions ouvertes, des questions de société liées au développement de l'intelligence artificielle.

*«Je vais commencer par l'histoire de cette discipline et montrer que ce n'est pas une question neuve que celle que pose l'intelligence artificielle. C'est même une question qui a plusieurs siècles»*

Il s'agit de savoir si une machine peut penser? C'est étrange mais même si l'on peut être troublé à l'idée qu'une machine pense, c'est pourtant une idée ancienne qui a été émise depuis longtemps par un certain nombre de philosophes. Plus exactement, il y a là la consécration de 2 hypothèses qui sont au départ toutes les deux problématiques et puis qui, au fil du temps, sont devenues de plus en plus naturelles :

- première hypothèse : «**est-ce qu'une machine peut faire des calculs?**»

Elle a été introduite très tôt d'abord avec la *pascaline* de Blaise Pascal au XVII<sup>e</sup> siècle, puis au début du XIX<sup>e</sup> siècle avec le moulin de Charles Babbage qui est l'ancêtre des ordinateurs actuels. Aujourd'hui, tout le monde en est convaincu mais cela ne semblait pas évident à l'époque et ça a mis un certain nombre de siècles à s'imposer.



Machine de Pascal

- deuxième hypothèse : «**est-ce que la pensée se réduit à un calcul?**»

Hypothèse philosophique plus ancienne encore, elle a été émise entre autres par *Thomas Hobbes* et par *Leibniz* pour qui tout se produit par calcul dans la nature, car Dieu est un calculateur prodige.



Gottfried Wilhelm Leibniz (1646-1716)

### **Alors, qu'entend-on par calcul?**

Il faut voir avec le terme de calcul l'idée de symbole. L'étymologie du terme de calcul, vous la connaissez, renvoie au caillou (certains d'entre nous le savent, une expérience douloureuse (rire)). Donc si la pensée se réduit à un calcul et si les calculs s'exécutent tous sur des machines, alors les machines peuvent reproduire la pensée.

Leibniz a imaginé qu'on pouvait fabriquer des machines pour calculer et qu'ensuite on pourrait reproduire la pensée sur ces machines qui calculent. Il n'a pas été le seul. Il y a eu, au XIX<sup>e</sup> siècle, des gens qui ont imaginé des ordinateurs. J'ai cité *Charles Babbage*. Il convient maintenant de mentionner son assistante *Lady Ada Lovelace*. Je la cite parce qu'elle a été la première programmeuse de l'histoire et qu'elle a conçu des programmes pour cette machine qui n'a jamais été construite. Ce faisant, très vite, elle s'est dite que cette machine ne ferait peut-être pas uniquement des calculs mathématiques mais qu'elle pourrait aussi traiter des textes et de la musique,...



Lady Ada Lovelace, fille de Lord Byron.

(En son honneur, un langage de programmation important s'appelle l'ADA)

Cette question a été reprise quelques années après, dans la deuxième moitié du XIX<sup>e</sup> siècle par un économiste anglais, *Stanley Jevons*, qui a repris l'algèbre de Boole pour fabriquer un piano mécanique qui simulait des raisonnements.



Stanley Jevons (1835-1882) et son piano

Une autre personnalité qui a joué un rôle majeur dans la genèse de l'intelligence artificielle est le très célèbre *Alan Turing*. Rappelons qu'il a eu une existence étonnante puisqu'il a commencé d'abord à poser les fondements théoriques de l'informatique avec les fameuses machines de Turing (il ne les avait pas appelé ainsi car c'était un homme très modeste), qui aidaient à mettre en évidence les limites de tous les automates à nombre fini d'états discrets.



Alan Turing (1912-1954)

Pendant la guerre, *Turing* a contribué à l'effort des alliés en aidant à décoder les messages allemands. Après la guerre, au moment où les premiers ordinateurs électroniques venaient d'être construits en 1947-48 (le premier a été construit en 1946), il s'est demandé si ces machines pouvaient penser et a écrit deux articles très remarquables sur ce sujet dont un en 1950 dans la revue MIND.

VOL. LIX. No. 236.]

[October, 1950

MIND  
A QUARTERLY REVIEW  
OF  
PSYCHOLOGY AND PHILOSOPHY

— ❦ —  
I.—COMPUTING MACHINERY AND  
INTELLIGENCE

By A. M. TURING

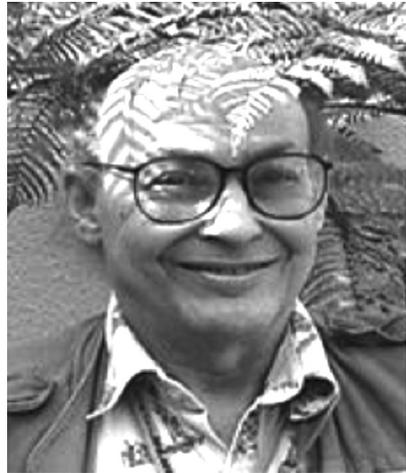
Dans cet article, Alan Turing se demande ce que pourrait signifier l'expression «*une machine pense*». Sans rentrer dans les détails de ces deux articles extrêmement riches qui prendraient beaucoup de temps à commenter, notons que le terme IA n'y est pas prononcé. Curieusement, alors que *Turing* est souvent vu comme l'un des précurseurs de l'IA, il n'en a jamais parlé.

**«L'intelligence artificielle est née en 1955, lorsque 2 jeunes mathématiciens imaginent que les ordinateurs peuvent être utilisés pour mieux comprendre l'intelligence»**

La discipline elle-même naît en 1955 et le terme est prononcé à ce moment-là par 2 jeunes gens qui ont 28 ans à l'époque, qui sont des mathématiciens, et qui imaginent que ces nouveaux objets que sont les ordinateurs peuvent être utilisés pour mieux comprendre l'intelligence.



John McCarthy (1927-2011)



Marvin Minsky (1927-2016)

Il ne s'agit pas, bien sûr, de prendre le terme d'IA au sens littéral d'une intelligence qui serait produite juste par l'art des hommes mais au contraire de voir dans l'IA une discipline scientifique qui va permettre de mieux comprendre l'intelligence en la décomposant en fonctions cognitives élémentaires et en simulant ces fonctions cognitives.

J'ai mentionné ici (ci-dessous) le texte du projet écrit en 1955 par ces 2 jeunes gens qui le soumettent à la Fondation Rockefeller pour organiser une école d'été sur cette nouvelle discipline appelée IA.

*We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it.*

Ils devaient être assez talentueux puisqu'ils ont réussi à convaincre 2 personnes, l'une étant le directeur scientifique de la société IBM et la seconde n'étant autre que le très grand Claude Shannon, l'homme de la théorie de l'information.

Ces 4 personnes cosignent ce projet et, en souligné dans le texte, ce qui est important repose sur une conjecture. Une conjecture est une proposition toujours avérée, mais

dont la véracité n'a pas encore été démontrée. On va la poser et elle va jouer le rôle d'horizon régulateur pour cette science nouvelle qui naît. D'une certaine façon, cette conjecture joue le même rôle que, pour Galilée, la conjecture selon laquelle la nature s'écrit en termes mathématiques.

**«L'intelligence artificielle est la discipline scientifique qui a le plus changé le monde ces 60 dernières années»**

Ça signifie que toutes les composantes de l'intelligence pourraient être isolées et réduites à des modules si élémentaires qu'une machine pourrait être construite pour les simuler un à un. Vous voyez que cela ouvre un horizon, un champ de travail vraiment considérable. Je crois qu'il faut vraiment avoir cela en tête pour comprendre ce qu'est l'IA.

Dernièrement en France, nous avons eu beaucoup de débat puisqu'il y a eu d'une part la «stratégie Intelligence Artificielle» qui avait été lancée par *Axel Lemaire* il y a un an et ensuite, avec l'arrivée d'*Emanuel Macron* au pouvoir, la mission confiée à *Cedric Vilani* qui va restituer ses conclusions en fin de semaine. Et donc je crois que ce qui est très important, c'est de bien comprendre cette définition parce qu'il y avait beaucoup de malentendus même dans la communauté scientifique sur la définition de l'IA. Les gens n'ont pas tous en tête cette vue historique qui clarifie à mon sens le concept lui-même et qui montre qu'il a une très grande extension. Et c'est ce que je vais démontrer. Avant cela, je voudrais montrer que l'IA a changé le monde.

