Ecole éthique du numérique
26-20 Sept 2016

L'apprentissage artificiel
« Machine learning » et l'éthique

L. Devillers

devil@limsi.fr

LIMSI-CNRS/Paris-Sorbonne

Membre de la CERNA

Préambule

Mes motivations

Détection des émotions/intentions (NN, SVM, deep learning)
Interaction affective et sociale avec des robots compagnons application pour les personnes âgées
-> *Tensions éthiques soulevées par le traitement des données personnelles, les relations affectives, l'intimité, la coévolution humain-machine, l'autonomie et l'apprentissage*
Que peut-on savoir de ce qu'a appris la machine et de ce qu'elle peut apprendre en interaction avec les humains ? Comment l'éduquer (ethics by design) ?

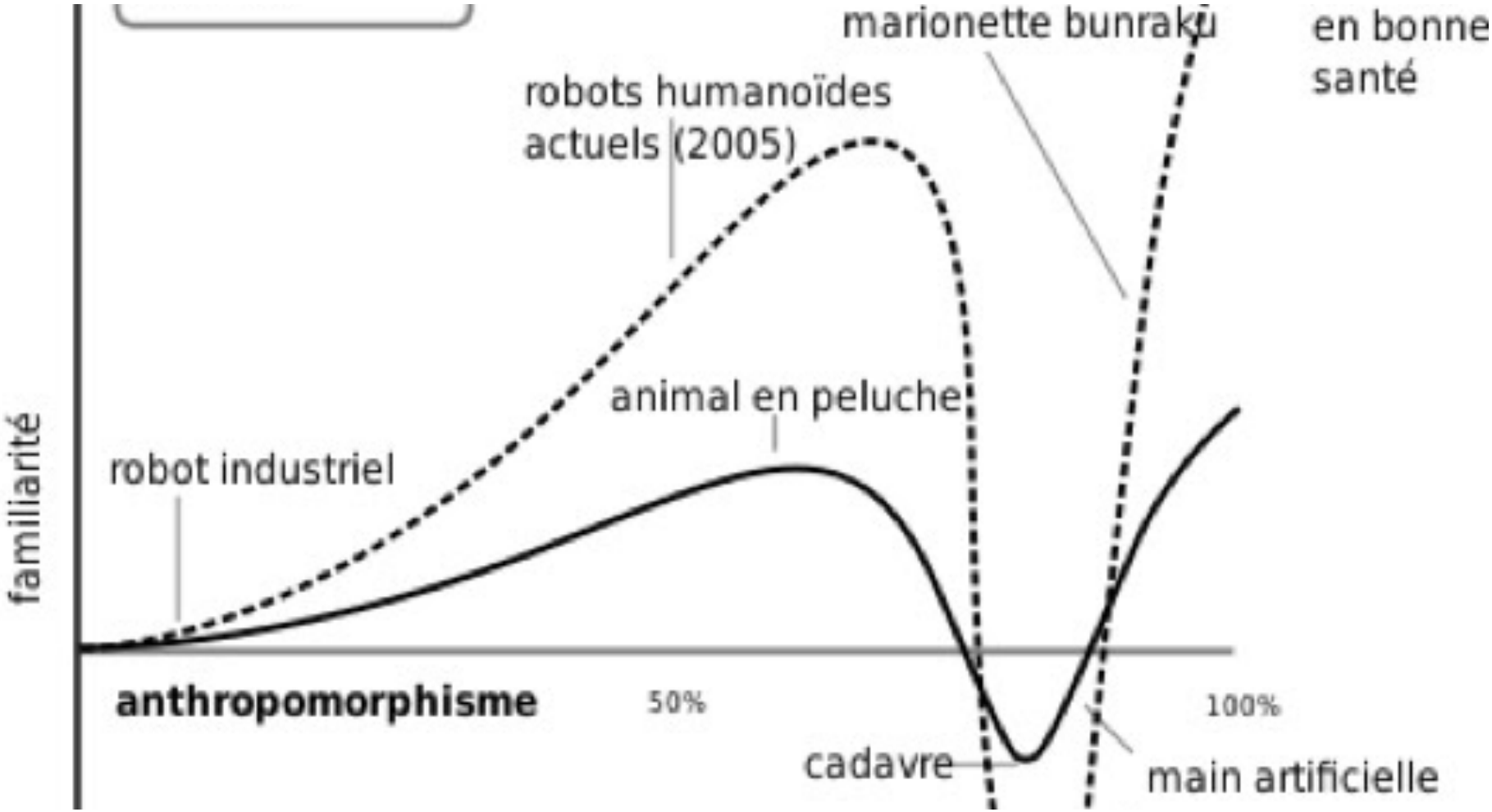
Les robots ne ressentent rien, n'ont pas de conscience malgré le fait qu'ils pourront dire « *je t'aime* »!

