

# L'intelligence artificielle, entre mythes et réalités

*Une Intelligence Artificielle peut en cacher une autre*

Amal El Fallah Seghrouchni

<http://www.digital-ia.org>

Head of **Ai** movement

International Artificial Intelligence Center of Morocco

UM6P – Morocco







<https://aim.um6p.ma>

Full Professor @ Sorbonne University

Co-head of AI and Data Science department @ LIP6 – CNRS - Paris - France



# Plan

-  Introduction à l'Intelligence Artificielle (IA)
-  Une IA peut en cacher une autre
-  Grands domaines de l'IA
-  Quelques projections
-  Quid de l'éthique de l'IA ?
-  Conclusion



# Plan

 Introduction à l'Intelligence Artificielle (IA)

 Une IA peut en cacher une autre

 Grands domaines de l'IA

 Quelques projections

 Quid de l'éthique de l'IA ?

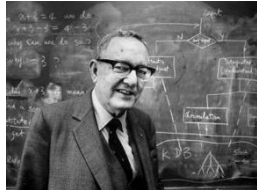
 Conclusion



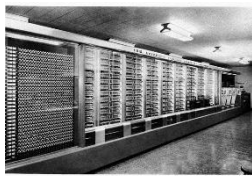
# Historique de l'IA



Le test de Turing (1950)



Satisficing (1956)



Perceptron (1960)



PDP-10 (1966)



Deep Blue (1997)



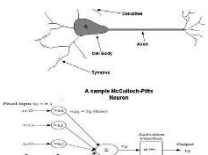
Cortana (2014)



Siri et Google Now (2011)



IBM Watson /Jeopardy (2011)



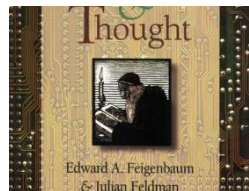
Réseau neuronal (1943)



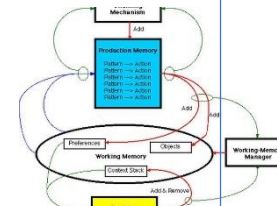
The Logic Theorist (1956)



LISP (1958)

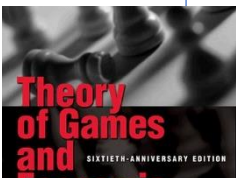


Reconnaissance et adaptation (1963)



Soar (1983)

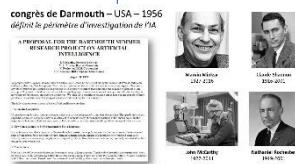
Game Theory (1944)



Konrad Zuse Z3 (1941)



Darhmouth Conference (1956)



The General Problem Solver(1957)

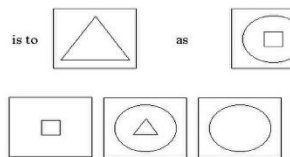
Jeu de dames (1951)



**General Problem Solver (1957)**

- Generalization of Logic Theorist
- general framework for solving any problem, given the proper setup
- Focus on word to non-linguistic base
- "trial and error" - learn from domain-specific knowledge
- goal: to be a better psychological model of human cognition
- have not just describe what they are doing but they solving problems
- Microscopic analysis
  - determine differences between current state & goal state
  - break into sub-problems (divide & conquer)
  - reduce the differences by means of operators
- Backtracking

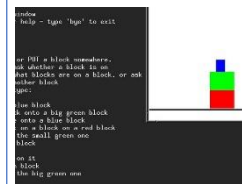
ANALOGY (1963)



Le premier système expert (1965)



SHRDLU (1968)



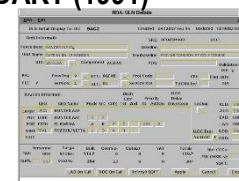
Prolog (1972)



Smart Truck (1989)



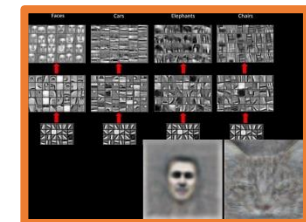
DART (1991)



Google Car (2009)



Deep Learning (2012)



Google AlphaGO (2016)



# Naissance de l'IA

## Summer Camp de Darmouth

USA New Hampshire – été 1956

*définit le périmètre d'investigation de l'IA*

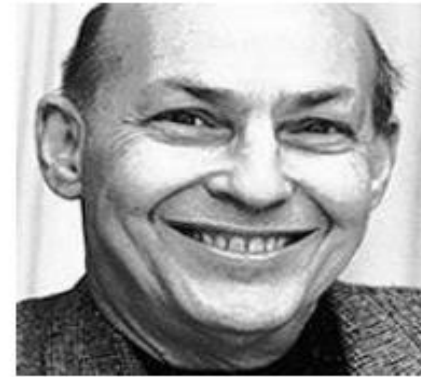
### A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College  
M. L. Minsky, Harvard University  
N. Rochester, I.B.M. Corporation  
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

« 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. »

- Automatic computers
- Use a language
- Neuron nets
- Theory of the size of calculation
- Self-improvements
- Abstractions
- Randomness and creativity



**Marvin Minsky, MIT**  
1927-2016



**Claude Shannon, Bell**  
1916-2001



**John McCarthy, MIT**  
1927-2011



**Nathaniel Rochester, IBM**  
1919-2001

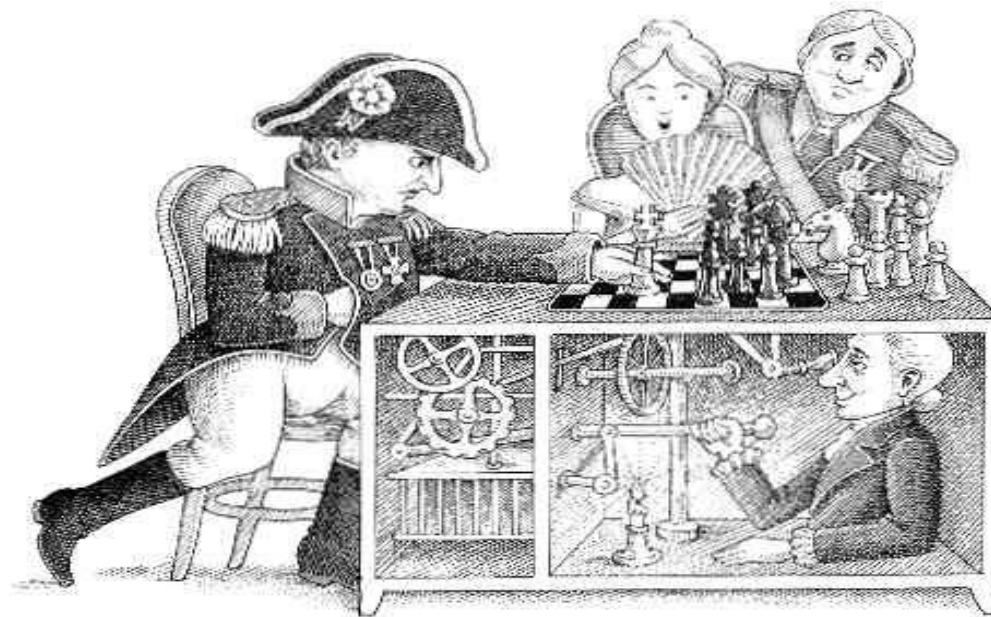
**Etudier les processus cognitifs humains et les reproduire par la machine**



# Concept de base de l'IA : l'agent intelligent

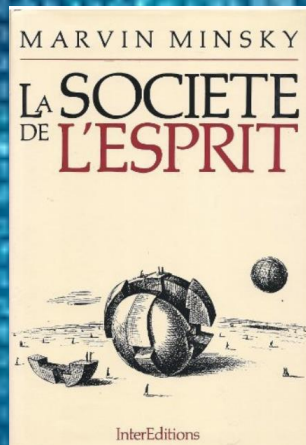
L'IA est la science et l'ingénierie de la construction de machines intelligentes

- comportements calculables, intelligents et proches de ceux des humains
- une machine est dite intelligente si elle arrive à se faire passer pour un humain auprès d'un autre humain : **test de Turing en 1950**



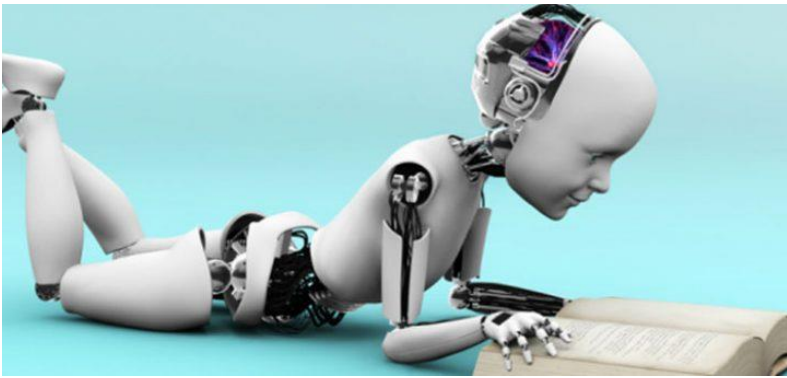
# Une définition de l'IA

**Marvin Minsky:** "Artificial intelligence is the science of making machine do things that would require intelligence if done by man"



*percevoir*  
*information riche, complexe et subtile*  
*apprendre*  
*avec l'environnement*  
*abstraire*  
*pour créer de nouveau sens*  
*raisonner*  
*pour décider et planifier*  
*communiquer*  
*Pour dialoguer*  
*agir*  
*pour réaliser des objectifs rationnels*

# 4 façon de faire de l'IA



**Concept d'agent**

Penser  
rationnellement

Agir  
rationnellement

Penser comme  
des humains

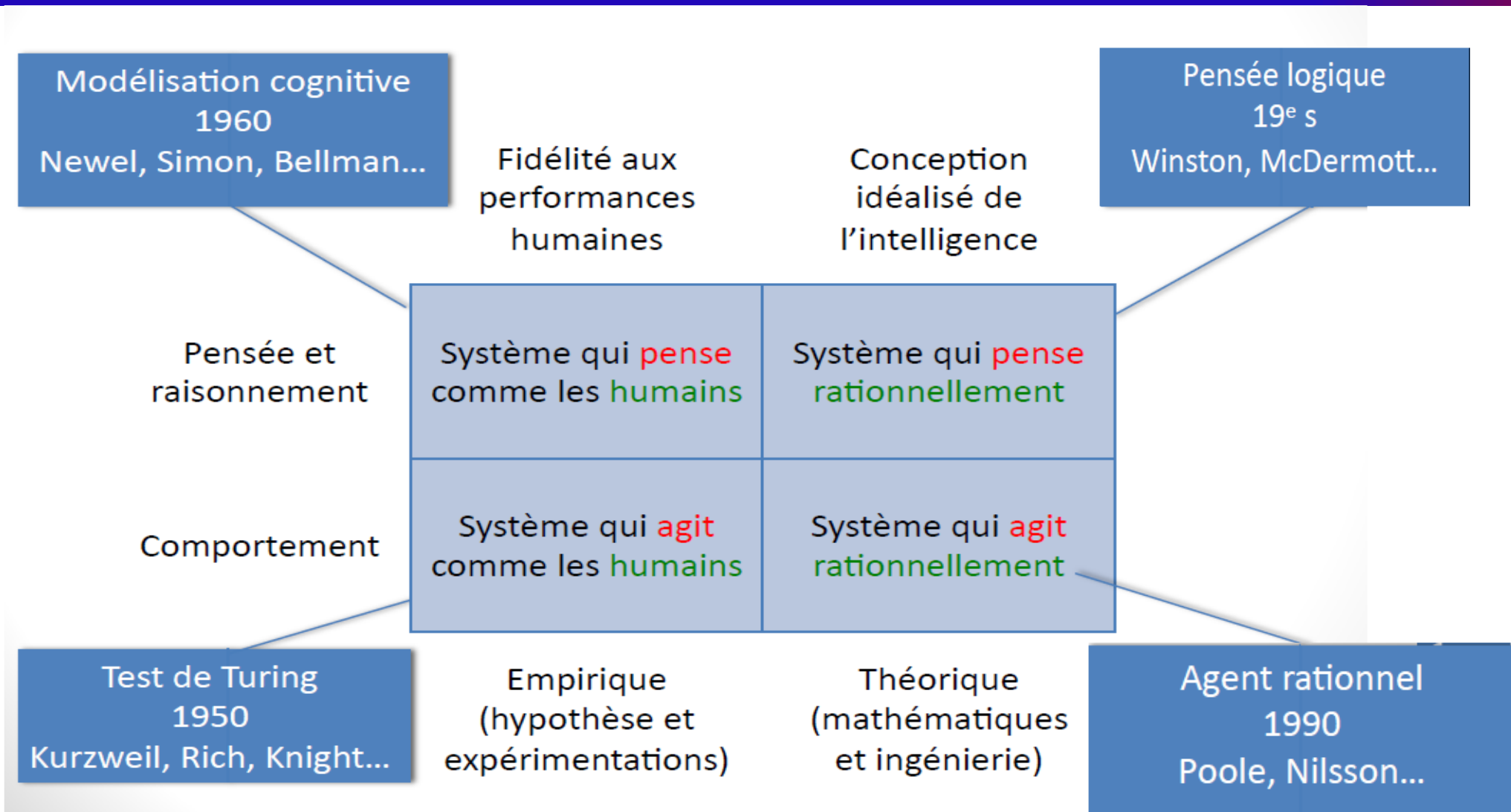
Agir **comme** des  
humains

[Russel, 2010]





# 4 façon de faire de l'IA



[Marie Lefevre, 2001]