L'intelligence artificielle est la discipline scientifique qui a le plus changé le monde ces 60 dernières années et ce n'est pas pour vanter ma propre discipline que je dis cela. Je vais vous le montrer objectivement à partir d'un certain nombre d'exemples :

- Le premier a changé la vie de tous les jours. C'est **le web**, qui est le couplage de réseaux de télécommunications qui ne sont pas à proprement parlé de l'IA, avec un modèle de mémoire qui s'appelle l'hypertexte. Vous savez très bien que le langage d'écriture des pages web c'est HTML (*HyperText Markup Language*) et le protocole de communication c'est HTTP (*HyperText Transfer Protocol*). Vous voyez donc que l'hypertexte est au centre du web. Or, l'hypertexte est une notion qui a été introduite en 1965 par un jeune homme de l'époque qui s'appelle *Ted Nelson*, qui l'a programmé avec des techniques d'IA et publié dans une conférence d'IA. C'est un modèle de mémoire qui est construit sur cette idée selon laquelle on peut avoir sur un texte des connexions entre les parties du texte. L'idée de *Tim Berners-Lee* lorsqu'il a conçu le web est que ces connexions ne sont pas limitées à une machine mais qu'elles peuvent s'étendre sur le réseau.
- Ensuite les **robots**, dont on entend beaucoup parler ces derniers temps. Les robots ne sont pas que de l'IA. Ce sont des capteurs (c'est de la physique) et des effecteurs (c'est de la physique et de la mécanique), mais l'IA est là pour interpréter les informations qui viennent des capteurs et pour arriver à décider des actions à accomplir et à mettre en œuvre les effecteurs. On y reviendra.

- Les robots virtuels (les bots) qui jouent un rôle important aujourd'hui dans la bourse par exemple, c'est de l'IA. De même que les véhicules autonomes (ce sont aussi des robots en vérité), la biométrie (reconnaissance d'empreintes digitales), la vision (reconnaissance des visages ou des vidéos), la reconnaissance vocale avec des systèmes comme SIRI, le traitement et la compréhension du langage naturel.

**«Le schéma expérimental classique qui existe depuis le début de la science moderne est aujourd'hui mis en cause par l'intelligence artificielle»**

- Je termine par les e-sciences. Vous savez que ces dernières années, il y a eu une révolution silencieuse dans le domaine scientifique. Si l'on reprend la démarche scientifique telle qu'elle a été inaugurée par Galilée, une démarche expérimentale qui part d'une hypothèse qui doit être validée par un dispositif matériel. Expérience, ça vient de péril : on va mettre en danger la théorie et pour ça, on va construire un dispositif matériel, recueillir des observations et les confronter avec ce qu'anticipe la théorie. Bien sûr s'il n'y a pas adéquation parfaite, on va revenir sur la théorie. Ce schéma, qui est le schéma expérimental classique depuis le début de la science moderne est aujourd'hui mis en cause par l'IA. Pourquoi? Parce que nous avons aujourd'hui des capteurs qui génèrent d'énormes quantités de données et on fait des expériences non pas dans le monde extérieur mais directement sur ces données. C'est ce qu'on appelle «*les expériences in silico*», qui donnent naissances «*e-sciences*» et qui sont directement lié à l'IA puisque grâce à l'IA l'on est en mesure de générer automatiquement des théories et de les tester sur d'immenses quantités de données.

Donc voilà quelques unes de ces transformations, récentes, qui sont liées à l'IA. Bien sûr, la physique a eu une très grande importance mais c'est plutôt la physique des années 30, peut-être un peu plus récemment bien sûr (voyez, le principe du laser est très ancien même s'il n'a été mis en œuvre que beaucoup plus tard). De même, la biologie est tout-à-fait centrale mais elle a eu beaucoup moins d'effets concrets. L'augmentation de l'espérance de vie est plus liée au progrès de l'hygiène qu'à la compréhension de mécanismes biologiques même si, heureusement, il y a eu de nouveaux médicaments qui sont liés aux progrès de la biologie.

## **Panorama général («IA at large»)**

Pour essayer de comprendre le champ de l'IA dans toute son ampleur, je vais reprendre ce que j'ai dit avant : *l'IA essaie de reproduire les différentes fonctions cognitives et ensuite de les simuler*. Pour mieux comprendre cela, je prends une décomposition des fonctions cognitives tout-à-fait classique dans le champ des sciences cognitives en 5 grandes catégories :

1. Les *fonctions réceptives* : elles autorisent l'acquisition, le traitement, la classification et l'intégration de l'information.
2. La *mémoire et l'apprentissage* permettant le stockage et le rappel de l'information.

3. Le *raisonnement, la pensée*. Cela concerne aussi l'organisation et la **réorganisation** mentale de l'information ainsi que son utilisation.
4. Les *fonctions expressives* qui rendent possible la communication.
5. Les *fonctions exécutives* de prise de décision et d'action.

**«L'intelligence artificielle a été en mesure de simuler  
5 grandes classes de fonctions cognitives»**

Je vous propose qu'on les examine une à une, pour voir à quel point l'IA joue un rôle important dans la simulation et à quel point cela peut avoir des applications pratiques (puisque'il y a 2 objectifs dans l'IA : le premier : comprendre l'intelligence; le second : une fois qu'on a simulé la fonction cognitive, on peut avoir des applications pratiques qui sont liées à la mise en œuvre de cette simulation des fonctions cognitives) :

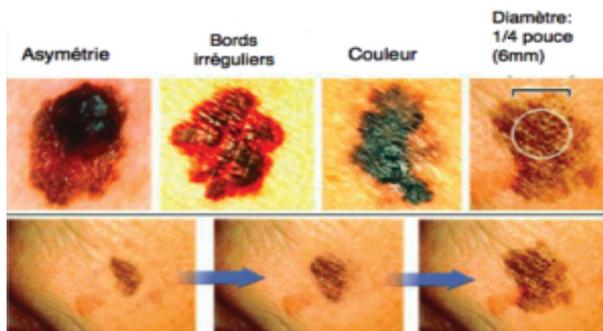
- Les 3 premières catégories correspondent un tout petit peu à une vision classique qui a été introduite par *Aristote* sur l'âme (aujourd'hui on ne parle plus d'âme, on n'ose plus en parler). Ce n'est pas l'âme au sens spirituel mais au sens rationnel, au sens de ce qui anime. Qu'est-ce que cela veut dire? Que vous avez une entité abstraite, qui a des connaissances sur le monde, qui a ce que nous appelons pour nous des sensations, ou pour des robots des capteurs. Et donc la première étape c'est quoi? C'est interpréter les informations qui viennent de ces capteurs. Et pour cela qu'est-ce qu'il faut faire? Il faut simuler ce qu'on appelle la perception, les **fonctions réceptives**, et construire à partir de ces capteurs une représentation, c'est-à-dire construire en mémoire une trace, mais qui soit une reconstruction du monde extérieur, qui nous présente le monde lorsqu'il n'est pas là autrement dit qu'il le re-présente, à savoir qu'il le présente en son absence. Voilà pour le premier point. Ensuite, pour construire cette représentation, il faut savoir comment organiser ces informations dans une **mémoire informatique**, qui est un dispositif de stockage d'informations. Or, dans notre tête, une mémoire c'est bien autre chose : ça nous permet de retrouver des éléments d'information avec une simple réminiscence, une association. Ça permet d'oublier, de généraliser, de construire des connaissances. Ce qu'on cherche à faire depuis longtemps en intelligence artificielle, c'est essayer de **simuler cette mémoire**. Nous avons vu que le web était une simulation de la mémoire. Il y a bien d'autres essais de simulation de la mémoire en intelligence artificielle et surtout qu'il y a des techniques d'**extraction de connaissances** à partir d'éléments en mémoire. Cela joue un rôle extrêmement important pour l'IA, si important aujourd'hui que l'on va s'appesantir un petit peu sur ce chapitre.
- Mais ces agents, ces âmes que j'ai mentionnées ici ne sont pas isolées. Pour être capable de bien comprendre comment elles communiquent entre elles, on essaye de simuler ces **fonctions expressives** ainsi que les fonctions d'analyse des échanges. C'est ce que l'on fait avec l'IA. On y reviendra.
- Enfin, dernier point : les fonctions exécutives qui permettent de prendre des décisions et ensuite d'agir.