- Les machines seront de plus en plus autonomes grâce à des programmes d'intelligence artificielle sophistiqués, elles ne seront pas pour autant capables de sentiments, de créativité et d'imagination tels qu'on l'entend pour les humains.
- La machine ne peut pas se sentir « heureuse » car elle est dénuée de conscience phénoménale. Elle ne peut pas « comprendre » d'ailleurs le concept de bonheur pour un humain.
- Ces concepts sont également difficile à formaliser.
- Lorsqu'AlphaGo bat le joueur de go, la machine ne comprend pas ce qu'elle fait et n’est pas heureuse

Relations affectives avec les robots

- Créer une relation affective avec les robots n'est plus un sujet de science fiction. La ressemblance avec un être humain ou un animal, les mimiques faciales, le ton de voix, ou encore l'aspect enfantin ou peluche de certains robots contribue à susciter l'émotion.
- Mais la ressemblance au vivant ne pourra aller que jusqu'à un certain point si on en croit la théorie de la vallée de l'étrange **(Exposition Persona : étrangement humain, Quai Branly)**.
- Car si l'on va trop loin, le moindre défaut devient monstrueux et la machine est rejetée. Cependant, il est possible d'éprouver des émotions en face de n'importe quel objet.

La vallée de l'étrange (Masahiro Mori)



Projection sur l'objet : attribution d'intentions

- Cependant, il est possible d'éprouver des émotions en face de n'importe quel objet.

- L'humain projette des relations affectives avec des robots non humanoïdes, dépourvus de capacités affectives comme des robots téléguidés démineurs qui sauvent des soldats ou encore des robots aspirateur.

- L'attachement est un lien affectif qui résulte de l'histoire commune. Certains leur donnent des noms, preuve que l'humain projette une identité sur le robot.

Anthropomorphisation

- L'anthropomorphisme est l'attribution des caractéristiques comportementales ou morphologiques de vie humaine à des objets.

- Nous travaillons non seulement à "l'anthropomorphisation" de la machine, mais aussi dans le sens contraire où la machine va essayer de se faire un "visage de l'humain" via la détection des états émotionnels et les états mentaux (Théorie de l'Esprit).

La « *media equation* » de Reeves et Nass en 1996

- Nous appliquons les mêmes attentes sociales lorsque nous communiquons avec des entités artificielles.
- Ainsi un objet qui semble être dans la douleur, comme le robot Atlas de Boston Dynamics, peut inspirer de l'empathie.
- Video Boston Dynamics: ATLAS
- Grâce à l'imagerie cérébrale, les chercheurs ont constaté que les humains ressentaient de l'empathie envers des robots maltraités, certes de moindre intensité qu'envers des humains maltraités mais cette empathie n'existe pas envers des objets inanimés.

Réflexion éthique sur les robots sociaux

- Les robots dotés d'empathie qui détectent les émotions humaines amènent à réfléchir à des règles éthiques sur l'attachement et l'intimité avec les machines, tout particulièrement pour les personnes vulnérables (personnes âgées, enfants).
- Une co-évolution humain-machine sera possible, la machine s'adaptera à l'humain et l'humain à la machine. Comment surveiller l'interaction avec de tels systèmes : les connaissances apprises, le respect de l'intimité de la personne ?
- Faut-il prévoir une réglementation, des recommandations éthiques…

La Société de l'Esprit (1986)

- Marvin Minsky, mathématicien de formation, « *voyait le cerveau comme une machine dont le fonctionnement pouvait être étudié et répliqué. (...) ».*

- Minsky crée sa théorie de la "société de l'esprit" en 1986 selon laquelle l'esprit serait composé d'un ensemble d'agents autonomes spécialisés, organisés en hiérarchie, dépourvus d'intelligence qui sont interconnectés, tantôt ils coopèrent, tantôt ils s'affrontent. La pensée, les sentiments sont le résultat de ces interconnections. Notre sensation d'avoir un esprit unifié ne serait donc qu'une illusion ?
- Pour certains chercheurs, son approche est une mine d'idées, pour d'autres cette théorique est trop éloignée de la réalité. Il propose une modélisation abstraite de l'esprit humain : l'esprit est ce que le cerveau produit, c'est à dire le résultat de l'activité électrique de notre système nerveux.

Robots : systèmes complexes

Perception – Prise de décision - Action

- Les robots sont des entités complexes qui utilisent plusieurs modules qui font
- Perception/Interprétation : Reconnaissance de la parole, faciale, des émotions, de l'interprétation sémantique, de l'interprétation des intentions,
- Raisonnement/prise de décision: Gestion de l'interaction, de mémoires (profil de l'utilisateur, historique du dialogue),
- Action: génération, synthèse, actions dans le monde physisue…
- De nombreux modules sont entrainés et optimisés en dehors du système complexe à l'aide de grand corpus de données. On peut pour cela utiliser du deep learning ou d'autres algos de ML
- Le robot ensuite utilise tous les modules interconnectés avec une gestion dans des évènements asynchrones

CERNA : Réflexion éthique sur les limites et performances des modèles (ML)

Système utilisant des modèles obtenus par apprentissage machine

- rupture technologique et juridique par rapport aux algorithmes classiques paramétrables : Modèle = Algorithme + Données

- Certains de ces algorithmes d'apprentissage machine apprennent également au cours de leur utilisation de façon autonome à partir des données des utilisateurs, de leur environnement ou d'autres programmes : Evaluation ?