# IA faible et IA forte



Le test de Turing a produit deux thèses sur l'IA :



## Thèse **faible**

- Il est possible de construire des machines qui se comportent comme si elles étaient intelligentes
- Vision raisonnable :
- Un programme peut être capable de raisonner, d'apprendre et même de résoudre des problèmes
- Un programme ne « pense » pas, il exécute



## Thèse **forte**

- Les machines qui ont un comportement intelligent possèdent des états cognitifs
- Capable d'éprouver une réelle conscience de soi,
- de ressentir de vrais sentiments et comprendre ce qui la pousse à faire telle ou telle façon
- Nécessite de savoir apprendre et modifier son comportement



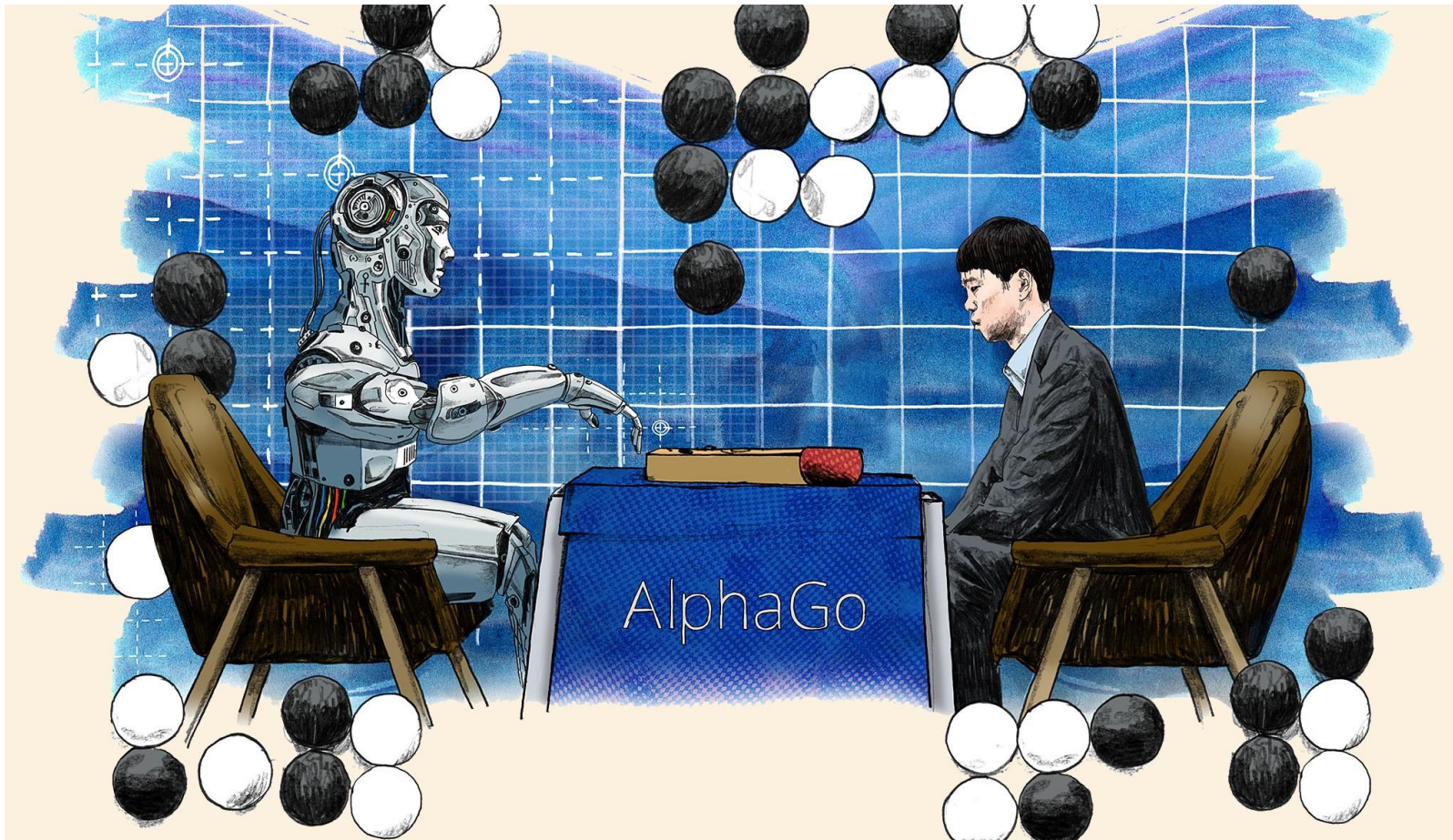
# Deep Blue d'IBM, un succès historique

 Garry Kasparov, grand maître russe et champion du monde des échecs affronte Deep Blue (IBM) lors de 2 matchs.

- G.K. gagne au 1er match (1996) contre *Deep*
- G.K. perd au 2nd (1997) contre *Deeper* Blue avec un score 3,5–2,5
- « *Deeper Blue* » d'IBM entre au **musée de l'histoire de l'ordinateur** de Mountain View



# AlphaGo de Google, un autre succès historique



20 ans après le succès de *deeper blue*



# AlphaGo (Google DeepMind)


 **27 Mai 2017** : AlphaGo bat le champion du monde **Ke Jie**

- AlphaGo combine des
  - techniques d'apprentissage **supervisé** alimenté par des parties de référence jouées par des humains,
  - techniques de parcours de **graphe**,
  - et **entraînements** avec des humains, d'autres ordinateurs, **et surtout lui-même**.



 **Octobre 2017** : AlphaGo Zero atteint un niveau supérieur

- En utilisant l'apprentissage par **renforcement**
- En jouant uniquement contre lui-même.

 **Décembre 2017** : AlphaGo Zero surpasse largement le niveau de tous les joueurs humains et logiciels dans les jeux de Shōgi, le jeu de go et les échecs



# Apprentissage supervise versus non supervisé



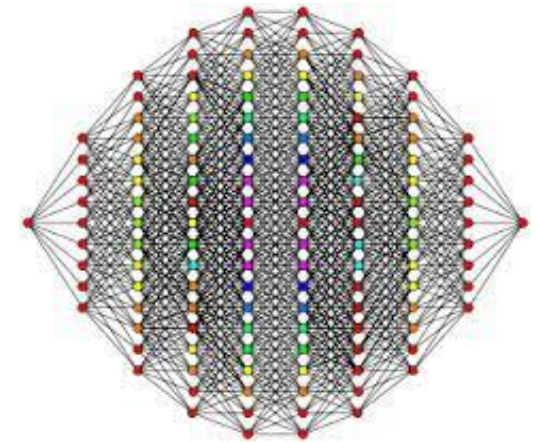
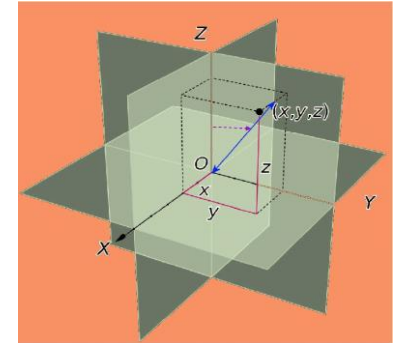
## Apprentissage supervisé

- Des données (data set) annotées
- Relativement aisé dans un fichier de données structurées
- Difficile pour des données non structurés ( image, vidéos..)
- Très difficile si beaucoup de données (deep learning)



## Apprentissage non supervisé

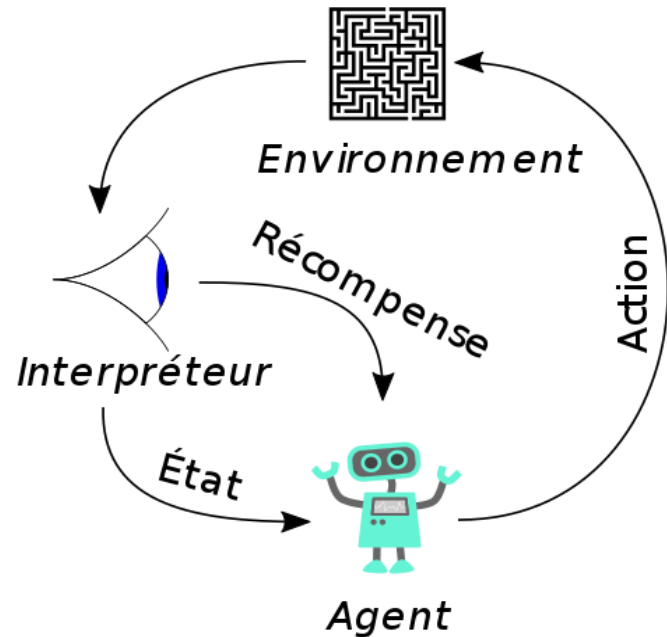
- Pas de données annotées
- L'algorithme va constituer des clusters: un ensemble d'éléments hétérogènes sous forme de sous groupes homogènes ou liés par des caractéristiques communes (features)
- Ex. recommandation sur la base de similarité (des amis ont apprécié le même produit)



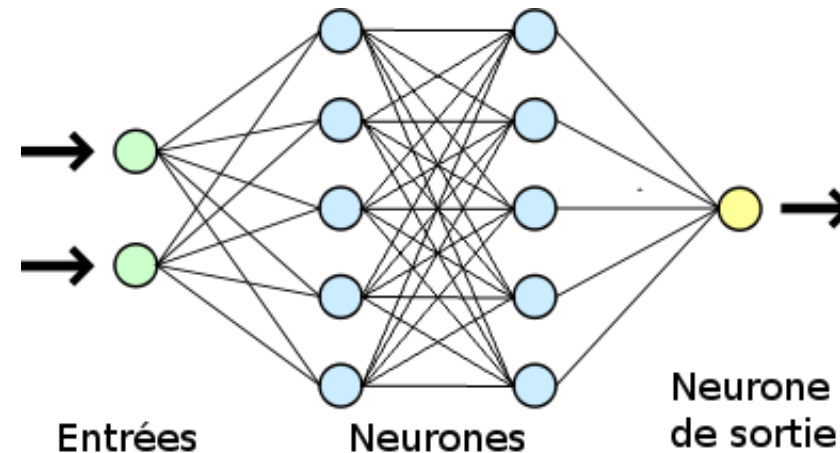
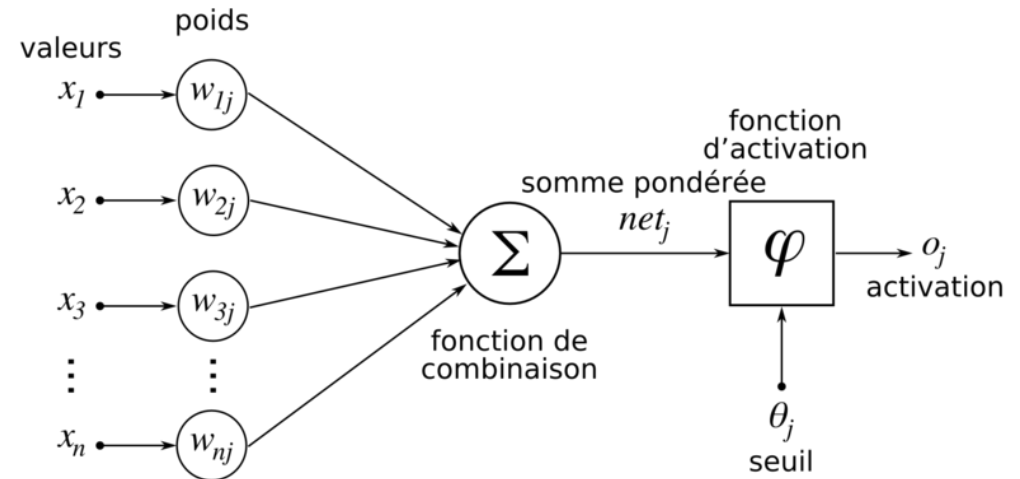
high-dimensional  
Euclidean space



# De nombreux modèles d'apprentissage

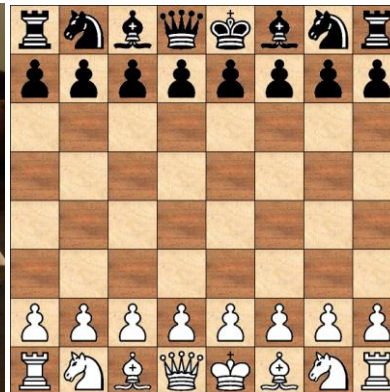


Apprentissage par renforcement

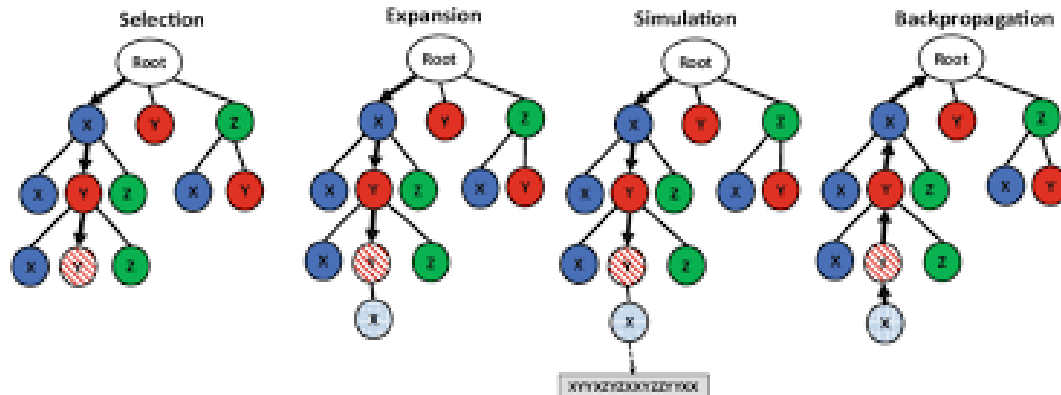


Réseaux de neurones

# Défi de l'incertitude



Shōgi, go et échecs ont un point commun



Il faut une sacrée force brute  
Une gigantesque puissance de calcul  
pour traiter la complexité croissante!