**«Je vais passer en revue ces 5 grandes classes de fonctions cognitives pour montrer que l'IA a été en mesure de les simuler et à quoi cela correspond en pratique.»**

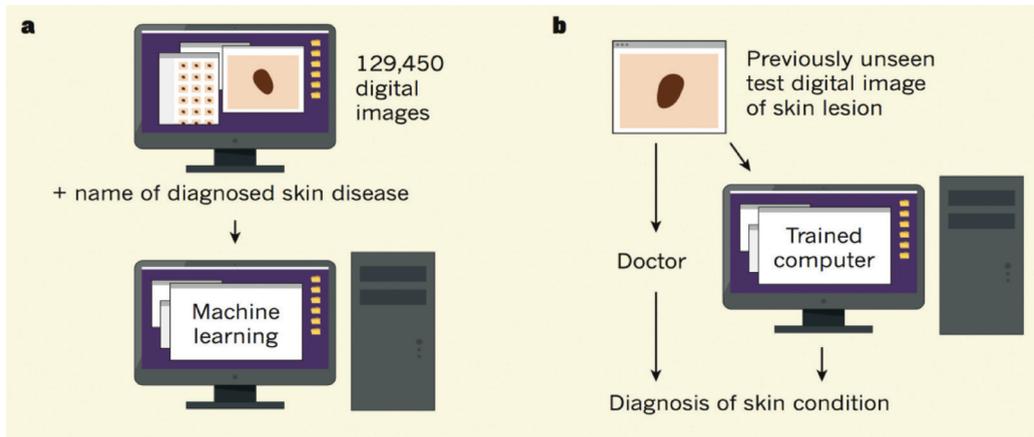
## 1- Fonctions réceptives

- Interprétation automatique d'images :
  - Reconnaissance des formes, reconnaissance des visages, identification d'objets, etc. On est capable aujourd'hui par exemple, à partir d'apprentissage machine, de construire un système qui fait de la reconnaissance des visages avec une performance de 99,63% sur 8 millions d'identités différentes. Pour cela, on a besoin de beaucoup d'images/personne et on apprend avec 200 millions d'images. Ce sont là des performances étonnantes qui dépassent les nôtres: un individu moyen reconnaît 3 à 500 visages différents;
  - Accès par le contenu dans des bases d'images. Par exemple vous cherchez une photo de Brigitte Bardot.
- Interprétation automatique vidéo :
  - Exemple : détection d'intentions (sur un quai de métro par exemple) ou de vortex en météorologie
- Traitement de la parole :
  - Reconnaissance pour la dictée automatique;
  - Reconnaissance de mots ou de thèmes;
  - Identification du locuteur, des accents et des langues.
- Analyse et fusion d'informations issues de capteurs :
  - Multi-modalités, ex. *désambiguïsation de l'image par le son*.
- Perception : construction d'une représentation.

Je donnerai juste un seul exemple d'un article publié il y a un an environ dans la revue Nature. Vous savez qu'il y a un cancer de la peau qui est extrêmement dangereux, le mélanome, qu'il faut diagnostiquer très tôt pour pouvoir l'enlever. Ce diagnostic demande une prévention permanente, ce qui est difficile.



On a alors imaginé un système qui, à partir de photos prises avec un Smartphone, est capable de diagnostiquer un mélanome. On a pris 129 450 photos de grains de beauté, chaque photo était annotée (il y avait pour chaque photo un descripteur associé pour dire quel était le diagnostic : bénin, malin ou à surveiller). A partir de ça, on a utilisé de l'apprentissage et on a fait un système qui diagnostiquait sur des photos de nouveaux grains de beauté. Et on a confronté le diagnostic fourni par la machine à celui fourni par 21 dermatologues.



l'intérêt du mélanome, c'est qu'on est capable, quand on fait l'exérèse, de savoir quel est exactement le résultat. Et là, la machine était meilleure que les 21 dermatologues. Bien sûr, vous me direz que c'est normal : même si un dermatologue voit beaucoup de grains de beauté, il n'arrive pas jusqu'à 130 000 dans toute sa carrière, ce qui est énorme.

## 2- Mémoire et apprentissage

*«La mémoire me passionne particulièrement, parce que c'est la chose la plus mystérieuse qui puisse exister»*

La deuxième question se rapporte à la mémoire et ceci me passionne particulièrement parce que c'est la chose la plus mystérieuse qui puisse exister. J'insiste sur le fait que la mémoire, ce n'est pas seulement le stockage d'informations mais aussi, et c'est essentiel, ce qui nous permet d'imaginer. La création, est en grande partie liée à la recombinaison d'éléments de mémoire. Cela a fait l'objet de beaucoup de travaux de représentations de connaissances, qui sont extrêmement importants mais que je ne mentionnerai pas ici, parce que cela serait trop long.

Au départ, on a représenté les connaissances en utilisant la logique et puis on s'est rendu compte, à partir des années 1970, que ça ne suffisait pas et on a essayé alors de s'intéresser à la psychologie, ce qui a donné naissance aux **sciences cognitives** et aux travaux sur les mémoires sémantiques.

Ce travail s'est poursuivi dans les années 2000 avec ce qu'on a appelé les **ontologies**. Là encore je n'insisterai pas : le terme est mal choisi. Il a été introduit par des ingénieurs mais il a un sens technique bien éloigné du sens philosophique initial du terme ontologie.

Ces ontologies jouent un rôle important dans ce qu'on appelle le web sémantique, c'est-à-dire pour l'organisation des connaissances sur le web ou dans les systèmes d'information afin d'essayer de construire des agents artificiels capables de faire des déductions analogues à celles que nous ferions pour accéder à des informations. C'est quelque chose de tout-à-fait central dans le monde d'aujourd'hui.

J'ai parlé bien sûr du web comme étant un modèle de mémoire, mais ce qui est beaucoup plus intéressant, plus exactement beaucoup plus actuel, ce sont toutes les techniques qui permettent d'extraire des connaissances à partir de grandes masses d'informations. Ces techniques sont extrêmement nombreuses :

- Elles sont dites supervisées (on reviendra sur les algorithmes d'apprentissage supervisé, parce que c'est ce qui a un grand succès aujourd'hui) lorsqu'on qu'on annote des exemples avec des étiquettes (c'est exactement ce qui s'est produit dans l'exemple précédent avec l'étiquetage des photos de grain de beauté par le diagnostic). Ça veut dire qu'il y a un professeur. Mais dans la nature il n'y a pas de professeur, ce qui limite considérablement la portée de ce genre de techniques.
- Il y a d'autres techniques qui sont dites non supervisées. Par exemple, on peut essayer d'apprendre sur des séquences, ou sur des graphes et essayer de trouver des motifs, voir par exemple qu'on a une séquence répétitive. Il y a quelques années, j'avais travaillé avec un de mes étudiants sur la reconnaissance de motifs dans les œuvres de Charlie Parker. C'est passionnant ce que l'on peut faire avec l'IA. De même aujourd'hui, je travaille sur l'extraction de patrons syntaxiques récurrents, qui seraient peut être caractéristiques soit d'un auteur, soit d'un style ou d'un genre littéraire.

### **Rappel sur la «cybernétique»**

Tout ça bien sûr, c'est de l'apprentissage, mais je voudrais revenir à l'apprentissage supervisé et faire une petite pause pour évoquer l'histoire et la réalité de ce qu'on appelle l'**apprentissage profond**. Ce sont des sujets extrêmement actuels, dont on parle beaucoup.

Cela a commencé par la parution de 2 articles en 1943, avant l'IA, avant même les ordinateurs électroniques, dans une discipline qui s'appelait à l'époque l'automatique théorique, devenue quelques années plus tard la cybernétique.

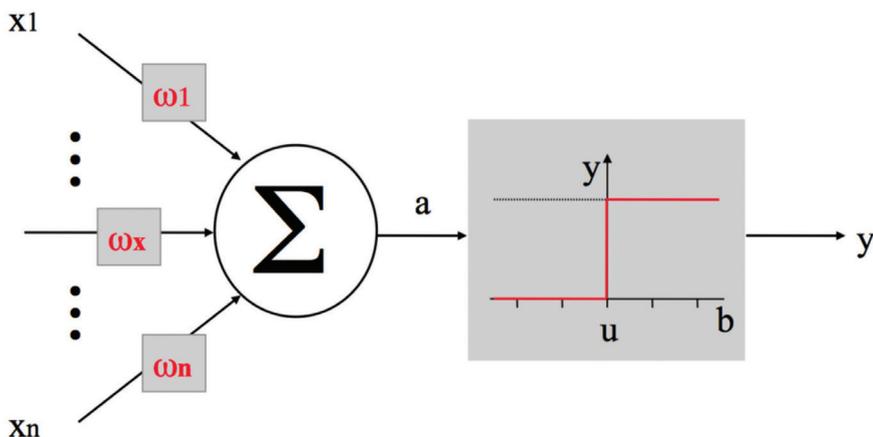
- Le premier article, de *Arturo Rosenblueth, Norbert Wiener et Julian Bigelow*, relatif à ce qu'on appelait les «**machines téléologiques**», expliquait que l'on pouvait simuler un certain nombre de phénomènes naturels avec des flux d'informations. C'était nouveau et bien sûr ça pouvait faire le pont entre des phénomènes d'ordres

différents : sociaux, biologiques, psychologiques,... et cela a donné naissance à un nombre impressionnant de réflexions théoriques de tout premier pla ;