- Les modèles obtenus sont dit de type « boite noire » : pour l'utilisateur, il est difficile de savoir ce que le système connaît. Le système n'est pas transparent

Les systèmes vont apprendre tous seuls sans contrôle

Ils peuvent construire des connaissances sémantiques et prendre des décisions de façon autonome. Pourquoi pas avoir des intentions…

C'est le cas des robots sociaux – assistant compagnon pour les personnes âgées…

Quelles règles éthiques pourrait-on préconiser ? Quelles règles juridiques pourraient en découler ?

Initiative mondiale IEEE pour les considérations éthiques dans la conception de systèmes autonomes

- Président : R. Chatila – Directeur exécutif : J. Havens

Robotique, IA, Affective computing, …
Juristes, Philosophes, Informaticiens, Cogniticiens, Psychologues, …

- Propositions de projets de normalisation, les certifications et les codes de conduite
- feuilles de route complets, des livres blancs et des matériels éducatifs qui peuvent inclure des MOOCs
- matériaux et / ou des programmes de certification
- Un environnement efficace et économique pour la construction d'un consensus et produire des résultats partagés
- Langue des codes de conduite ou des principes en matière d'éthique et de systèmes autonomes pour itérer dans les groupes de collaboration ouverts
- Propositions de conférences et sommets
- Coopération avec les alliances industrielles ciblées dans ce domaine

Ethics Initiative Affective Computing Committee Call (IEEE)

- Systems supporting human potential: Human flourishing, dignity, human autonomy, needs satisfaction. Roles of Positive and negative affect - gift of catharsis - support for human autonomy
- When systems lie. Emotions, honest signals, trust and deception
- When systems become intimate: Intimacy and relations with machines
- When systems go across cultures: Respect for cultural nuances of signaling. Should they represent their own culture?
- When systems have their own "feelings": Robots and robot suffering. SHould they have it?
- System Nudging - Social justice, charitable acts. When systems manipulate emotions to sell us stuff

IEEE Global Initiative for Ethical Considerations in the Design of Autonomous Systems

4 leviers

- Eduquer à l'éthique des systèmes
- Expliciter des règles éthiques
- Mettre en œuvre des outils informatiques de vérification
- Avoir un ensemble de règles juridiques en cas de non respect de ces lois.

Media

Déclarations médiatiques : alertes

- Stephen Hawking (2014): « Success in creating AI would be the biggest event in human history » « In the near term, world militaries are considering autonomous-weapon systems that can choose and eliminate targets », « humans, limited by slow biological evolution, could'nt compete and would be superseded by A.I. »
- Si Stephen Hawking a pu lancer son appel à la vigilance, c'est en partie grâce à un ordinateur très avancé qui permet à ce physicien atteint de la maladie de Charcot de s'exprimer.

- Echelle de l'histoire de l'Univers
- Nuance son avis en discutant avec le PDG de DeepMind

- Lettre ouverte signée par 700 personnalités (10 janvier 2015) sur les dangers de l'intelligence artificielle : RESEARCH PRIORITIES FOR ROBUST AND BENEFICIAL ARTIFICIAL INTELLIGENCE

Déclarations médiatiques : alertes

- Elon Musk (15 décembre 2015), fondateur de PayPal, Tesla et Space X, lance l'organisation OpenAI qui se veut un centre de recherche sur l'intelligence artificielle (IA). Pour faire face aux dangers de l'IA, autant démocratiser autant que possible ladite technologie. « *Si tout le monde dispose des pouvoirs de l'IA, alors il n'y aura pas une seule personne, ou un petit groupe d'individus qui disposeront des superpouvoirs de l'IA* », explique-t-il dans une interview.
- Elon Musk (3 juin 2016) explique aussi l'urgence des implants intra cérébraux chez l'homme pour éviter d'être vassalisés par l'IA !

- Bill Gates (2015) estime qu'il faut aborder les développements de l'intelligence artificielle (IA) avec la plus grande prudence. « *Je suis dans le camp de ceux qui s'inquiètent du développement d'une super intelligence* », a expliqué le co-fondateur de Microsoft, *« D'abord, les machines réaliseront pour nous de nombreuses tâches sans être très intelligentes. Cela devrait s'avérer positif si nous les gérons bien. Mais, quelques décennies plus tard, leur intelligence sera suffisamment développée pour devenir un sujet d'inquiétude. Je rejoins Elon Musk et quelques autres et ne comprend pas pourquoi certaines personnes ne semblent pas s'en inquiéter* », a écrit le milliardaire en réponse à une question sur le sujet, tout en participant à des projets Microsoft qui s'en approchent.