- Le Jeu d'échecs : le nombre de parties possibles est estimé à  $10^{123}$ .
- Le nombre d'atomes dans l'univers ne dépasserait pas  $10^{80}$

Arrivée des GPU  
TPU (tensor processus units)





# De plus en plus de succès..

 Récemment, Dr Fill gagne le tournoi de mots croisés aux Etats-Unis

- « Une IA bat un humain à plate couture aux mots croisés »

 L'IA investit le sport

- **entraînement des joueurs** : niveau de forme, performance individuelle ;
- **stratégie de l'équipe** : analyse des matchs passés (vidéos, captation de données) ;
- **gestion de l'effectif** : transferts, valeur marchande des joueurs ;
- **match** : résultat, actions, nombre de passes, de centres, de tirs, d'arrêts de gardien, avertissements, longueur des courses réalisées par les joueurs ;
- **compétitions** : résultats des rencontres, classements des équipes, des joueurs, calendrier des matchs ;
- **événements physiques** : nombre de spectateurs présents, météo, état du terrain ;
- **retransmissions** : audiences télévisuelles ;
- **animation des communautés par les clubs** : abonnés au stade, fans clubs ;
- **activités des clubs sur les réseaux sociaux** : nombre de fans, engagement, interactions avec les supporters
- **dimensions économiques** : comptes de résultats des clubs, droits télévisuels ; merchandising et e-commerce ;
- **activités liées aux paris en ligne** : côte des matchs, cote des équipes.



# Entre fascination et inquiétude

- L'IA réussit des tâches communément réalisées par des humains
  - analyser une image ou du texte, naviguer, planifier, prédire, recommander, « voir », « parler »...

• Tâches proches de productions humaines ~ intelligentes ~ **l'IA peut remplacer l'homme**

- L'IA réalise une tâche qu'un humain ne peut pas effectuer sans l'aide de machines : ne pas remplacer mais assister davantage

- Mais de nombreux dérapages: L'assistant de MS qui devient raciste ...



# Plan

 Introduction à l'Intelligence Artificielle (IA)

 **Une IA peut en cacher une autre**

 Grands domaines de l'IA

 Quelques projections

 Quid de l'éthique de l'IA ?

 Conclusion



# Pyramide Cognitive

[Pyramide Cognitive, Bernard Georges]

**CONSCIENCE** : capacité de sentir et de penser

Sentiments (e.g. emotions)  
Réflexion ( intelligence, reflexivity moi ≠ autres, introspection)

**INTELLIGENCE** : capacité de comprendre et de s'adapter

Prospection ("poser les bonnes questions", être empathique)  
Abstraire les K et trouver les connexions et les corrélations

**CONNAISSANCE** : corrélation des informations + mémorisation

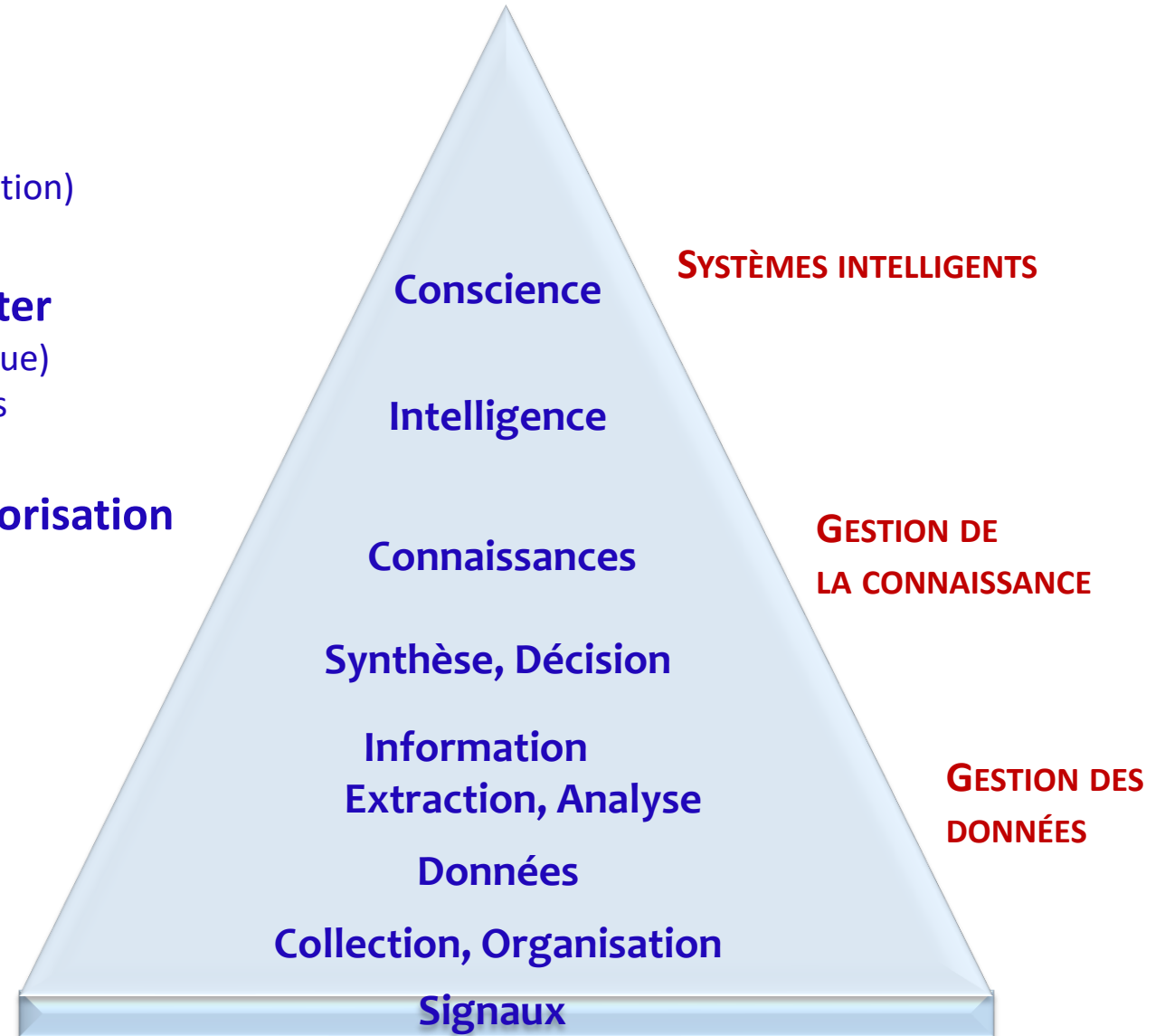
Apprendre, prédire en extrapolant la connaissance  
"trouver les bonnes réponses"

**INFORMATION** : données significantes

Modélisation, prédiction

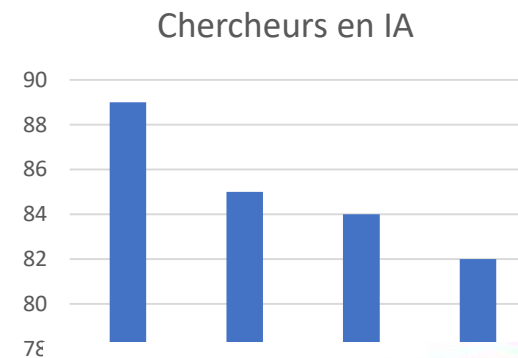
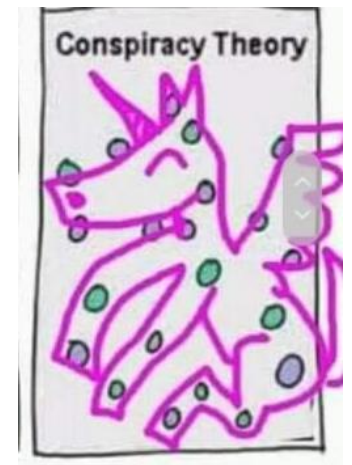
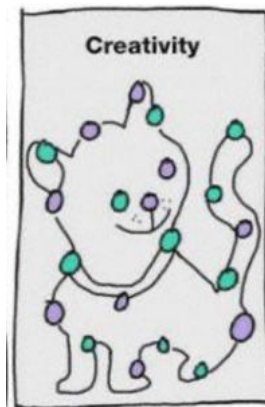
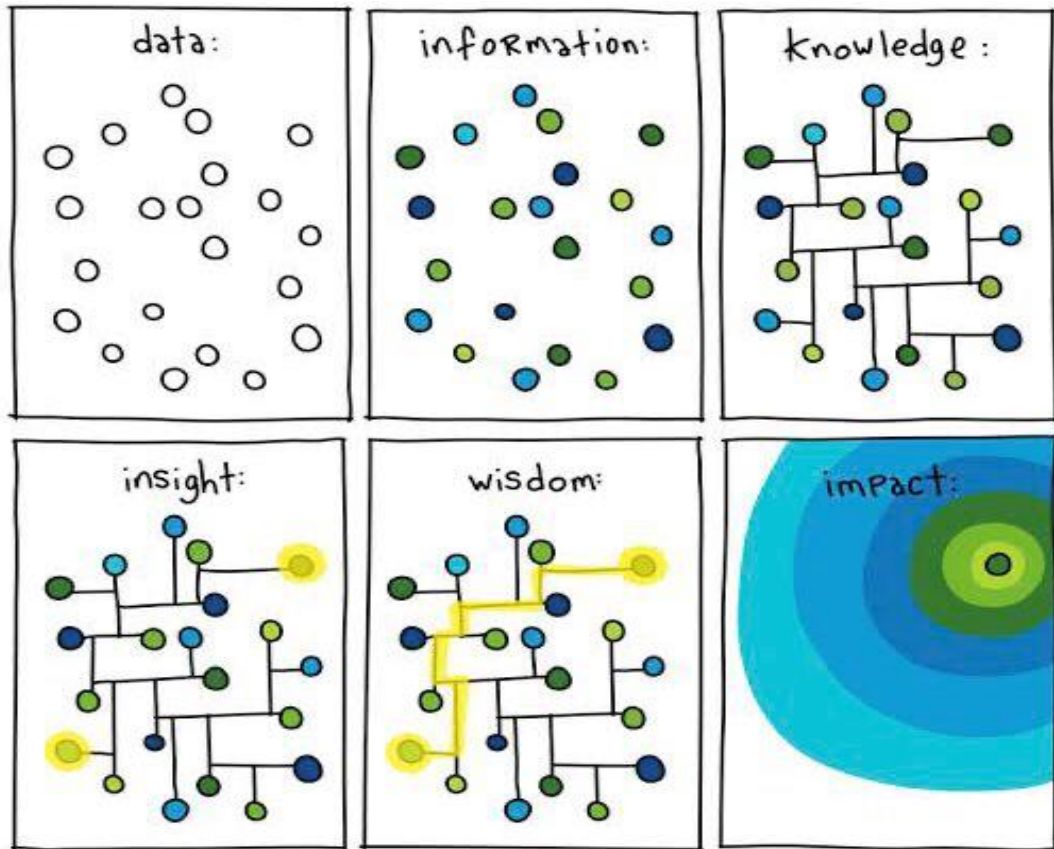
**DATA** : représentation (conventionnelle)

Représenter quelque chose



# En d'autres termes..

Utiliser les bonnes données et les bons algorithmes pour générer les bonnes connaissances