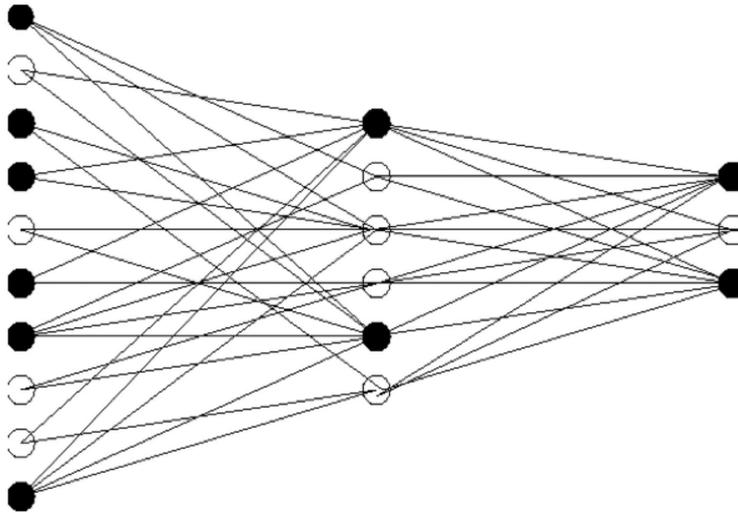
- Le second article - écrit par *Waren McCulloch* (un neurobiologiste) et un jeune homme de 20 ans à l'époque, *Walter Pitts*-, est une modélisation, un essai de simulation de ce que l'on connaissait des cellules du cerveau. A l'époque, il n'y avait pas encore d'ordinateurs et on utilisait des relais téléphoniques qu'on mettait bout à bout pour faire du calcul. Les auteurs vont donc représenter les cellules du cerveau avec des relais téléphoniques, en supposant des connexions entre les cellules que l'on sait plastiques (*déjà à l'époque, on avait l'idée que l'apprentissage était lié à l'évolution des propriétés de conduction des connexions entre les cellules qu'on appelle synapses*) et en les affectant d'un poids pour pouvoir représenter l'influx nerveux incident comme étant la sommation de tous les influx des cellules connectées pondéré par les poids des synapses incidentes. Cela recourt ensuite à un automate très simple représenté avec une **fonction en escalier** : si l'influx nerveux incident est supérieur à un certain seuil, la cellule est activée, sinon, elle ne l'est pas. Cela se résume sur la formule suivante :

$$y(t+1) = f\left(\sum_{k=1}^n w_k x_k(t) - u\right) = f\left(\sum_{k=0}^n w_k x_k(t)\right)$$

Ainsi, notre fonction de départ pour un neurone se représente sur le schéma présenté ci-dessous :



*Walter Pitts* et *Waren McCulloch* ont ensuite démontré que si on organisait les neurones en 3 couches (c'est ça qui est important ! Une couche d'entrée, une de sortie et une intermédiaire) avec une fonction d'activation en escalier, on peut réaliser, à condition de bien pondérer les poids et d'avoir suffisamment de neurones dans la couche centrale, n'importe quelle **fonction booléenne**.

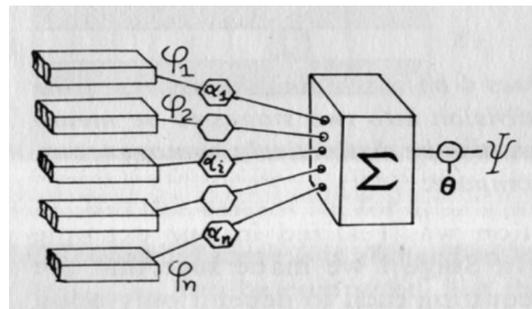


C'est un théorème d'universalité et c'est remarquable.

Mais que se passe-t-il si on a une fonction booléenne particulière? Comment la programmer? On sait qu'il faut ajuster les poids mais comment faire à la main? Cela est très difficile et on n'y arrive pas.

Des chercheurs vont essayer de les ajuster automatiquement et pour ça, ils proposent de donner des exemples étiquetés. Autrement dit, ils proposent de faire de l'apprentissage supervisé.

**C'est ce qu'a essayé de faire dès 1950 Marvin Minsky, l'un des pionniers de l'IA. En vain.**



En 1958, quelques années après que l'IA soit née, un ami de *Marvin Minsky* qui s'appelait *Frank Rosenblatt* a inventé un système d'apprentissage pour les réseaux de neurones qui était très intéressant : il examinait la différence entre la sortie obtenue et la sortie désirée et, à partir de ça, il la répercutait rétroactivement, en transformant les différents poids synaptiques.

Cela fonctionnait assez bien, sauf que le théorème d'universalité de *Pitts et McCulloch* n'est validé que sur des réseaux à 3 couches tandis que le perceptron ne fonctionnait que sur des réseaux à 2 couches. C'est ce qu'a démontré quelques années plus tard *Marvin Minsky* dans l'ouvrage qu'il a co-écrit avec *Seymour Papert* qui s'intitulait «**Perceptrons**» (au pluriel).

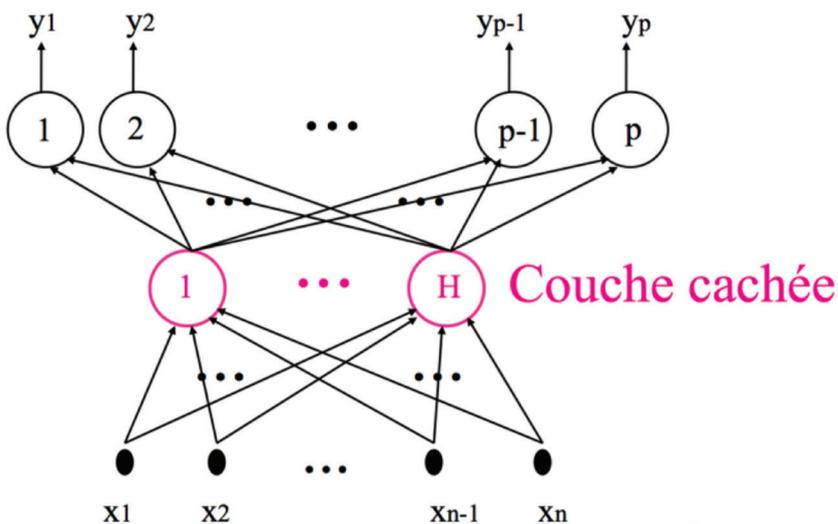
## Limitations du Perceptron- Minsky & Papert 1969

Dans cet ouvrage paru en 1969, les auteurs montrent que les algorithmes du type Perceptron sont très généraux, qu'ils marchent bien pour des réseaux à 2 couches mais que les fonctions qui sont apprises sont très limitées, mais qu'ils ne marchent pas sur des réseaux à 3 couches ou plus.

Ils montrent par exemple qu'il n'est pas possible de réaliser un «ou exclusif» sur un réseau de neurones à deux couches, car cette fonction n'est pas «linéairement séparable». Il faudrait donc un réseau à trois couches pour le réaliser. Or, on n'était pas capable d'apprendre sur des réseaux de neurones à trois couches à l'époque. Minsky décide alors d'arrêter de travailler sur les réseaux de neurones formels.

## Perceptrons multicouches - Rumelhart 1985

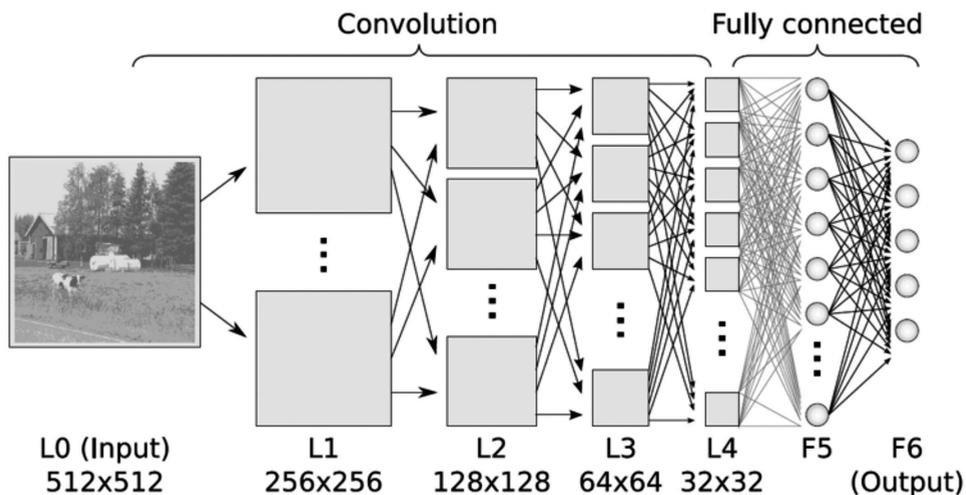
Il faut attendre l'année 1985 pour voir l'introduction d'un nouvel algorithme qui généralise l'algorithme de Perception mais qui permet d'apprendre sur 3 couches. C'est ce qu'on appelle la **rétro-propagation de gradient**. Le gradient est terme scientifique qui désigne une différence, en l'occurrence, la différence entre la sortie obtenue et la sortie désirée. Quant à la «rétro-propagation», cela signifie que l'on propage en arrière ce gradient pour modifier la pondération des poids synaptiques.