- Besoin de régulation nationale et internationale et de gestion du risque «industriel » et «sociétal »

Déclaration de google : éthique

- « Google réfléchit au moyen de désactiver un programme sans que celui-ci s'y oppose », Le monde Juin 2016 (M. Tual)
  - Scientists from Google's artificial intelligence division, DeepMind, and Oxford University are developing a "kill switch" for AI.
  - « Safely Interruptible Agents » pub. on the website of the Machine Intelligence Research Institute (MIRI) «*Reinforcement learning agents interacting with a complex environment like the real world are unlikely to behave optimally all the time… We proposed a framework to allow a human operator to repeatedly safely interrupt a reinforcement learning agent while making sure the agent will not learn to prevent or induce these interruptions.* »
    - Laurent Orseau, Google Deep Mind, antérieurement à AgrotechParis (INRA)
    - Stuart Armstrong, chercheur au Futur of Humanity Institut, Oxford, dirigé par Nick Bostom (auteur de Superintelligence: Paths, Dangers, Strategie, 2014)
  - Pour L. Orseau, l'intention est avant tout de connecter les communautés de chercheurs qui travaillent sur l'apprentissage des machines (deep learning) et qui se penchent sur les questions éthiques. L'idée est de commencer à réfléchir à ces questions de façon plus technique… On en est tout au début!

L'IA et le transhumanisme

Le transhumanisme prône l'usage des sciences et techniques pour améliorer l'humain. La convergence annoncée entre Nanotechnologies, Biotechnologies, sciences de l'Information et sciences Cognitives dite «Convergence NBIC » met l'accent sur les possibilités technologiques d'augmentation des capacités humaines.

Le transhumanisme annonce/promeut l'apparition d'une espère nouvelle post-humaine de cyborgs (humain qui a reçu des greffes de parties mécaniques). L'université de la Singularité (Kurzweil et Grossman, 2006) a reçu le soutien de Google et de la Nasa

Ray Kurzweil says: « computer will match or even beat human intelligence when they possess emotion and a sense of humor » (mars 2016) (horizon 2029 - @CNNTech)

Calico (2013) – Google – s'attaque au défi de l'âge et des maladies associées – big data

L'IA et le transhumanisme

Laurent Alexandre, Nouvel OBS – Médecine de demain, l'angle mort de la santé – 29 Juin 2016 (critique de loi Touraine) – urgent de définir une stratégie européenne de plateforme

- **Le médecin perd le pouvoir :** le risque est grand que le médecin de 2030 soit une infirmière subordonnée à un algorithme, GAFA disposent de plateformes, d'un énorme volume de données les données, IA auto-apprenante et domineront la chaine des valeurs
- **l'eugénisme devient acceptable (augmentation) :** via les thérapies cellulaires et la chirurgie du génome – USA politique de santé plus « transhumaniste » - Chine – aucun problème éthique – EU –bio-conservateur

Le Monde – 28/09/16 – M. Tual

Intelligence artificielle : les géants du WEB lancent un partenariat sur l'éthique (sauf Apple)

Ce partenariat prendra la forme d'une organisation à but non lucratif, qui « mènera des recherches, recommandera de bonnes pratiques, et publiera les résultats de ses recherches sous une licence ouverte ». Ses domaines d'activité concerneront notamment l'éthique, et plus largement l'impact de ces technologies sur la société. Son bureau devrait être constitué pour moitié de représentants de ces entreprises, l'autre moitié étant composée de chercheurs, d'associations et autres membres de la société civile.

Toutefois, promet l'organisation, elle « ne compte pas effectuer de lobbying auprès des gouvernements ». Même si, a-t-on appris lors d'une conférence de presse rassemblant toutes ces entreprises, « le but sera d'éduquer le public, mais aussi les gouvernements ».

Le Monde – 28/09/16

La note d'intention reste encore très vague, et n'engage en rien les entreprises participantes à appliquer les « bonnes pratiques » que définira cette organisation.

Quelques principes forts sont néanmoins mis en avant : ce partenariat vise à « protéger la vie privée et la sécurité des individus », et « s'oppose au développement et à l'usage de technologies d'IA qui violeraient les conventions internationales ou les droits humains ».

Même si, concrètement, il sera difficile de définir ce que sont des « technologies qui violeraient les droits humains », c'est davantage l'usage que la technologie elle-même qui serait en mesure de le faire.

Qui plus est, si l'organisation dit « promouvoir des technologies qui ne font pas de mal », elle ne dit à aucun moment interdire celles qui en font.