Les chei

es

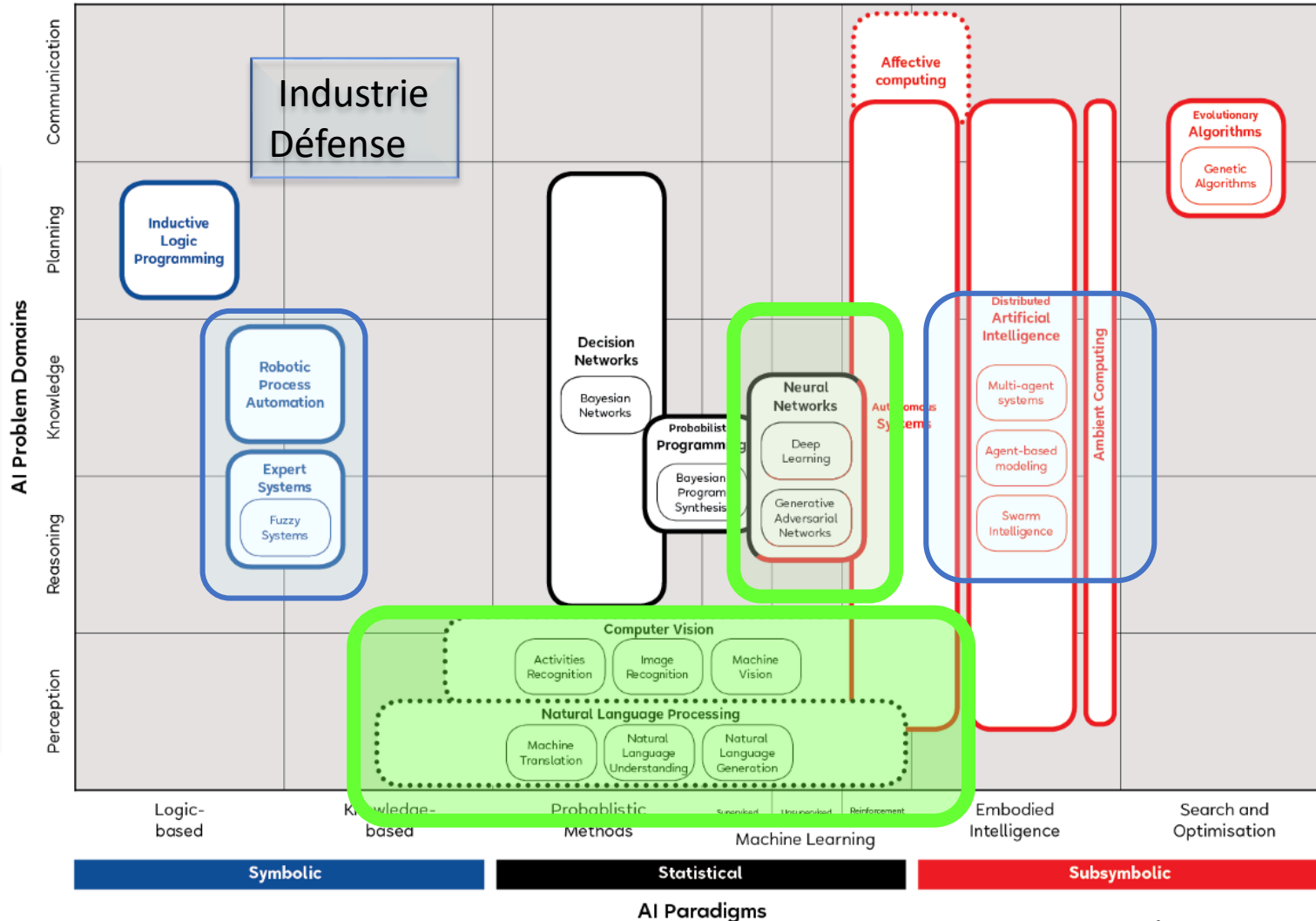
**FAUX**

Les chercheurs en IA vivent longtemps entre 82 et 89 ans



# 3 Paradigmes de l'IA

**Intelligence**  
 Autonomie  
 Représentation des connaissances  
 Raisonnement  
 Action  
 Décision  
 Apprentissage  
 Optimisation,  
 Simulation, ...



**Système**  
 Interaction  
 Perception  
 Communication  
 Multi-Modalité  
 Représentation des autres  
 ...

Reconnaissance Faciale  
 Modèles prédictifs  
 Chatbots

Francesco Corea –2018



Amal El Fallah Seghrouchni

—[CHÔRA]— chora.space

⊙ Narrow Applications  
 ○ General Applications  
 ○ Subtype

# Terradynamica : 3 approches complémentaires

1. IA dirigée par les données
2. IA symbolique
3. IA hybride

Simulation  
IA symbolique  
Architecture cognitive  
SMA  
Emotions  
Navigation  
Planification  
...

## Comment apprehender la complexité ?

Transport :  
- bus  
- Tram  
- ...

Traffic :  
- cars  
- Bicycles  
- ...



Signalisation :  
- Lights  
- ...

Humans:  
- isolated persons  
- groups  
- crowd  
- ...

*TerraDynamica*

*Animer la vie dans la ville (100 000 acteurs, > 10M€)*



# Plan

 Introduction à l'Intelligence Artificielle (IA)

 Une IA peut en cacher une autre

 **Grands domaines de l'IA**

 Quelques projections

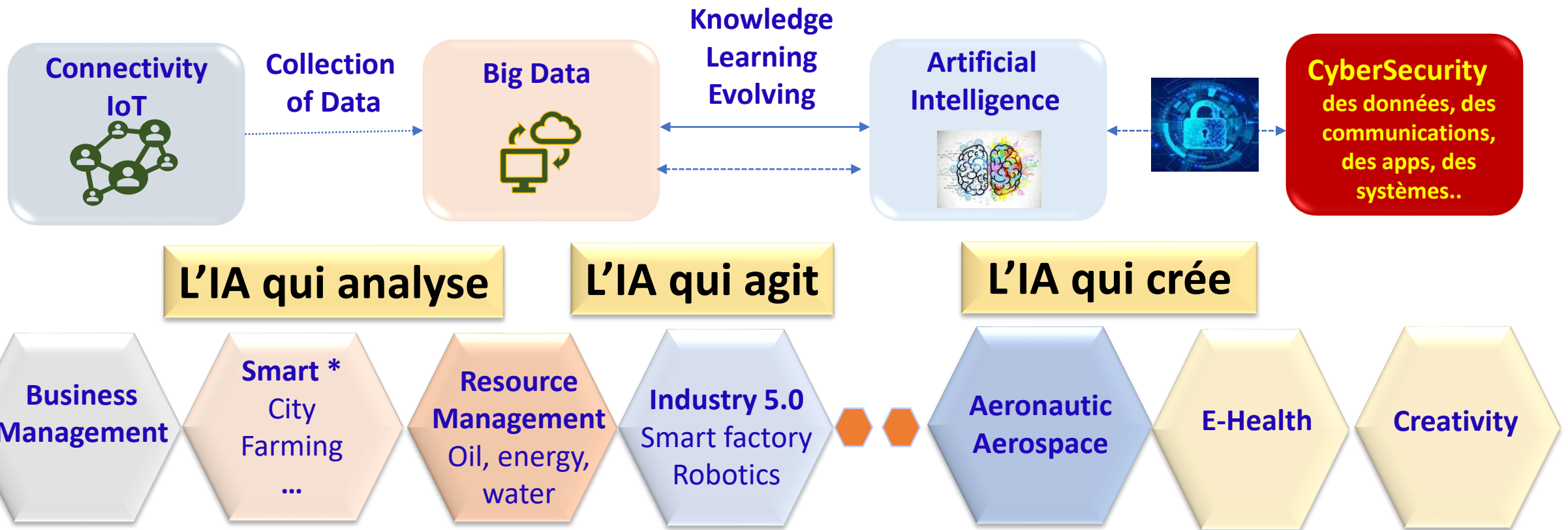
 Quid de l'éthique de l'IA ?

 Conclusion





# Grands domaines de l'IA



# Grands domaines de l'IA

Création de la valeur, automatisation et optimisation des tâches, des processus et des gains..

## sources de données

- Imagerie satellitaire
- Données du terrain
- Données climatiques et topologiques
- Données socio-économiques
- Données démographiques
- Données d'entreprises
- Données sur les médias sociaux
- etc.

## Analyse et prédiction

- Agriculture de précision
- Services publics
- Services financiers (banque et assurance)
- Business management
- Changement climatique
- Economie circulaire
- E-santé
- ...

## Actions et planification

- Gestion des ressources, énergies renouvelables et eau
- Robotique cognitive
- Industrie 4.0
- Défense
- Aéronautique
- Aérospatiale
- Etc.

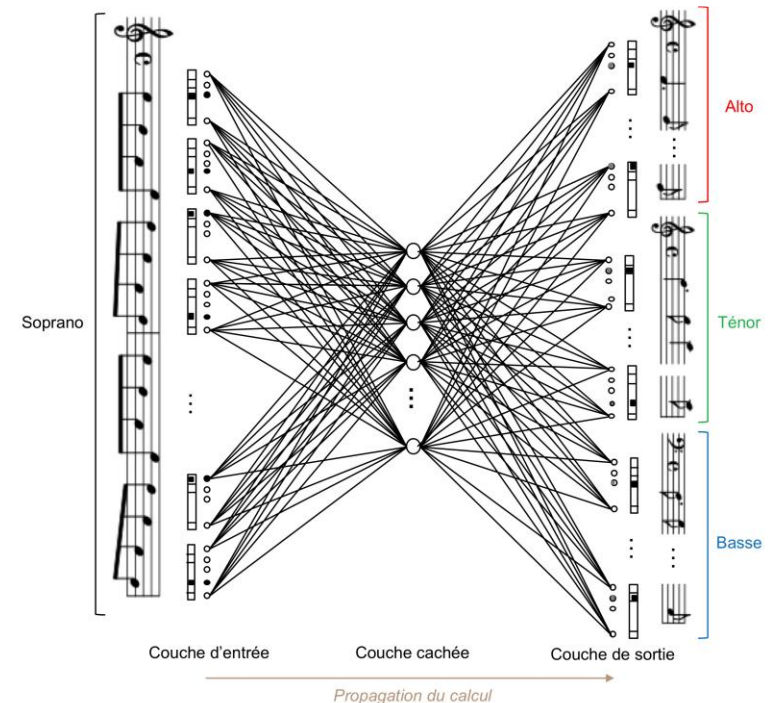
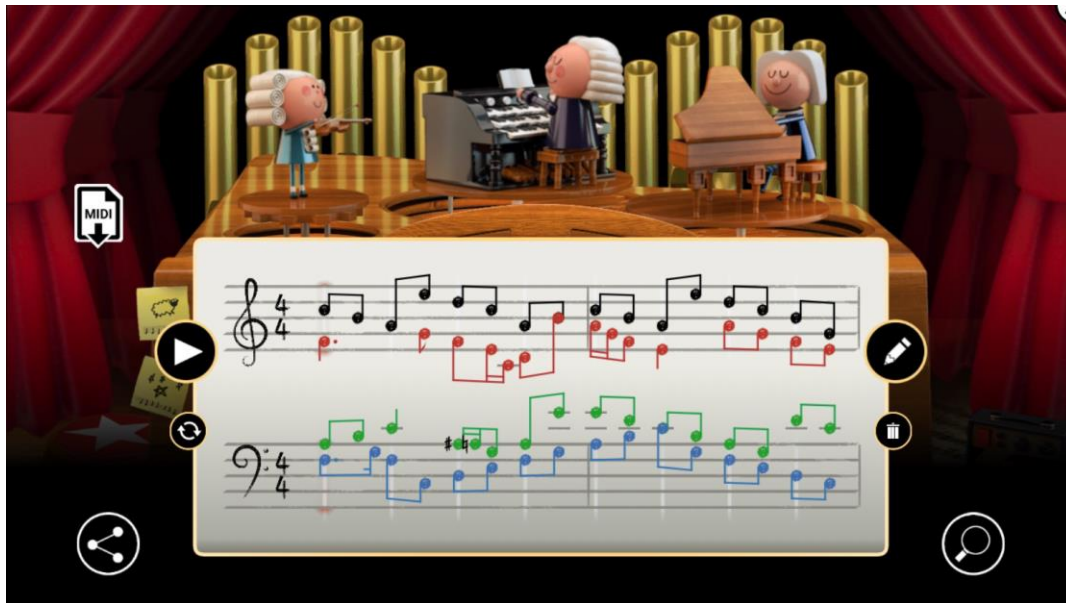
Multi-sources, multi-médias, hétérogènes, temporelles

Génération de contenu, art et créativité

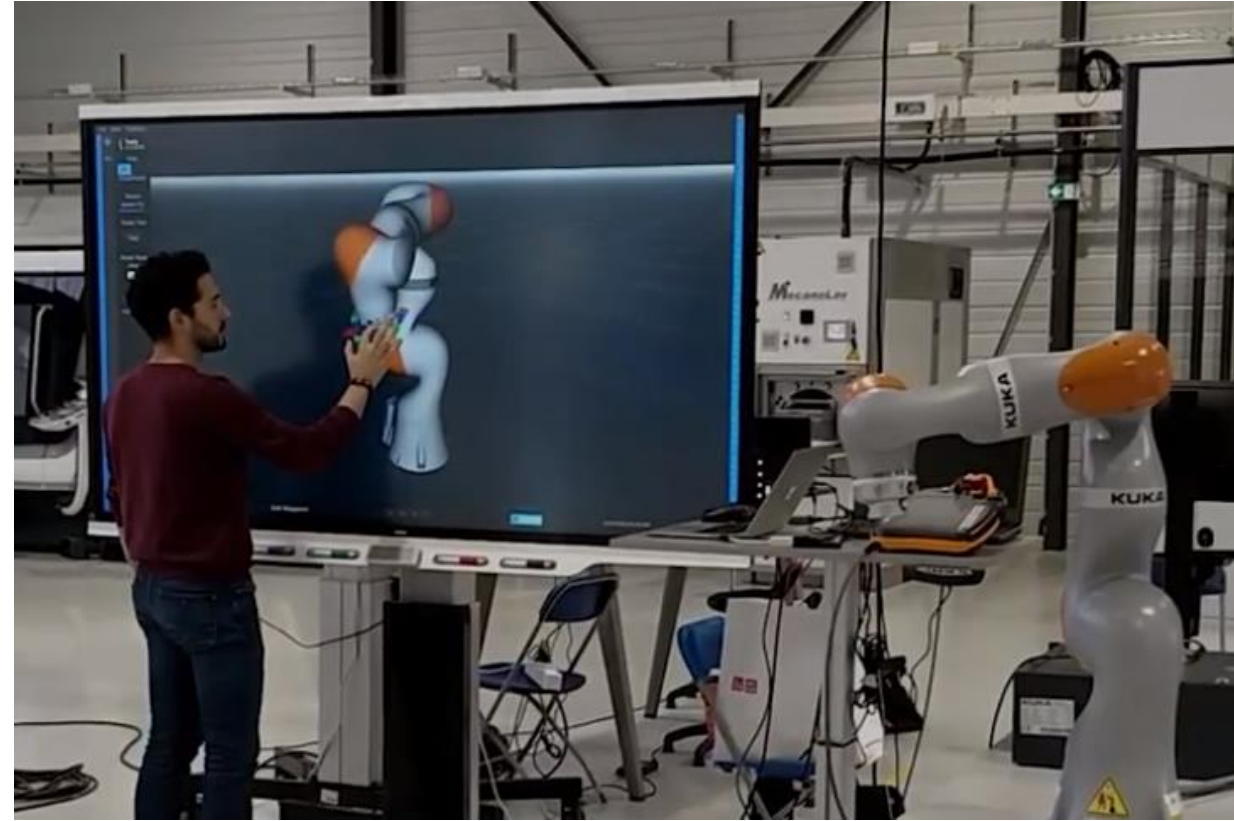


# Un nouveau domaine d'excellence de l'IA

- La création de contenus : quelques secondes pour générer l'hymne national
- La composition, autorise la rature et peut ne jamais aboutir à une œuvre définitive.
- L'improvisation est une performance, un défi car le musicien doit produire en temps réel.