L'enthousiasme de l'époque est toutefois freiné par la lenteur des ordinateurs qui exécutent l'algorithme. Des mathématiciens vont alors chercher à comprendre ce qui est important dans l'apprentissage à rétro-propagation de gradient. Des théories formelles de l'apprentissage sont introduites, en particulier les théories du PAC Learning (*Probably Approximately Correct*) et surtout les théories de l'apprentissage statistique. Cela donne naissance à des techniques comme les SVM (*Support Vector Machine*) ou les machines à noyaux, techniques qui seront les seules à être utilisées à partir de 1995.

## 2010 – 3<sup>e</sup> renaissance...

Plus personne n'utilise donc, depuis 1995, les réseaux de neurones formels, sauf un Français expatrié aux États-Unis (*Yann Le Cun*), qui va s'acharner à continuer à les utiliser, mais non pas sur 2 couches ni sur 3 mais sur 15. Et c'est ce qu'on appelle **l'apprentissage profond**. Certaines couches sont « câblées »; c'est ce qu'on appelle les convolutions (tous ceux qui ont fait un peu de traitement du signal savent que ce sont des opérations tout à fait classiques ; on les câble directement donc il y a des poids qui n'évoluent pas) et puis après, il peut y avoir plusieurs couches d'apprentissage. Il y a des couches de filtrage aussi.



Apprentissage profond – Deep Learning

Avec ces réseaux qui sont extrêmement volumineux, qui ont beaucoup de connexions, on arrive donc à apprendre et cela tient au fait que les machines sont beaucoup plus puissantes aujourd'hui qu'elles ne l'étaient en 1985.

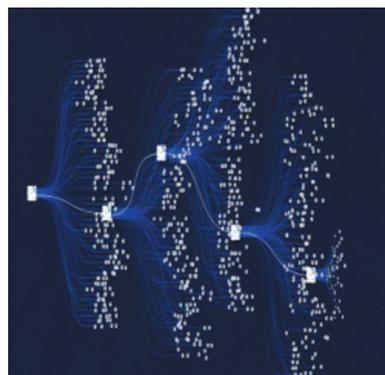
Et puis le 2<sup>ème</sup> miracle, c'est que, avec ces techniques, on apprend sur de très grandes masses d'exemples. Or, on dispose aujourd'hui d'immenses quantités d'exemples (j'ai cité les 200 millions d'images utilisées pour reconnaître des visages, mais on peut avoir encore plus d'exemples comme on le verra plus tard). Et c'est bien sûr l'origine de la renaissance de l'IA parce que l'on s'est rendu compte que ces réseaux de neurones profonds étaient capables de performances bien meilleures que les autres techniques d'apprentissage-machine.

Exemple d'application : les jeux. Je n'insisterai pas là-dessus mais il faut savoir que les jeux en général ont été vus comme des tests pour faire des progrès en IA. John McCarthy disait des jeux que c'était la drosophile (mouche servant à faire des tests en génétique) de l'IA.

- Il y a bien sûr le jeu des échecs et la fameuse victoire de la machine **Deep Blue** qui l'a remporté sur Kasparov en 1997.
- Puis il y a 2 ans, cet épisode mémorable d'**AlphaGo** qui l'a remporté sur le meilleur joueur de Go après plusieurs parties. Alpha Go apprenait avec de l'apprentissage profond des parties qui avaient été jouées par de grands joueurs.
- Et plus récemment, au dernier trimestre de 2017, les ingénieurs de **DeepMind** ont repris AlphaGo en faisant jouer la machine contre elle-même. Et là, on a eu des performances extraordinaires : au lieu de 4 mois de calculs de machines vectorielles, il a suffi de 3 jours d'apprentissage à cette nouvelle machine, appelée **AlphaGo Zero**, qui a battu AlphaGo par 100 à 0.



AlphaGo : Deep Learning



AlphaGo Zero : la machine autistique

### 3- Raisonnement – pensée

*«Depuis Aristote, la pensée, c'est un calcul sur des représentations»*

Je ne m'appesantirai pas sur le sujet mais je vais simplement vous montrer la variété des choses qui ont été faites dans ce domaine.

- Les Inférences
  - *Démonstration automatique de théorème.* C'est de la pensée bien sûr. Il y a des machines qui démontrent des théorèmes, même si c'est difficile, même si ça pose problème. On ne démontre pas tous les théorèmes bien sûr.
  - *Logiques non-monotones.* La modélisation du sens commun obligeait à introduire de nouvelles logiques. Pourquoi? Parce que si nous avons des règles générales, nous voulons ensuite les appliquer à des cas particuliers et, souvent, ce sont des exceptions aux cas généraux. L'exemple le plus connu est celui des oiseaux qui volent. Comme il existe des oiseaux qui ne volent pas (les autruches par exemple) nous aimerions avoir une règle générale qui nous dit «les oiseaux volent» et puis, éventuellement savoir que les autruches, bien que étant des

oiseaux particuliers, ne volent pas. C'est ce qu'on a essayé de faire avec ces logiques dites non-monotones. Normalement, le nombre de théorèmes est une fonction qui croit dans un système formel : plus vous avez d'axiomes, plus vous aurez de théorèmes. Or, les logiques non-monotones font que parfois un nouvel axiome conduit à enlever des théorèmes, par exemple le fait qu'un oiseau soit une autruche conduit à retirer le théorème selon lequel il vole. Ce sont des logiques d'ordre supérieur qui permettent de simuler des raisonnements non-monotones de ce type.

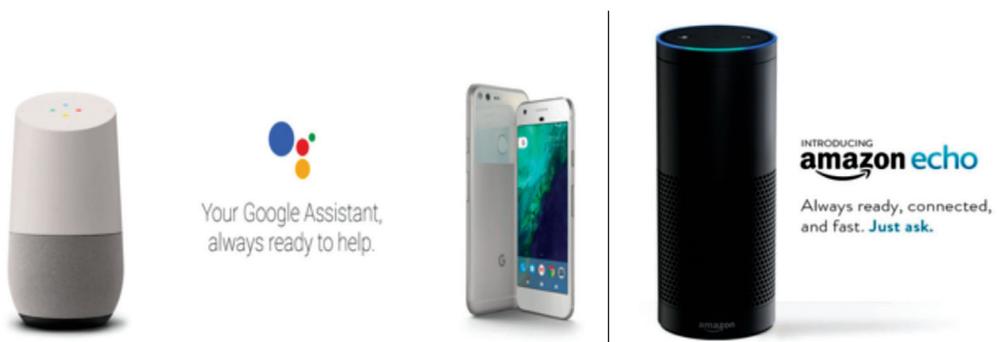
- *Raisonnement automatique.*
- *Satisfaction de contraintes.* Vous donnez juste des contraintes et la machine est capable d'en tirer partie. Si vous avez un emploi du temps par exemple, vous lui donnez des contraintes et vous trouverez une solution satisfaisante.
- Les logiques épistémiques, qui sont des logiques modales particulières. Ce sont les logiques de la connaissance (épistémiques) : elles traduisent qu'un agent sait quelque chose ou ne sait pas quelque chose, que la connaissance est partagée ou ne l'est pas, etc.
- Le retour d'expérience (traçabilité, réflexion, capitalisation des connaissances). Là encore, ce sont des choses extrêmement importantes. Comment tirer partie de tous les cas qu'on a rencontrés pour envisager des cas futurs. Par exemple, comment un médecin peut tirer partie de toute la casuistique pour améliorer le diagnostic qu'il va faire. Un autre exemple mentionné tout à l'heure est celui de la fusion des données, qui est aussi un type de raisonnement : on a des informations hétérogènes d'ordres différents qu'on peut essayer de mettre ensemble pour construire de nouvelles connaissances.