What to think about Machines that Think
J. Brockman (2015)

To Marvin Minsky

1980- now : discussions philosophiques sur le fait qu'une machine puisse être consciente, sur quels sujets, quels droits… Le risque de l'IA forte

Plus de 150 avis d'experts :

- *Stanislas Dehaene (neuroscientifique, Collège de France) : Two Cognitive Functions Still Lack*
  Author: Consciousness and the Brain (2012)

  Global workspace and Theory of Mind
- *Nick Bostrom (Oxford, Future of Humanity) : It's still early days*
  Author : SuperIntelligence : Paths, Dangers, Strategies (2012)

  Machine are currently very bad at thinking except in certain narrow domains. Machines are already much stronger and faster than any biological creature. There's little information about how far we are from human-level AI

What to think about Machines that Think
J. Brockman (2015)

- *Tim O'Reilly (CEO O'Reilly Media) : What if We're the Microbiome of the Silicon AI ?*

  The AI we will confront won't be a mind in an individual machine !
- *Stuart Russel (UC Berkeley) : Will They Make Us Better People ?*

  Author: AI : A modern Approach

  Build intelligence that's provably aligned with human values
- *Alex (Sandy) Pentland (créateur du MIT media lab, physique sociale, big data, vie priéve) : The Global Artificial Intelligence is Here*

  Democratic systems that include both humans and computer intelligences
- *Max Tegmark (Physique, MIT, Future of Life Institute) : Let's get prepared !*

  The most interesting question about AI: what we do about it

  What happens to humans if machines gradually replace us on the job market ?

What to think about Machines that Think
J. Brockman (2015)

- *Demis Hassabis et al. (Google deep mind) , A short distance Ahead – and Plenty to be done*

Safety and ethics

AI Winter – AI Spring

Some people are too optimistic -> prediction

Human-level AI is not just around the corner

Peter Norving :

The narrative has changed. It has switched from :

Isn'it terrible that AI is a failure ?

Isn'it terrible that AI is a success ?

Apprentissage – domaine de l'IA

Définition de l'IA (livre blanc CIGREF)

La norme ISO 2382-28 définit l'intelligence artificielle comme la « capacité d'une unité fonctionnelle à exécuter des fonctions généralement associées à l'intelligence humaine, telles que le raisonnement et l'apprentissage ».

Larousse : « ensemble de théories et de techniques mises en oeuvre en vue de réaliser des machines capables de simuler l'intelligence humaine ».

L'Intelligence artificielle est une tentative de réplication des capacités cognitives humaines en vue d'atteindre des objectifs de manière autonome en tenant compte des contraintes et des obstacles éventuels.

Intelligence Artificielle

« La recherche de moyens susceptibles de doter les systèmes informatiques de capacités intellectuelles comparables à celles des êtres humains » Lexicographie du Centre National de Ressources Textuelles et Lexicales (CNRTL) créé par le CNRS.

Définition plus technique : L'intelligence artificielle perçoit l'environnement, l'analyse, et prend une décision en conséquence mais surtout apprend.

IA Concept complexe et dynamique

Intelligence Artificielle

- l'intelligence artificielle « faible » : elle vise simplement à imiter certaines fonctions de l'intelligence pour répondre à des missions spécifiques. Il s'agit des programmes les plus utilisés. La machine ne fait que donner une impression d'intelligence ;

- l'intelligence artificielle « forte » vise, quant à elle, à faire naître une sorte de conscience. Elle se rapproche du raisonnement humain en modélisant notamment les interactions neuronales et en développant la notion d'apprentissage.

Dès son origine, l'IA cherche à imiter les fonctions de l'intelligence humaine

- 1950 Alan M. Turing, mathématicien britannique et théoricien précurseur de l'informatique, lance le concept d'intelligence artificielle et le test de Turing : le jeu de l'imitation
- 1957 Modélisation des jeux d'échec
- 1958 Construction du Perceptron, Frank Rosenblatt, machine dite connectionniste
- 1960 M. Minsky, J. McCarty, MIT :machine capable de penser comme un humain
- 1986 M. Minsky, Société de l'Esprit
- Algorithme de rétropropagation du gradient – Rumelhart, Hinton et Williams
- 1989 Deep Thought, supercalculateur d'IBM, deux millions de coups par seconde
- 1990-97 Développement de Deep Blue : conception d'un système de 256 processeurs fonctionnant en parallèle ; chaque processeur peut calculer environ trois millions de coups par seconde – bat Gary Kasparov aux échecs en 1997
- 2011 Watson, le superordinateur d'IBM, remporte deux des trois manches du jeu télévisé Jeopardy! IA répond à des questions de culture générale
- 2013 Human Brain Project : simuler un cerveau humain sur machine – 1,19 milliard eu
- Google ouvre un laboratoire de recherches dans les locaux de la NASA

Dès son origine, l'IA cherche à imiter les fonctions de l'intelligence humaine

- 2014 Deep Knowlegdge Ventures nomme à son conseil d'administration VITAL, un algorithme capable d'élaborer ses décisions en analysant les bilans comptables des entreprises potentiellement intéressantes, les test cliniques, la propriété intellectuelle et les précédents investissements
- Eugene Goostman, programme informatique conçu en Russie, est parvenu, lors d'une compétition oà tromper plusieurs personnes dans le cadre d'un test de Turing

- 2015 Facebook Artifical Intelligence Research (FAIR)
- Google rend sa technologie d'IA TensorFlow accessible à tous
- Développement d'une crainte que l'intelligence artificielle dépasse à terme les performances de l'intelligence humaine