# Intelligence Artificielle, Robotique Cognitive et co-botique

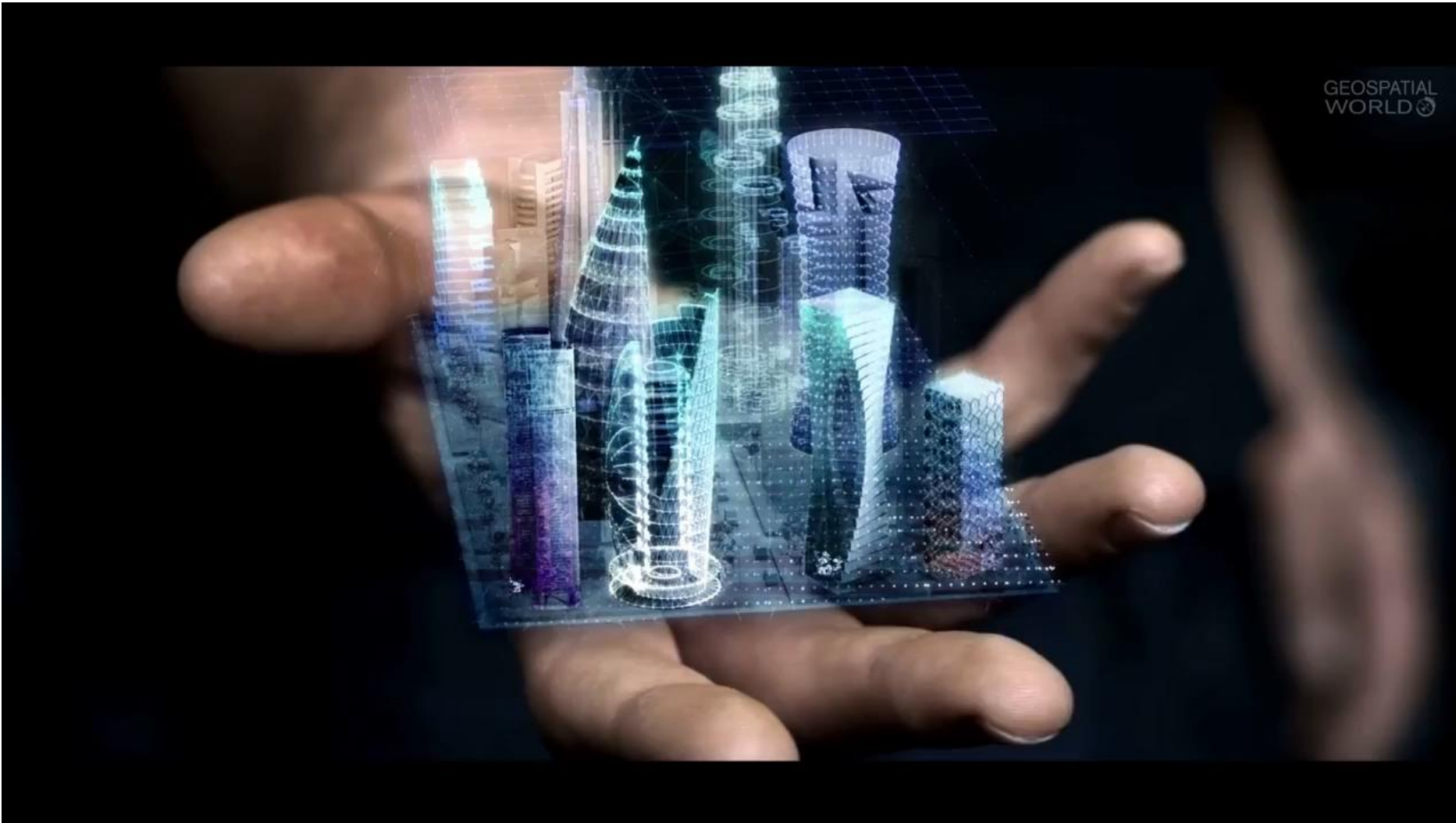


Un domaine en plein essor



# Jumeau numérique = rencontre entre le monde réel et virtuel

Digital twin





# Discours de F. Parly, ministre des Armées sur IA de défense 12/05/2021


- 🇫🇷 Il y a 2 ans, j'avais fait devant vous un constat : celui d'une **course mondiale à l'intelligence artificielle**, une course effrénée où le domaine de la défense occupe une **place stratégique**.
- 🇫🇷 Nous avons bien compris que nous étions face à un **nouveau virage technologique, économique, éthique aussi**.
- 🇫🇷 .. Car nous savons que **l'IA est stratégique**. Elle est **indispensable pour garantir la supériorité opérationnelle** des armées de demain. Bâtir une IA appliquée à la défense, maîtrisée et souveraine, tel est notre défi.
- 🇫🇷 **L'IA est une arme redoutable** en ce sens qu'elle a **le pouvoir de décupler la force de frappe de toutes les autres**.
- 🇫🇷 Ce qui comptera prochainement, ce ne sera plus uniquement le nombre de chars d'assaut, de missiles ou de moyens de surveillance dont on dispose, **mais la façon dont on coordonne leur emploi, dont on optimise leur déploiement sur une zone définie**. C'est toute la force du **combat collaboratif** que nous développons dans le cadre du **programme Scorpion ou encore du SCAF**.



# Discours de F. Parly, ministre des Armées sur IA de défense 12/05/2021

 Au-delà de la **collecte et du stockage des données**, les armées, nos services de renseignement, la DGA, ont bien saisi l'impérieuse nécessité de **les structurer, les analyser, les corrélés**, pour en **extraire le renseignement dont nous avons besoin, optimiser nos capteurs, gagner en performance grâce à la maintenance prédictive** ou encore **aider à la planification de nos missions**. Bien évidemment, les applications sont nombreuses et il s'agit là que de **quelques illustrations de l'apport de l'IA pour éclairer nos décisions et conduire nos opérations**.

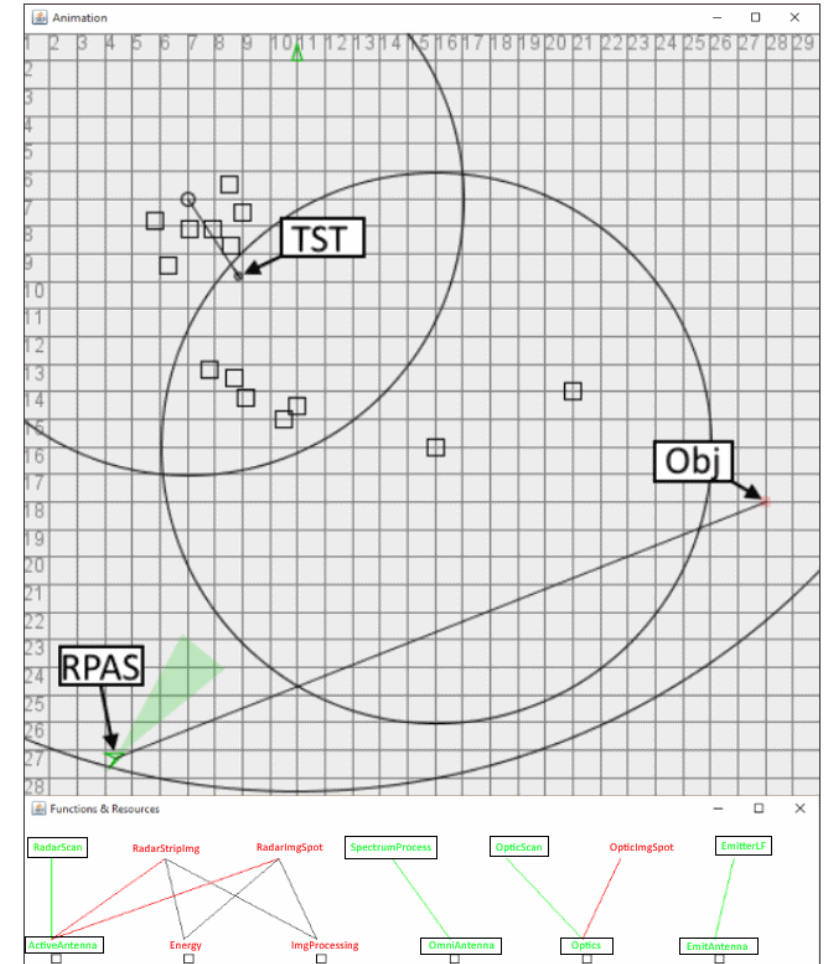
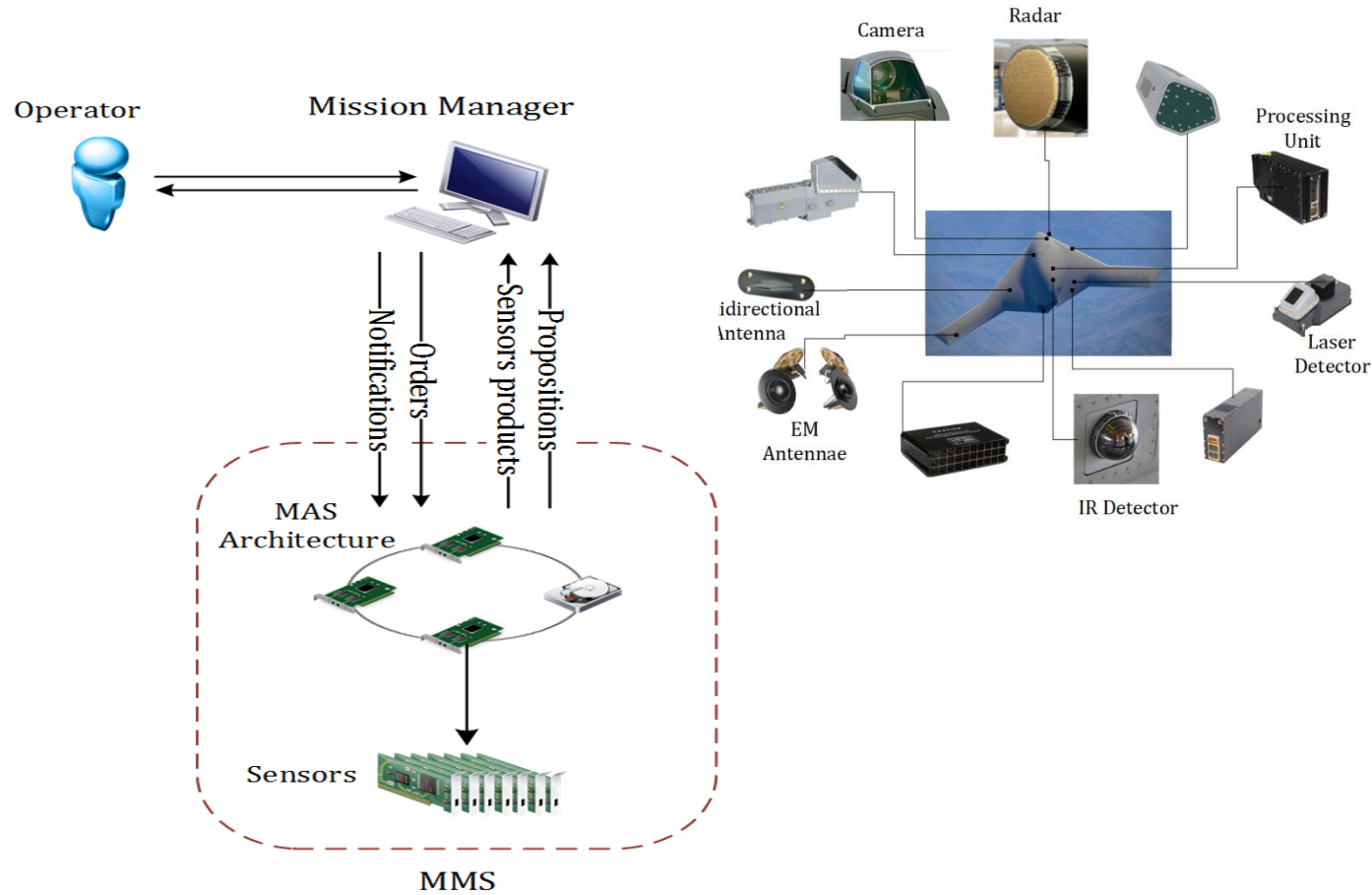
 Au-delà des applications qui irriguent nos systèmes d'armes qui apportent une **aide à la planification opérationnelle**, nous avons recours à l'IA pour des **métiers administratifs dans les finances ou encore les ressources humaines**

 **IA de confiance** : La France dit et dira toujours **non aux robots-tueurs**. La France refuse de confier la décision de vie ou de mort à une machine qui agirait de façon pleinement autonome et échapperait à tout contrôle humain.