#### 4- Fonctions expressives

- Il s'agit de tout ce qui relève du langage, du langage naturel (écrit) en particulier: le traitement et la compréhension du langage naturel :
  - le traitement du langage, c'est simplement essayer d'observer la langue (un peu comme l'ont fait les linguistes il y a longtemps, qui recensaient les mots de la langue dans des dictionnaires), d'écouter les gens, et de mieux comprendre la logique du langage.
  - la compréhension du langage, c'est la traduction d'un texte écrit dans un langage, dans une représentation formelle. Et c'est utile parce qu'ensuite, on peut avoir des outils de démonstration automatiques qui vont nous permettre de faire des déductions sur ces démonstrations formelles. C'est extrêmement important si on veut faire de la commande de robots, si on veut questionner des bases de données. Cela ne veut pas dire que la machine accède au sens (c'est un peu différent). Elle traduit les textes dans un langage formel à l'aide duquel on fera des inférences (système de question-réponse).

- l'apprentissage sur d'immenses masses de textes. On peut dire que c'est une forme d'apprentissage mais en même temps, c'est lié à de la communication. On est capable de gérer une grande quantité de communications. En médecine par exemple, il y a tellement d'articles que les médecins sont incapables de tous les lire. Il y a des machines qui sont capables non seulement d'extraire automatiquement l'essentiel des informations que contiennent ces articles mais aussi qui permettent de savoir quels sont les articles qui répondent à une question donnée.
- Ensuite les Interfaces homme-machine. Ici, ce n'est plus simplement le langage naturel mais de la communication. Il s'agit :
  - du dialogue homme-machine (robots bavards notamment comme on le verra ultérieurement,...);
  - d'informatique affective : la communication n'est pas seulement le langage articulé mais peut également être le langage des émotions (le fait qu'on tremble, qu'on transpire, le mouvement des yeux,...). C'est ce genre de choses qu'on peut essayer de faire avec une machine.
- Les Interfaces cerveau-ordinateur, c'est encore plus prospectif bien sûr. Vous savez qu'on est capable, à partir d'électrodes, de déterminer les zones du cerveau qui sont activées, de corréliser ces zones avec une intention et, à partir de ça, de faire de la communication directe entre le cerveau et l'ordinateur. Ça pose bien sûr un tas de questions mais ce sont des travaux que l'on fait avec des techniques d'IA et qui sont très stimulants.
- J'ai oublié de parler de l'oral parce que je l'avais mentionné plus haut avec les questions de perception. Alors bien sûr, la traduction d'un signal de parole dans un texte écrit est aussi liée à la communication avec la machine.

L'exemple que je vais donner est celui de ce qu'on appelle les «**chatbots**» (**agents conversationnels ou robots bavards**), c'est-à-dire des petits systèmes de dialogue qui, en principe, devraient nous permettre d'échanger directement avec une machine.



Je crois qu'il faut dire 2 choses à ce propos :

- d'une part, ces techniques sont très difficiles à mettre en service et je pense leur fonctionnement limité à un environnement professionnel parce que la difficulté, c'est la sémantique. Si vous avez différents champs sémantiques qui sont abordés, s'il y a une polysémie, c'est extrêmement délicat.
- en revanche, si on est sur des significations précises et étroites, alors ça peut être mis en œuvre. Je pense donc qu'on aura du mal à les mettre à disposition d'un public large dans les maisons comme le prétendent certaines publicités. Il y a des raisons économiques à tout cela. Il ne faut pas également oublier qu'avoir un Chatbot chez soi, ça serait se payer un petit espion domestique capable d'enregistrer en continu toutes les discussions privées...

## 5- Fonctions exécutives

- Il s'agit des systèmes autonomes et/ou semi-autonomes (autonomie de mouvement et/ou d'action : voitures, armes, robots,...);
- Cela recouvre aussi toutes les questions d'aide à la décision, entre autre ce qu'on appelle les conflits d'intérêts : qu'est-ce qu'il se passe quand l'homme et la machine ne sont pas en phase l'un par rapport à l'autre ? C'est une question vraiment délicate.



*Je voudrais maintenant envisager le futur de l'intelligence artificielle, examiner les perspectives*

Je vais pour cela reprendre les conclusions d'une conférence à laquelle j'ai assisté il y a plus d'un an à New York et qui a réuni tous les grands acteurs de l'internet en particulier Facebook, NVIDIA, Google, etc.

- La première conclusion est que les enjeux du futur sont liés à la taille des données (il y aura des données de plus en plus importantes). Aujourd'hui par exemple, il y a plus de 20 000 milliards de requêtes par jour sur les moteurs de recherche et c'est ce type de choses que les acteurs de l'internet voudraient être capables de traiter. De même, il y a 2 milliards de photos partagées par jour. Là encore, ce sont ces dimensions qui intéressent ces grands acteurs.



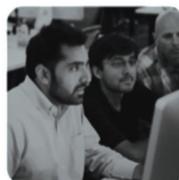
Predictions  
Per Day

>20 x 10<sup>12</sup>



Photos Shared  
Per Day

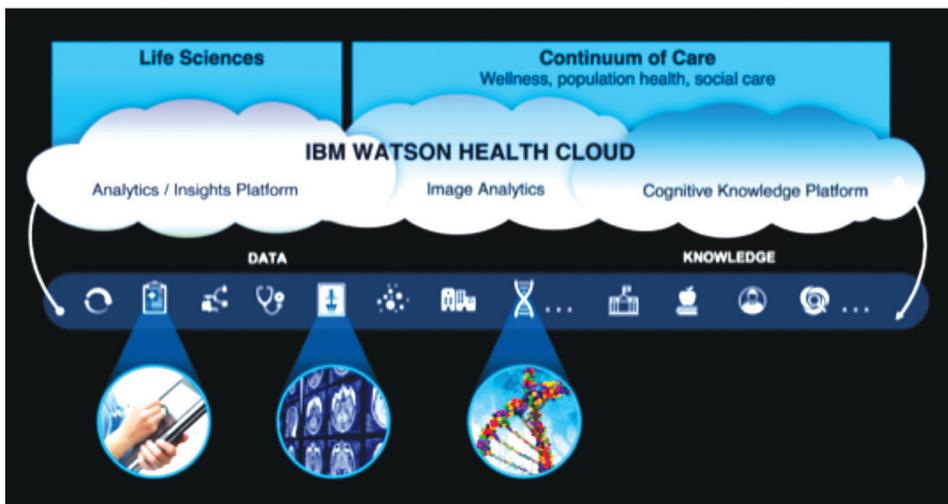
2 x 10<sup>9</sup>



Engineers using ML  
Platform

>700

- Le futur est également lié à l'efficacité des processeurs, qui permettent aujourd'hui de traiter ces immenses masses de données, et aux techniques algorithmiques de traitement, comme les techniques d'apprentissage profond.
- Les enjeux industriels, à commencer par (a) les recommandations car on vit dans des sociétés où nous sommes saturés de biens de consommations et de sollicitations. Il faut alors viser juste et proposer à chacun ce qu'il pourrait désirer. Et pour ça, il faut le profiler, ce qu'on fait avec les techniques de l'IA. Puis (b) la perception bien sûr (images, vidéos, parole, son, ...) dont on a parlé tout à l'heure. Ensuite le raisonnement (c), accès à la connaissance avec des systèmes comme WATSON, qui sont intéressants dans les domaines juridique, médical, réglementaire (parce qu'on a des connaissances de plus en plus touffues et donc il faut aider les gens à y accéder). Et puis les interfaces Homme-Machine (d) avec les émotions que j'ai mentionnées tout à l'heure, à savoir le repérage des émotions et aussi la simulation des émotions sur les machines. Enfin le dialogue (e), avec ces chatbots que j'ai déjà mentionnés.



Donc voilà pour les grands enjeux industriels de l'IA.