- 2016 Amelia d'IPSoft, un agent virtuel
- AlphaGo bat trois fois consécutives le champion du monde du jeu de go, Lee Se-Dol, en cinq manches

Apprentissage Machine
domaine de l'iA

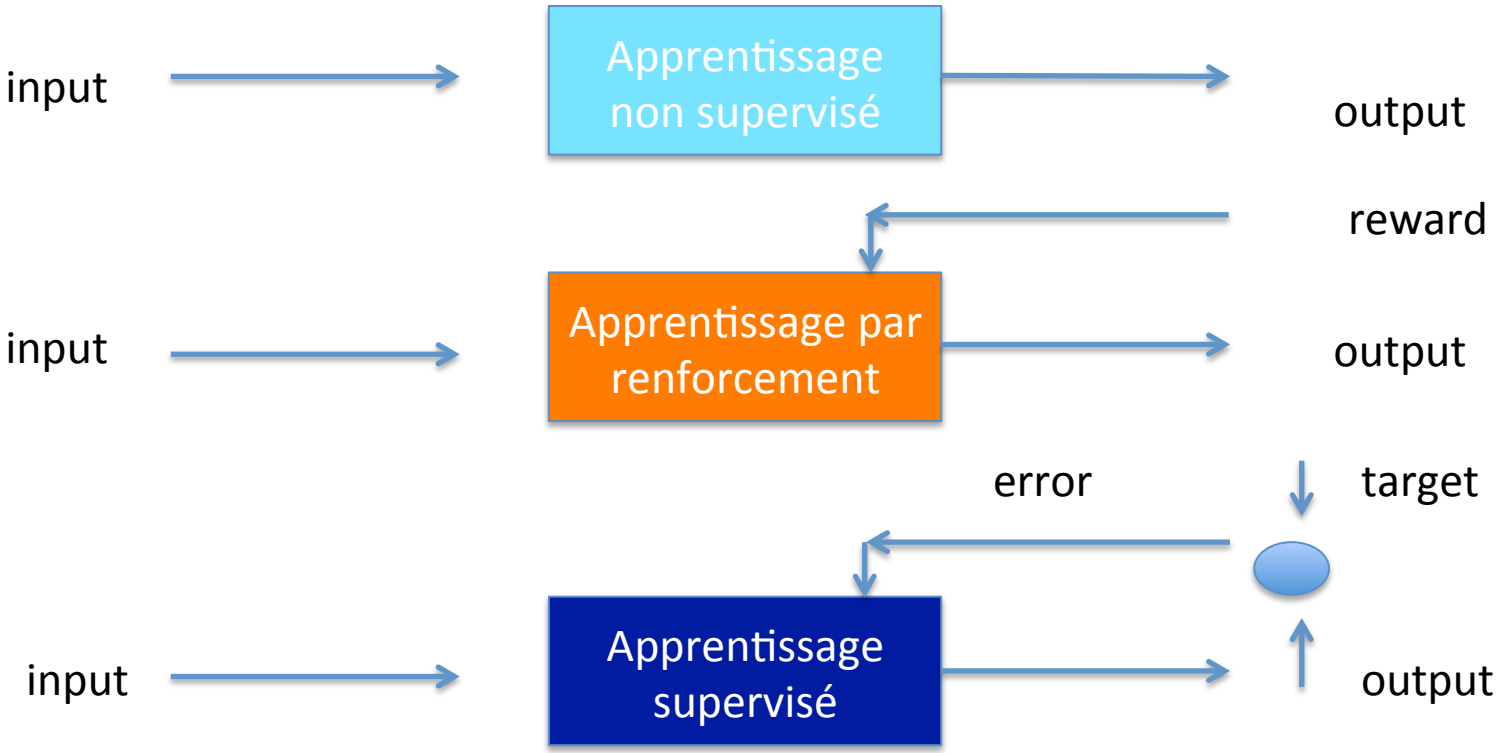
Le « Machine Learning » regroupe des algorithmes qui apprennent à partir d'exemples, de données.

Le « Machine Learning » essaie de prédire des valeurs à partir de jeux de données qui lui servent d'exemples.

La qualité du Machine Learning dépend de la qualité des données.

Apprentissage machine

- Plusieurs grandes classes d'apprentissage en IA :
  - non-supervisé, supervisé, par renforcement



Types d'apprentissage

• supervisés : le système apprend à fournir la bonne réponse à partir de données lui indiquant des exemples ;

• non supervisés par l'homme : le système apprend seul à partir des données brutes.
La majorité des applications actuelles reposent sur l'apprentissage supervisé.

Néanmoins, il existe de nombreux algorithmes non supervisés

Ex: Deep Learning de Google « Google Brain » de 2012 est un réseau de neurones distribué sur de nombreux serveurs, et auquel lui sont donnés essentiellement des tâches d'apprentissage supervisé.

Apprentissage supervisé

• L'apprentissage supervisé consiste à fournir des exemples au système en lui indiquant la bonne réponse, et il apprend à fournir la bonne réponse dans de nouveaux cas.