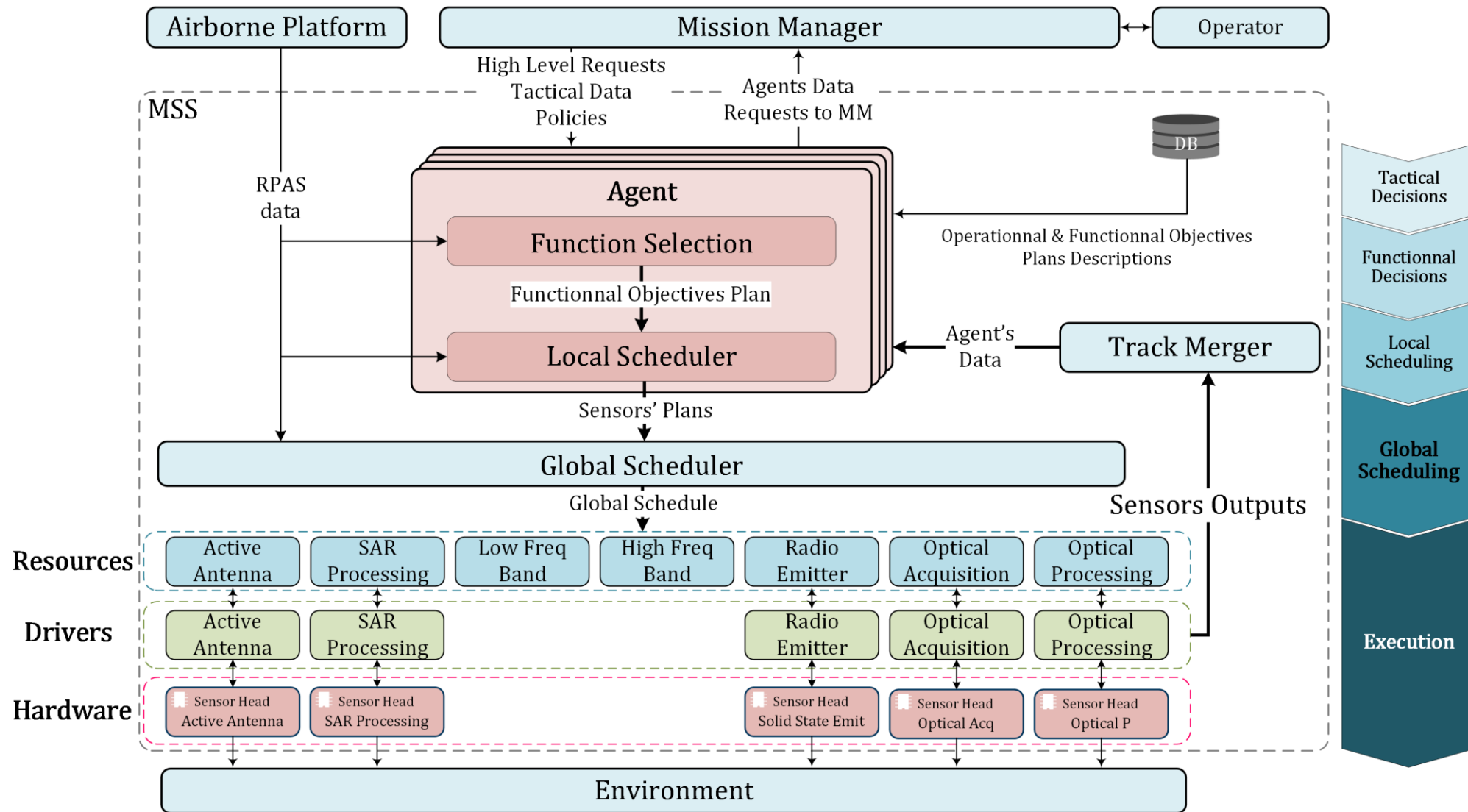
# RAMSES collaboration avec Thales (2014-2017)

Architecture of the new generation of airborne platforms - Optimization of resources (sensors)

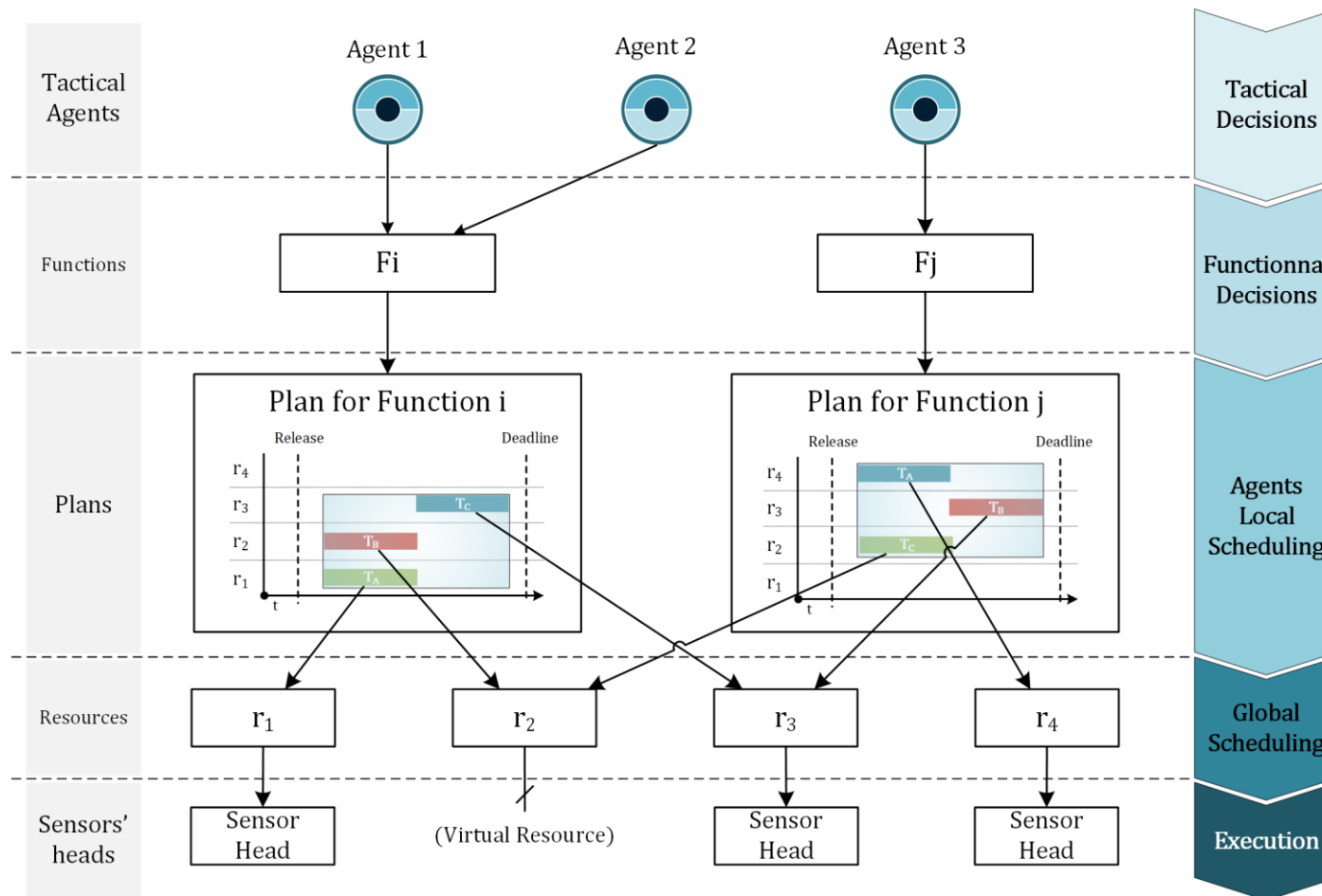




# Systeme multi-senseurs multi-fonctions



# Ordonnancement des Plans des Fonctions



Intelligence artificielle distribuée ou SMA

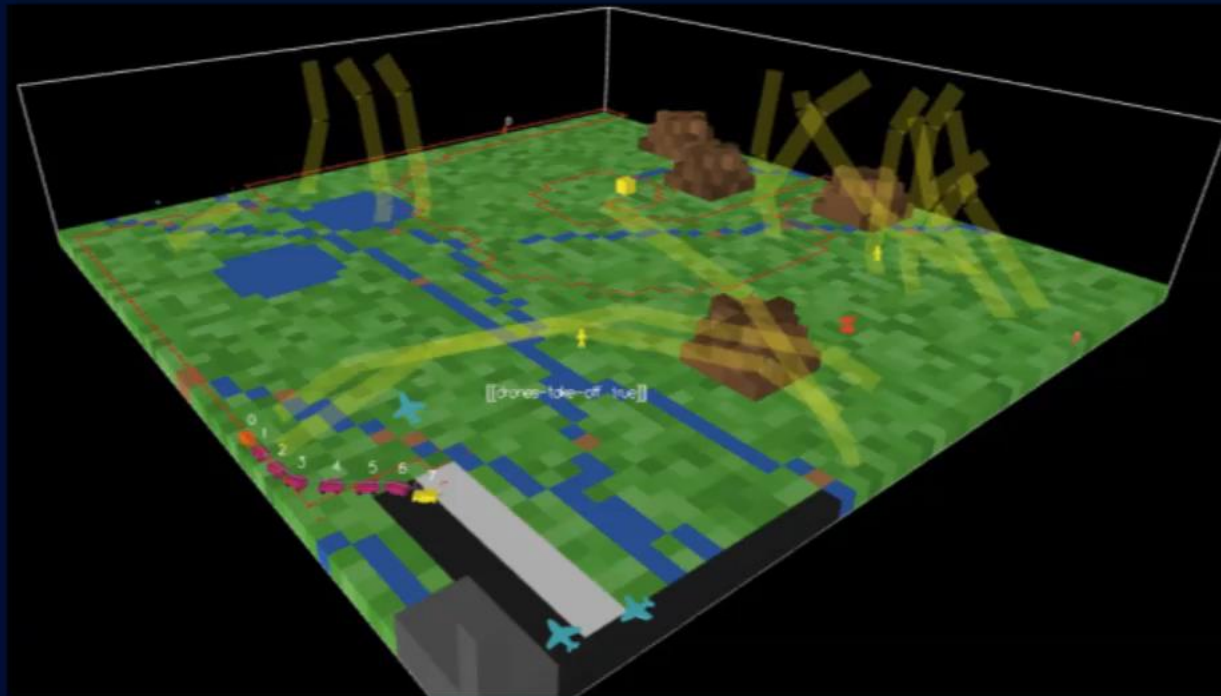


UE Systèmes Multi-Agents (CoCoMa)

Responsable: Mme Amal El Fallah Seghrouchni

## Mission de drones pour l'escorte de convois







Projet de simulation multi-agents en NetLogo

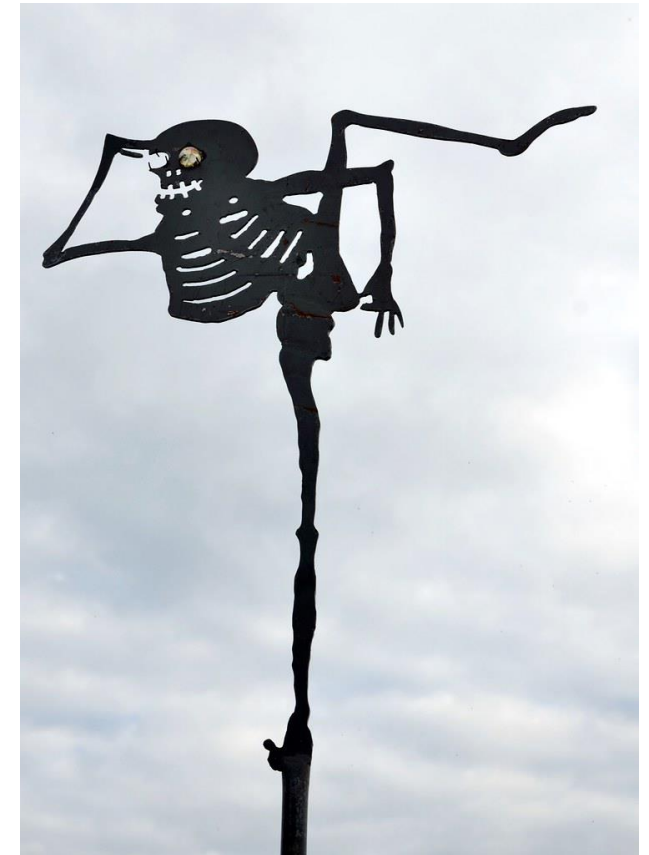


Issam BANI, Olivier HOTEL, Stephen LARROQUE

[github.com/LRQ3000/netlogo-mas-drones](https://github.com/LRQ3000/netlogo-mas-drones)

# Plan

-  Introduction à l'Intelligence Artificielle (IA)
-  Une IA peut en cacher une autre
-  Grands domaines de l'IA
-  Quelques projections
-  Quid de l'éthique de l'IA ?
-  Conclusion