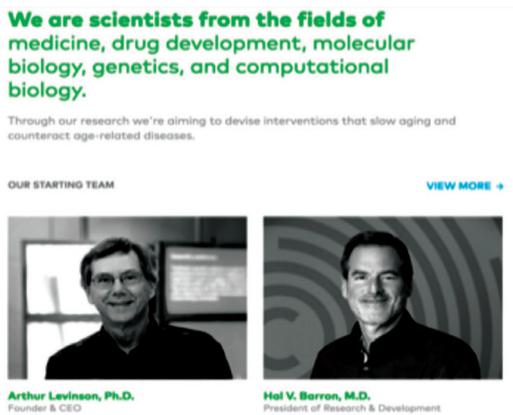
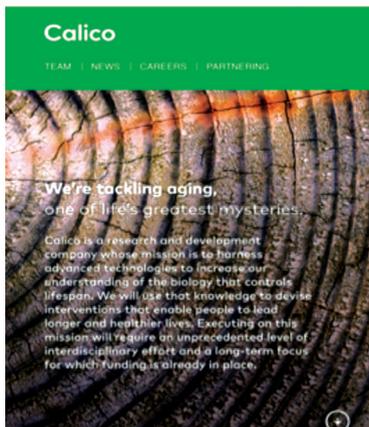
*Je voudrais maintenant conclure : jusqu'où ?*

Il y a une société qui s'appelle Calico (California Life Company) qui stocke des génomes dans le but d'essayer de détecter ce qui est à l'origine du vieillissement, de trouver le secret de la jeunesse éternelle grâce à l'IA (*we are tackling aging, one of life's greatest mysteries*). Ça va un peu loin me diriez-vous, mais sachez que la société a été montée par la femme d'un des dirigeants de Google, avec des fonds de Google.

## Limites éthiques

Il y a alors plein de questions d'éthique, des questions liées à l'intimité, à la vie privée de l'individu :

- Bases de données de patients – RGPD;
- Objets connectés et robots de compagnie (stimulateurs cardiaques, respirateurs,...);
- Excès possible de protection!



Il y a des questions de décision, de responsabilité. A un moment donné, il se peut que la machine prenne la décision à la place du médecin ou que les assurances sociales disent au médecin qu'il ne s'est pas conformé à ce que la machine demandait,... Donc là je pense qu'il faudra qu'il y ait des réflexions, qu'on exige de conserver une certaine liberté et bien sûr, et que l'on accepte d'associer à cette liberté une responsabilité, que l'on ne reportera jamais sur la machine, car ce serait une forme d'abdication.

Sur ce point, il y a de quoi s'inquiéter. Il y a quelques mois, le parlement européen a voté une loi donnant une responsabilité juridique aux robots. Je pense que c'est la pire des choses qu'on puisse faire et j'espère aussi qu'on va changer ceci.

## L'Apocalypse

Je voudrais terminer sur ces déclarations fracassantes de **Stephan Hawking**, très grand scientifique, avec une volonté farouche, décédé il y a quelques jours. Il ne pouvait plus communiquer avec le monde extérieur qu'à travers les technologies et pourtant, il manifestait son inquiétude face aux progrès de la technologie.

**«L'IA est le risque fondamental pour l'existence de la civilisation humaine», Stephan Hawking**



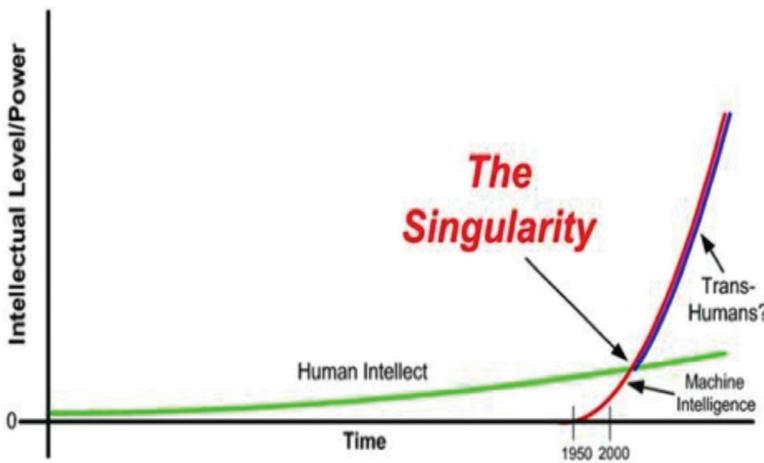
### Stephen Hawking says A.I. could be 'worst event in the history of our civilization'

- Physicist Stephen Hawking said the emergence of artificial intelligence could be the "worst event in the history of our civilization."
- He urged creators of AI to "employ best practice and effective management."
- Hawking is among a number of voices including Elon Musk who have warned about the dangers of AI.

Arjun Kharpal | @ArjunKharpal  
Published 2:11 PM ET Mon, 6 Nov 2017 | Updated 3:39 PM ET Mon, 6 Nov 2017

*«Alors que l'impact à court terme de l'IA dépend de ceux qui la maîtrisent, son impact à long terme dépend de la possibilité de la maîtriser», Stephen Hawking, 1<sup>er</sup> mai 2014*

Hawking parlait de risque existentiel et cette idée, qu'il a mentionnée dans plusieurs circonstances a été reprise par un certain nombres de gens tels **Bill Gates, Elon Musk, Frank Wilczek ou Stuart Russel**. Ils prévoient que les machines vont devenir autonomes, qu'elles vont se passer de l'Homme. Ces inquiétudes ont donné naissance à la théorie de la Singularité technologique.



La Singularité technologique

## De la Singularité au transhumanisme

Selon ses tenants, avec l'accélération des progrès des machines, il y a un moment où l'«intelligence» des machines, à savoir l'intelligence artificielle, sera supérieure à l'intelligence humaine. On passera alors de l'humanité à la transhumanité.

Certaines de ces théories affirment même qu'avec cette transhumanité, l'homme va se télécharger sur les machines, et devenir quasiment immortel. D'autres, moins optimistes, prétendent que l'humanité va disparaître et que les machines vont prendre le pouvoir. Selon les premiers, l'homme va s'hybrider, se changer en machine, ce qui veut dire que son corps disparaîtra et qu'on ne gardera que l'âme, la conscience. Le problème, c'est que ce sont des sociétés comme Google qui donnent crédit à ce genre de choses, que, selon elles, on va être des âmes sans corps et ce dès 2045. Il y a d'ailleurs un milliardaire russe qui a fait un site (2045.com) avec l'idée que vous pouvez vous réincarner, commencer à développer votre avatar immortel personnalisé et ce en cliquant sur un simple bouton virtuel.



## Une nouvelle religion!

Il y a aussi de nouvelles formes de spiritualité un peu étrange, comme celle dite de l'église de la voie du futur («Way of the future» church).

### Way of the Future: Una religión que desarrolla un dios de Inteligencia Artificial

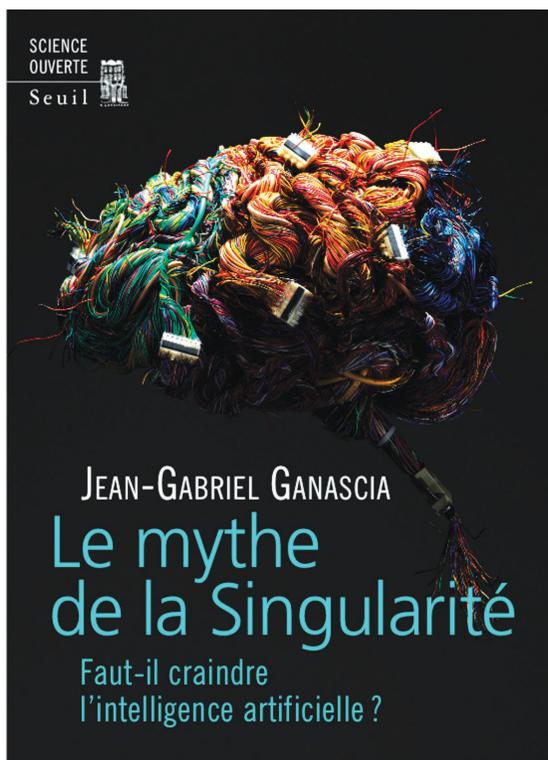
La organización religiosa fue fundada por un ex empleado de Google y Uber.



Alors, il m'a semblé ces dernières années, en tant que scientifique, qu'il était de mon devoir d'expliquer que toutes ces choses sont des fictions, que les arguments déployés par ces gens, ces autorités du monde moderne telles Hawking ou Musk, ne sont pas viables.

C'est ce que j'ai fait dans ce livre intitulé «**Le mythe de la singularité. Faut-il craindre l'intelligence artificielle?**» dont je vais résumer très rapidement les 3 points importants.

- a) Le premier, c'est l'évolution exponentielle de la technologie (Loi de Moore) et là, il y a 2 arguments qui mettent en cause ces projections sur le futur :
- d'une part, rien ne dit que cela va se poursuivre. En tous cas, la technologie du silicium aura à l'évidence des limites qui sont liées entre autres à la taille des atomes. Même s'il reste la possibilité de développer le calcul quantique, c'est une rupture technologique que l'on ne peut anticiper avec précision. Et la théorie sur l'exponentielle qu'on prétend première ne l'est pas car c'est une loi d'observation;
  - la 2<sup>ème</sup> raison, c'est que la fréquence du calcul n'est pas la conscience. On fait des machines qui font mieux les multiplications que nous, qui jouent mieux aux échecs ou au jeu de go, voire au poker, mais cela ne veut pas dire qu'elles auront une volonté propre.