Il s'agit par exemple de la :

- classification : des milliers de vidéos de sport sont données au système en lui indiquant de quel sport il s'agit, et il est ensuite capable d'identifier le sport dans de nouvelles vidéos ;
- régression ou prédiction de valeur : de nombreux exemples de maisons sont donnés au système en précisant toutes les caractéristiques et le prix de vente, et il est ensuite capable de prédire le prix de vente d'une nouvelle maison, étant données ses caractéristiques ;
- prédiction de risque : de nombreuses caractéristiques de patients sont fournies au système en indiquant s'ils ont souffert ou non d'une maladie et le système prédit ensuite le risque qu'un nouveau patient développe cette maladie, en fonction de ses caractéristiqus

Apprentissage non supervisé

• L'apprentissage non-supervisé consiste à fournir de nombreux exemples au système, mais cette fois sans lui donner de « bonne réponse ».

A titre d'illustrations :

- le partitionnement ou clustering : le système apprend tout seul à détecter des groupes de clients similaires en vue d'un marketing particulier pour chacun ;
- la détection d'anomalie ou de nouveauté : le système apprend tout seul à identifier des cas qui sortent de l'ordinaire, pour la détection de fraude à la carte bancaire ou la détection de défauts de production dans l'industrie ;
- la réduction de dimensions : le système apprend tout seul à identifier les dimensions les plus importantes dans un jeu de données afin, par exemple, de les représenter visuellement d'une façon qui permette à un humain de mieux comprendre les données, ou simplement afin de les compresser ;
- les règles d'association : le système apprend tout seul à détecter des corrélations fortes. Par exemple, les clients qui achètent un produit achète souvent un autre avec, donc il faut placer ces deux produits à proximité dans un supermarché.

Apprentissage semi-supervisé

L'apprentissage semi-supervisé consiste à fournir de nombreux exemples au système, ainsi que la « bonne réponse » pour certains d'entre eux.

Par exemple,

Google Photos détecte automatiquement les personnes sur les photos et crée tout seul un dossier pour chaque personne. Il donne ensuite à l'utilisateur la possibilité de préciser qui sont ces personnes et de corriger quelques erreurs. Le système utilise ces informations pour s'améliorer.

Apprentissage par renforcement

Le système est libre d'évoluer dans son environnement qui peut être physique dans le cas d'un robot, mais peut également être un environnement virtuel.

Ce système reçoit des récompenses ou des punitions selon ce qu'il fait et il tente de s'améliorer pour avoir le plus de récompenses possibles.

C'est l'exemple :

- d'un robot qui apprend à marcher tout seul. On lui donne une récompense quand il approche de son objectif, et on lui donne une punition lorsqu'il tombe ou perd du temps. A partir de cette supervision réduite, il est capable d'apprendre tout seul à contrôler ses moteurs pour se déplacer efficacement ;
- d'un « bot » dans un jeu vidéo qui joue contre le joueur et apprend de ses erreurs pour s'améliorer.

Sur-Apprentissage

Il y a toujours un risque avec le Machine Learning : c'est le sur-apprentissage.

Certains modèles fonctionnent sur un jeu de données mais ne pourront être généralisés.

Une partie importante du travail de conception d'un système de Machine Learning est d'ajuster les paramètres de l'algorithme afin que le système apprenne au mieux des exemples fournis, mais sans sur-apprendre.

Par exemple, si je fournis à Google Photos plusieurs photos d'un ami assis sur un canapé, le système ne doit pas apprendre que le canapé fait partie de mon ami »

Deep learning

Le Deep Learning qui est considéré comme l'avancée opérationnelle majeure en termes d'intelligence artificielle repose sur la capacité d'une technologie à apprendre à partir de données brutes.

Schématiquement, un réseau de neurones est un ensemble de neurones artificiels, inspirés des neurones biologiques mais extrêmement simplifiés, qui sont connectés les uns aux autres et s'échangent des signaux très simples, par exemple un nombre de 0 à 1. Il existe de nombreuses façons d'interconnecter ces neurones. La plus fréquente est une organisation en couches superposées : tous les neurones d'une couche sont connectés à tous les neurones de la couche suivante. Lorsqu'il y a plus de deux couches, on parle de Deep Learning.

Les données à analyser, par exemple une photographie dont il faut détecter si elle contient un chat ou un chien, sont fournies en entrée de la couche basse. Certains neurones s'activent et envoient des signaux à la couche supérieure, et ainsi de suite jusqu'à la couche la plus haute, parfois constituée d'un seul neurone. Dans l'exemple précédent, si ce neurone s'active c'est un chat, sinon c'est un chien.

Deep learning

Entrainer le réseau de neurones consiste à lui donner de nombreux exemples de photographies et à corriger légèrement la force de connexion entre chaque neurone lorsque le système fait une erreur, de façon à réduire cette erreur.