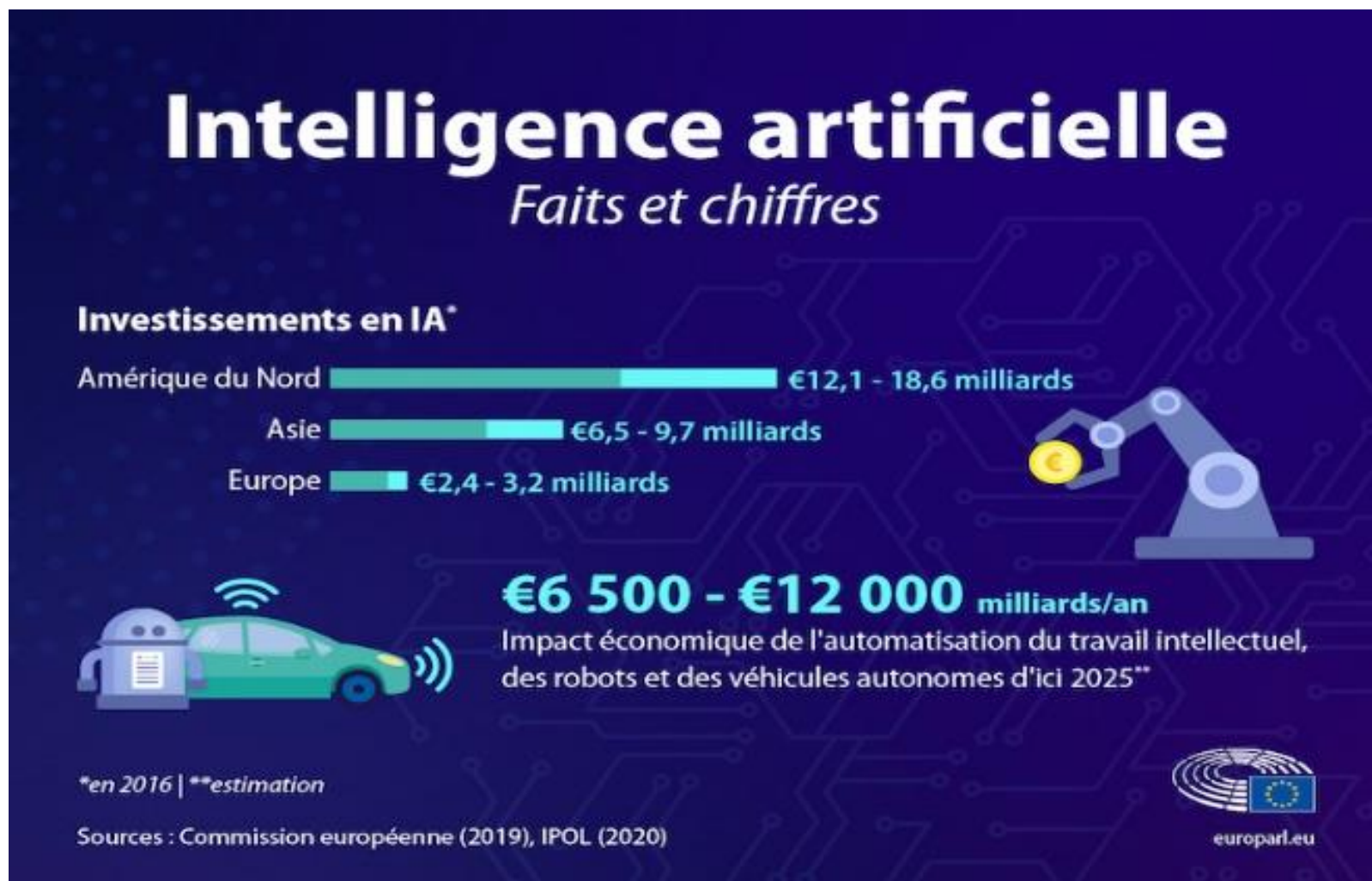
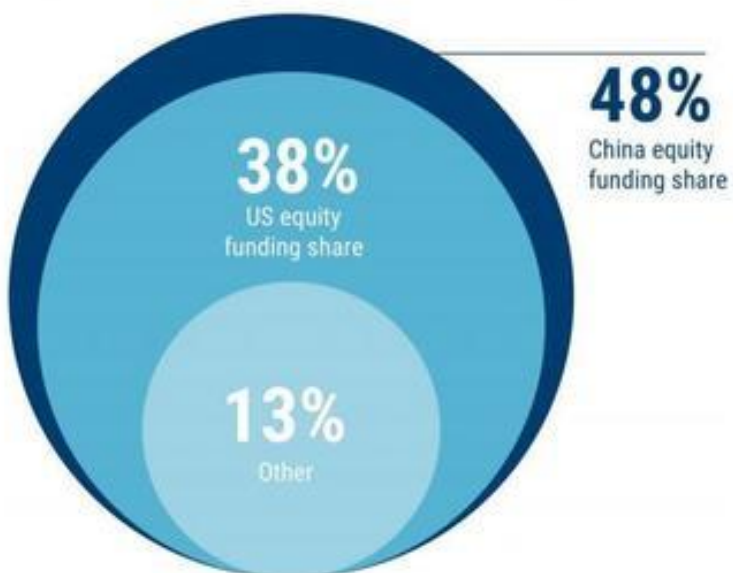


# Investissements en IA

- Feuille de route en IA de la Chine :
- **150 milliards de dollars** pour son économie d'ici **2030**
- Le MIT / USA a investi 1 Md\$

## China dominates global AI funding

US vs. China total equity funding to startups in 2017

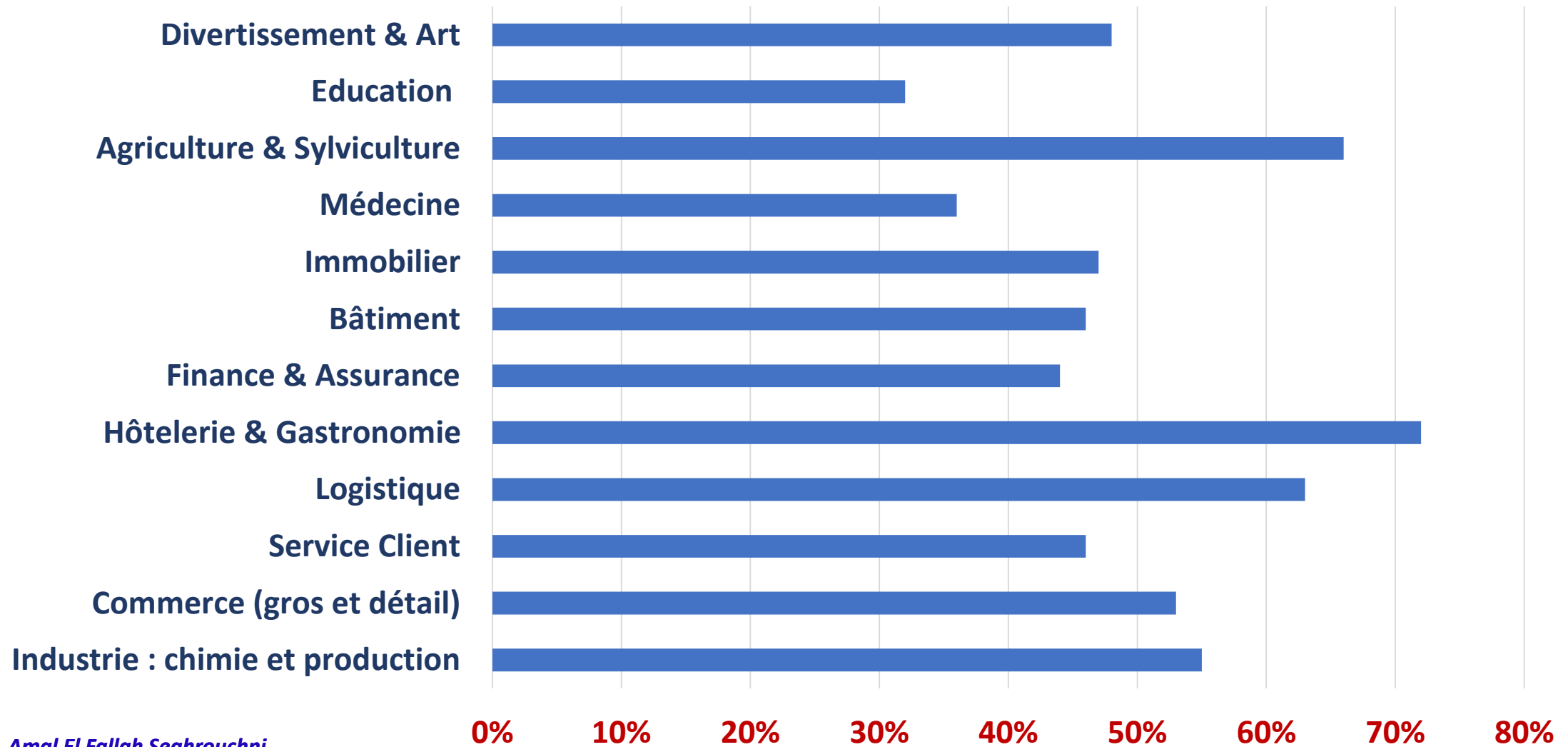


# Création d'emplois







- L'IA **crée de nouveaux emplois** et améliorera notre niveau de vie global
- **De nouveaux métiers seront également créés**
- De **nombreuses opportunités** économiques se créent, ouvrant de nouvelles perspectives d'emplois et d'amélioration du niveau de vie global.
- Les machines dotées d'IA pourront certes remplacer les hommes dans certains métiers, mais cela se fera sur **des tâches pénibles ou répétitives, dangereuses** ou sans réelle valeur ajoutée.



# Automatisation des tâches par secteur (Horizon 2055)



# Enjeux sociétaux

-  Comme à chaque révolution majeure, un temps de transformation est nécessaire afin de mettre en adéquation le marché du travail et les nouveaux besoins.
-  Augmenter le niveau de qualification humaine
-  Accompagner les populations dans ces transformations et réfléchir collectivement aux nouveaux modes de vies : un des enjeux majeurs de notre siècle
-  Des changements organisationnels **majeurs** à intégrer dès aujourd'hui !!
-  Transparence des algorithmes et explicabilité
-  Définition d'un cadre juridique

=> vers une IA organisationnelle





# Plan


- Introduction à l'Intelligence Artificielle (IA)
- Une IA peut en cacher une autre
- Grands domaines de l'IA
- Quelques projections
- Quid de l'éthique de l'IA ?**
- Conclusion



# Consensus : besoin d'un cadre éthique de l'IA

S'en remettre à des machines et à des systèmes d'aide à la décision peut poser de nombreux problèmes éthiques



 Même si les programmes reposent sur une logique «froide», ils ne sont pas exempts de préjugés et de **biais**

- Des algorithmes qui implémentent une cognition humaine avec ses biais
- Des données qui souffrent de biais de sélection, d'endogénéité, des variables omises,..
- Du déploiement qui peut cacher des biais économiques et d'autres manipulations
- Instrumentalisation consciente de l'IA

 Algorithmes de ML opaques (qualifiés de **boîtes noires**),

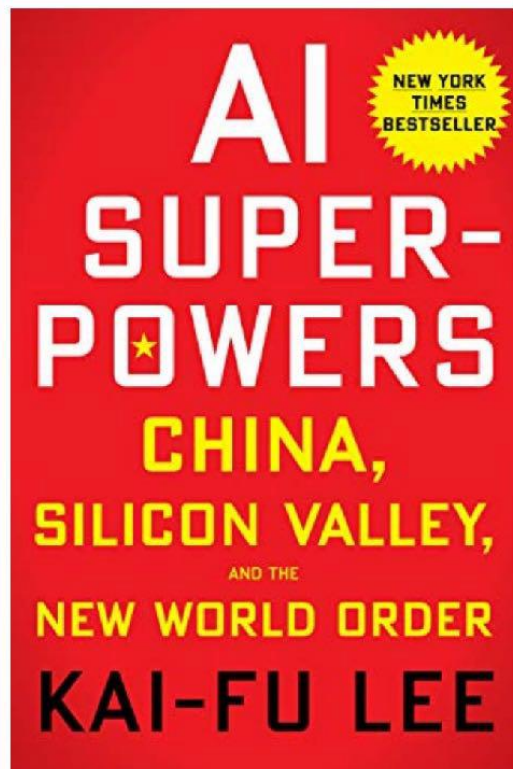
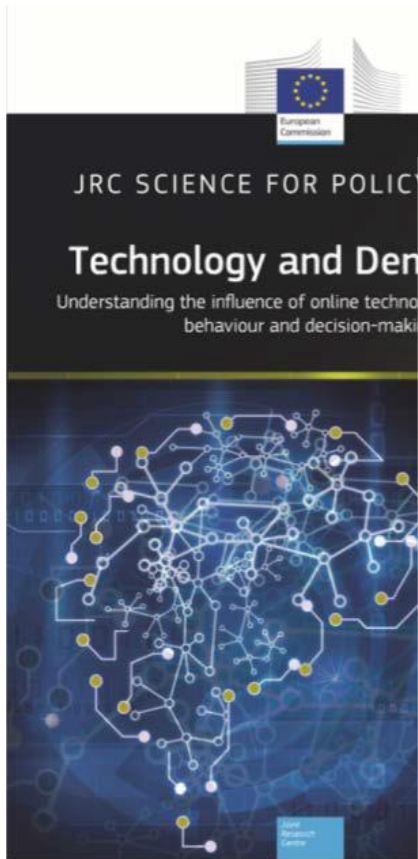
- Inexplicables, difficilement interprétables ou certifiables

 Question **d'alignement de valeurs** entre Humains et Machines

 L'IA est **énergivore** (GPT3 – 175 milliards de paramètres – entraînement sur des pages web)



# Littérature foisonnante



## INVESTIGATION OF COMPETITION IN MARKETS

OFF REPORT AND  
FOUNDATIONS  
ON ANTITRUST  
ADMINISTRATIVE  
ON THE JUDICIAL

Committee on the Judiciary  
Chairman, Subcommittee on  
Administrative Law and  
Procedure



STATES  
20

## THE AGE OF SURVEILLANCE CAPITALISM

THE FIGHT FOR A  
HUMAN FUTURE  
AT THE NEW  
FRONTIER OF POWER

SHOSHANA  
ZUBOFF



OFFICIAL SELECTION 2020  
sundance  
Festival  
The Social Dilemma



# De nombreux projets de réglementation







## Organisations intergouvernementales

- 🏛️ Conseil de l'Europe, et travaux de son Comité *ad hoc* sur l'intelligence artificielle (CAHAI)
- 🏛️ Union européenne et travaux traitant de l'IA
- 🏛️ Groupe d'experts de haut niveau de la Commission européenne sur l'IA - IA digne de confiance ;
- 🏛️ Premier Groupe d'experts de l'Organisation de coopération et de développement économiques (OCDE) (AIGO) et de son successeur, le réseau d'experts de l'OCDE sur l'IA (ONE AI),
- 🏛️ Recommandation du Conseil de l'OCDE sur l'IA
- 🏛️ Observatoire des politiques de l'IA de l'OCDE (OECD.AI) ;
- 🏛️ Principes sur l'IA du G20, inspirés des travaux de l'OCDE et présentés dans la Déclaration ministérielle du G20 sur le commerce et l'économie numérique ;
- 🏛️ Vision commune de Charlevoix sur l'avenir de l'IA adoptée par le G7 ;
- 🏛️ Travaux du Groupe de travail de l'Union africaine sur l'IA ;
- 🏛️ Groupe de travail de la Ligue des États arabes sur l'IA, ...
- 🏛️ la recommandation de l'UNESCO en 2 temps
- 🏛️ ...