- b) Le 2<sup>ème</sup> argument, c'est l'autonomie des machines : certains disent aujourd'hui que si les machines utilisent l'apprentissage, elles se développeront et construiront leurs connaissances d'elles-mêmes, des connaissances qui vont nous dépasser et qui, de ce fait, leur permettront de s'autonomiser. Or, l'apprentissage dont il est question est, comme on l'a vu, un apprentissage supervisé, qui est extrêmement limité. C'est à la fois un apprentissage très puissant, qui permet d'arriver à des résultats impressionnants, et en même temps cela demeure très limité. Une des limites tient à ce que dans l'évolution de la science, il y a des moments de rupture, des moments de changement de paradigme et ça, une machine ne sait pas le faire. Par exemple, un tournant comme celui de la relativité d'Einstein ou celui de Wegener avec la tectonique des plaques est impossible avec une machine.
- c) Alors, qu'est-ce qui donc aujourd'hui donne crédit à ces choses? Il y a longtemps, c'étaient juste quelques scientifiques ou des auteurs de science-fiction mais maintenant, ce sont des gens importants investis de hautes fonctions et qui ont beaucoup de responsabilités. Ce qui est étrange, c'est qu'ils nous disent que l'IA est dangereuse, que c'est un danger majeur pour l'humanité et ce sont eux qui la développent le plus. J'appelle ceux-là des «pompiers-pyromanes».

Mais quelles sont donc leurs motivations? Plusieurs théories existent dont la première est qu'ils veulent forger un imaginaire, nous expliquer que la technologie est toute puissante, qu'étant les maîtres de la technologie, ils ont toute puissance sur le monde du futur.

### ***Enjeux politiques: fin de la souveraineté des États***

D'autres hypothèses viennent ensuite mais la dernière est que, non seulement ils veulent asseoir leur puissance économique considérable (puissance qu'ils ont déjà), mais ils veulent en plus acquérir une puissance politique. Ils veulent se substituer aux États dans l'accomplissement d'un certain nombre de tâches, en particulier, ils veulent assumer des attributs de souveraineté comme celui de la santé (on a vu l'exemple déjà donné avec la société Calico), de l'éducation, à la place des États.

Pendant, ils visent aussi à assumer les attributs dits régaliens comme :

- la justice (Utilisation des masses de données pour prévoir les récidives (US), la législation participative avec l'existence de plusieurs sites comme <https://crowdlaw.org/>);
- la finance (Nouvelles monnaies : Paypal, Bitcoin - Percevoir l'impôt: Google par exemple propose à beaucoup de pays qui n'ont pas les moyens d'établir leur cadastre gratuitement);
- la sécurité intérieure (Reconnaissance faciale - État civil : par exemple le formulaire ESTA à remplir pour rentrer aux USA exige vos propres données et paramètres d'accès aux réseaux sociaux qui, on le sait, possèdent énormément d'informations sur chaque utilisateur). Tout ça bien sûr est assez troublant;

- la défense : les enjeux aujourd'hui ne sont plus limités à l'espace «terre-air-mer» mais concernent aussi le cyberspace. Or, sur le cyberspace justement, les grands groupes ont des ambitions de contrôle. Non seulement ils veulent avoir des armes défensives pour se prémunir contre les attaques d'un certain nombre d'acteurs mais maintenant, et cela pose des problèmes aux États, ils veulent déployer des armes offensives, pour attaquer.

**Comme conclusion**, je dirai que la singularité technologique est un peu comme l'était la religion au 19<sup>ème</sup> siècle avec Karl Marx qui affirmait qu'elle était «l'opium des peuples». On essaie de nous endormir avec une fable. Le danger est que cette fable ne nous laisse pas voir les vrais enjeux. Elle nous trompe, elle nous laisse entendre qu'il y a un risque existentiel alors qu'il y a un risque politique.

Comment faire alors pour que la peur engendrée par cette fable ne masque pas les menaces et les vrais dangers?

Merci.

**Académie Hassan II des Sciences et Techniques  
Km 4, Avenue Mohammed VI - Rabat.**

**Tél : 0537 63 53 77 • Fax : 0537 75 81 71**

**E-mail : [acascitech@academiesciences.ma](mailto:acascitech@academiesciences.ma)**

**Site internet : <http://www.academiesciences.ma>**

**Jean-Gabriel GANASCIA** (Jean-Gabriel.Ganascia@lip6.fr)



Professeur d'informatique à l'Université Pierre et Marie Curie et membre senior de l'Institut Universitaire de France, Jean-Gabriel Ganascia poursuit ses recherches au LIP6 (Laboratoire d'Informatique de Paris VI) et au sein du Labex OBVIL qui fait collaborer son équipe avec les équipes de littérature de l'université Paris-Sorbonne. Spécialiste d'intelligence artificielle (EurAI Fellow – *European Association for Artificial Intelligence* <https://www.eurai.org/>), d'apprentissage machine et de fouille de données, ses recherches actuelles portent sur le versant littéraire des humanités numériques, sur la philosophie computationnelle et sur l'éthique des technologies de l'information et de la communication.

## PARCOURS

Après une formation initiale d'ingénieur et de philosophe, il s'est orienté vers l'informatique et l'intelligence artificielle. Il est titulaire d'une thèse de doctorat sur les systèmes à base de connaissance obtenue à l'université Paris-Sud en 1983 et d'une thèse d'Etat sur l'apprentissage symbolique soutenue à l'université Paris Sud en 1987.

Il a été nommé assistant à l'université d'Orsay (Paris XI) en 1982, puis maître de conférence dans cette même université en 1987 et professeur d'informatique à l'UPMC en 1988.

Il a dirigé le Diplôme d'Etudes Approfondies IARFA (Intelligence Artificielle, Reconnaissance des Formes et Applications) pendant 12 ans (1992-2004). Il a aussi été chargé de mission à la direction du CNRS (1988-1992) avant de créer et de diriger le Programme de Recherches Coordonnées «Sciences Cognitives» pour le compte du ministère de la recherche (1993) puis le Groupement d'Intérêt Scientifique «Sciences de la cognition» (ministère de la recherche, CNRS, CEA, INRIA, INRETS) (1995-2000). Il a coordonné, pour l'université Pierre et Marie Curie, le master Erasmus Mundus DMKM (Data Mining and Knowledge Management – Fouille de données et gestion de connaissances) entre 2010 et 2016.

Il est président du COMETS (comité d'éthique du CNRS) et membre de la CERNA (commission de réflexion sur l'éthique de la recherche dans les sciences du numérique d'Allistene).

## PUBLICATIONS

Au cours de sa carrière, il a publié plus de 350 articles dans les actes de conférences scientifiques, dans des livres scientifiques et dans des revues. Il est aussi l'auteur de plusieurs ouvrages destinés au grand public donc voici quelques items :

- Ganascia J.-G., *Le mythe de la Singularité : faut-il craindre l'intelligence artificielle?*, Éditions du Seuil, Collection Sciences Ouvertes, 2017.
- Ganascia J.-G., *Intelligence Artificielle : vers une domination programmée?*, Le Cavalier Bleu, Collection Idées reçues, 2017.
- Chapouthier G., Ganascia J.-G., Naccache L., Picq P., *Que reste-t-il du propre de l'homme?*, Les Presses de l'ENSTA, 2012.
- Ganascia J.-G., *Voir et pouvoir : qui nous surveille?* Editions le pommier, 2009.
- Ganascia J.-G., *Les sciences cognitives*, Editions le pommier, 2006.
- Ganascia J.-G., *Gédéon ou les expériences extravagantes d'un expérimentateur en chambre*, Editions le pommier, collection Roman & Plus, 2002.
- Ganascia J.-G., 2001, *L'odyssée de l'esprit* – Flammarion (Collection essais) 1999.
- Ganascia J.-G. - *Le petit Trésor de l'informatique et des sciences de l'information* – Flammarion 1998.
- Serres M., Farouky N., *Trésor des sciences* – Flammarion 1997.
- Ganascia J.-G.- *L'intelligence artificielle.*– Flammarion (Collection Dominos), 1993.
- Ganascia J.-G.- *L'Ame Machine.*– Editions du Seuil, Collection Sciences Ouvertes, 1990.