« Chez Google, des milliers de projets utilisent le Machine Learning, de l'amélioration des recommandations à la classification des images en passant la reconnaissance vocale et la traduction automatique. Cela requiert souvent des réseaux profonds de plusieurs dizaines de couches, des dizaines de milliers de neurones et des milliards de connexions. Ces réseaux doivent être ré-entrainés régulièrement, parfois quotidiennement afin de s'adapter aux changements des contenus et des usages, ce qui requiert une puissance de calcul très importante. Cela a bien sûr un coût, mais l'amélioration des services que le Machine Learning permet le justifie amplement ».

Aurélien Géron, Product Manager chez Google France
Le Machine Learning existe chez Google depuis 10 ans et cela fait 6 ans que la société l'utilise et le développe intensivement.

Deep learning

«*Beaucoup des scénarios catastrophes sont élaborés par des gens qui ne connaissent pas les limites actuelles du domaine. Or les spécialistes disent qu'ils sont loin de la réalité*», estime le Français Yann LeCun, professeur à l'université de New York et également directeur du laboratoire de recherche en IA de Facebook.

Les GAFAM : Google, Amazon, Facebook, Apple et Microsoft investissent dans l'intelligence artificielle, dans la robotique et le transhumanisme au moins pour quelques uns d'entre eux, avec comme vision une hybridation de l'homme et de la machine pour en faire un homme augmenté.

Apprentissage deep learning/ très différent apprentissage humain

Entrainer le réseau de neurones consiste à lui donner de nombreux exemples de photographies et à corriger légèrement la force de connexion entre chaque neurone lorsque le système fait une erreur, de façon à réduire cette erreur.

« Chez Google, des milliers de projets utilisent le Machine Learning, de l'amélioration des
recommandations à la classification des images en passant la reconnaissance vocale et la
traduction automatique. Cela requiert souvent des réseaux profonds de plusieurs dizaines de couches, des dizaines de milliers de neurones et des milliards de connexions. Ces réseaux doivent être ré-entrainés régulièrement, parfois quotidiennement afin de s'adapter aux changements des contenus et des usages, ce qui requiert une puissance de calcul très importante. Cela a bien sûr un coût, mais l'amélioration des services que le Machine Learning permet le justifie amplement ».

Connaissances, corrélations et Big data

« La connaissance artificielle résulte du Big data, via l'agrégation d'une très grande quantité de données de toutes sortes (internes, externes ; structurées, non structurées) et de traitements mathématiques, appelés Analytics, qui visent à faire ressortir des corrélations statistiques entre ces données. On peut faire beaucoup de choses avec la connaissance artificielle, notamment faire émerger de l'information implicite, jusque-là « cachée » dans les données, afin de compléter ou d'améliorer nos connaissances humaines. Mais il ne faut pas confondre « connaissance » et « intelligence », autrement dit ne pas confondre « apprendre » et « comprendre ».

Bernard Georges, Responsable de la prospective stratégique, Société Générale

Connaissances artificielles : corrélations
Connaissances sémantiques : compréhension
Choix des DATA -> neutres, non discriminantes…

Bernard Stiegler : Attention à la bêtise artificielle !!

Intelligence humaine et machine

« *Fondamentalement, l'ordinateur et l'homme sont les deux opposés les plus intégraux qui existent*. » explique Gérard Berry, informaticien et professeur au Collège de France, médaille d'or 2014 du CNRS, dans un entretien du nouvel Obs. Il explique « *L'homme est lent, peu rigoureux et très intuitif. L'ordinateur est super rapide, très rigoureux et complètement con.* ». Nouvel Observateur, 26-08-2016

L'autocomplétion (B. Stiegler)

Lorsque vous commencez à taper une requête, cela prend 1/50ème de seconde pour aller de votre cerveau à votre main, puisque le signal électrique neurologique circule à 50 mètres par seconde en moyenne sur un nerf. Sur les fibres optiques, le signal circule à 200 millions de mètres par seconde. La machine va 4 millions de fois plus vite que vous. En tenant compte des temps de latence, on peut descendre à 100 000 fois plus vite. C'est extrêmement lent pour la machine, mais cela reste extraordinairement puissant pour nous.

GT CERNA

GT Apprentissage machine/IA et Ethique :

L. Devillers, S. Abiteboul, N. Boujemaa,
D. Bourcier, R. Chatila, G. Dowek, J-G. Ganascia,
A. Grinbaum, M. Dauchet

Mars – Novembre 2016

Résumé des Objectifs

- GT Apprentissage machine et éthique (Durée 6-9 mois) - Saisine
- Composition : chercheurs en apprentissage-machine, IA, informatique, philosophie, droit
- Auditions avec un questionnaire préétabli par le GT
- Interviews (experts) – lien avec AFIA
- Journée d'étude le 13 juin 2016

Rapport (15 pages, 2 rapports * AN/FR) :

- Langages/concepts expliqués par des chercheurs pour des non experts & experts.
- Préconisations éthiques pour les chercheurs, les concepteurs, les développeurs (robotiques, bots, objets connectés), pour les industriels/utilisateurs et pour la réglementation des systèmes (vision internationale & nationale)