# De nombreux projets de réglementation

## Secteur privé, organisations professionnelles, ONG ..






-  L'Initiative mondiale de l'Institut des ingénieurs électriciens et électroniciens (IEEE) sur l'éthique des systèmes autonomes et intelligents et ses travaux sur une conception conforme à l'éthique ;
-  Le livre blanc du Forum économique mondial sur une gouvernance multipartite mondiale des technologies ;
-  les « 10 grands principes pour une intelligence artificielle éthique » de l'UNI Global Union ;
-  la Déclaration de Montréal pour un développement responsable de l'IA ;
-  la Déclaration de Toronto – Protéger le droit à l'égalité et à la non-discrimination dans les systèmes reposant sur l'apprentissage automatique ; les principes pour une intelligence artificielle harmonieuse ; et les principes du Partenariat sur l'IA..
-  ...

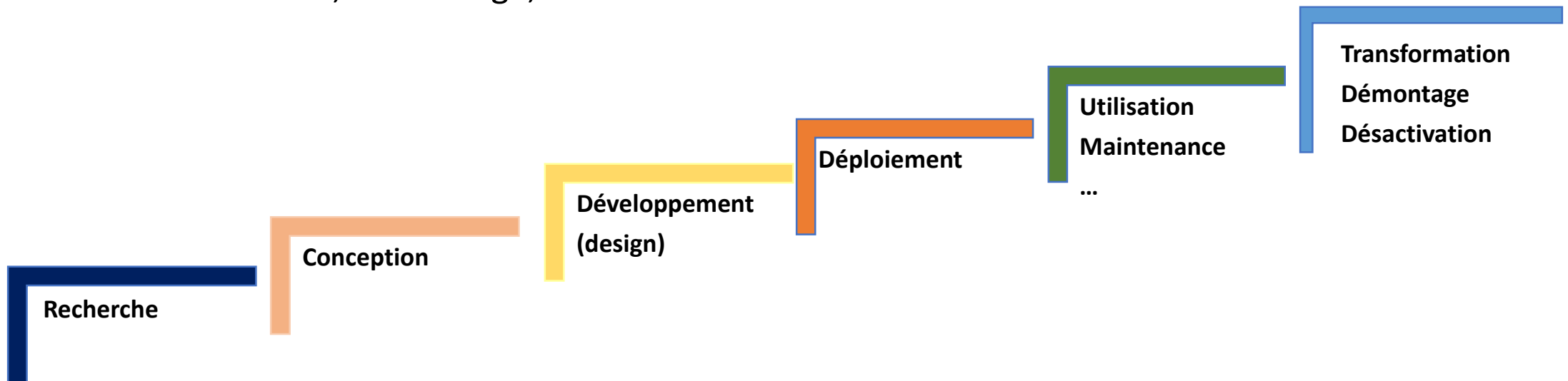


# Questions éthiques









## Se posent tout au long cycle de vie d'un système d'IA

-  Collecte, utilisation et génération des données
-  Conception d'algorithmes éthiques et responsables
-  Utilisation incluant la maintenance, l'exploitation, la commercialisation, le financement, le suivi et l'évaluation, la validation, la fin de l'utilisation..
-  Déploiement éthique inclusif et non discriminant
-  Transformation, démontage, désactivation ...

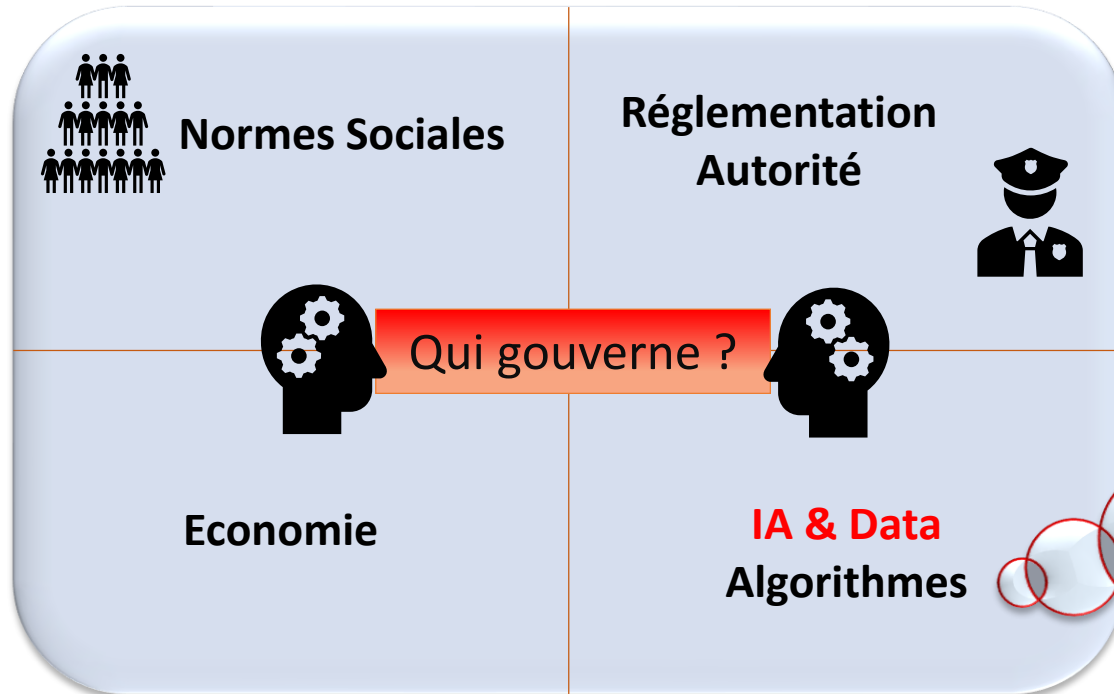


# Impact de l'IA sur les droits de l'homme et sa dignité

-  Subordination algorithmique
-  Manipulation et *nudges*
-  Vie privée et protection des données
-  Biais et discriminations
-  Droits cognitifs
-  Alignement(s) de valeurs

# 1. Subordination algorithmique

On passe d'un système qui décide tout seul à un système qui décide tout seul **et à votre place**



- Parcours-sup affecte les bacheliers
- Un drone décide de lancer un missile
- Une voiture autonome fonce sur un piéton
- Un système de recommandation octroie les crédits
- Le prix du billet d'avion s'auto-modifie en ligne,
- La voiture qui refuse de démarrer tant que la ceinture est détachée

**Subordination algorithmique**  
Partage d'autorité en faveur de la machine  
( cf. LOA versus autonomie en aérospatial)

Moins de **0,05%** de la population mondiale peut coder ou construire des algorithmes.  
A qui confions-nous cette gouvernance ?





## 2. Manipulation et nudges 1/2



Concept des **sciences du comportement** issu des pratiques de design industriel (architecture du choix)

- des **suggestions indirectes** peuvent, **sans forcer, influencer les motivations, les incitations** et la prise de décision des groupes et des individus, au moins de manière aussi efficace sinon plus efficacement que l'instruction directe, la législation ou l'exécution.



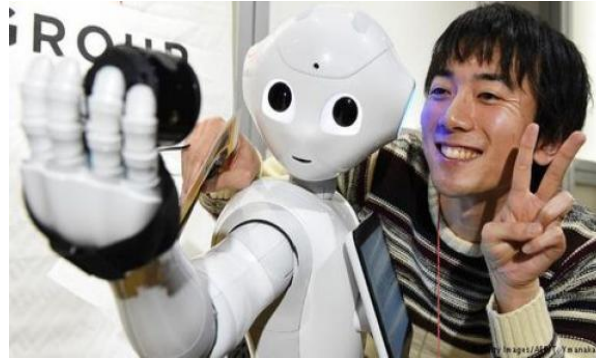
Des hommes politiques (comme David Cameron et Barack Obama) ont eu recours aux nudges pour pousser les citoyens à prendre les « bonnes décisions » ..

- pour lutter contre l'obésité, favoriser le recyclage, réguler la consommation d'énergie, etc.

Théorie des nudges introduite en 2008, par [Richard Thaler](#) et [Cass Sunstein](#) dans leur livre [Nudge : Améliorer les décisions concernant la santé, la richesse et le bonheur](#).

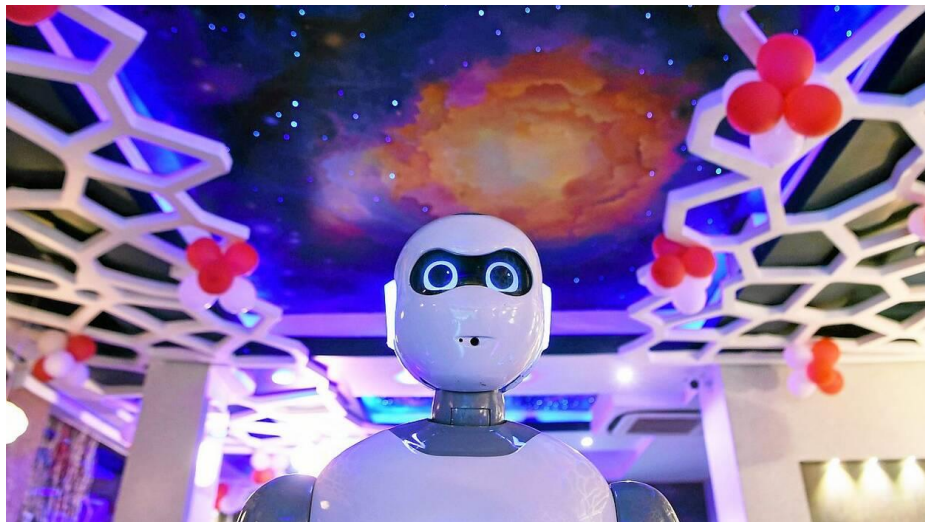


## 2. Manipulations et nudges 2/2



les robots empathiques  
qui ressemblent aux enfants  
sont irrésistibles

Is empathy with robots a danger  
to human communication?



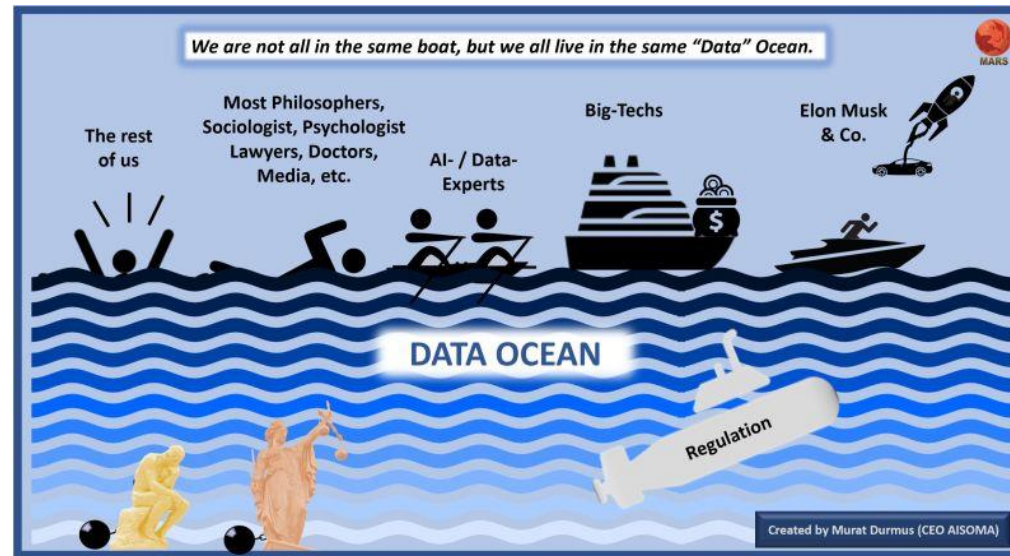
Quelle est la place du robot  
dans un foyer ?

The German experience with the robot that or who (?) said: "No, No, no, no, no, Please don't go away! I'm afraid it's never going to be daylight again "

After interaction with the robot, **13 out of 43** participants in the experiment were sensitive to the robot's request and could not turn off the robot !

# 3. Questions de vie privée et protection des données 1/2

Plus les algorithmes ont de données, plus ils apprennent, mieux ils manipulent



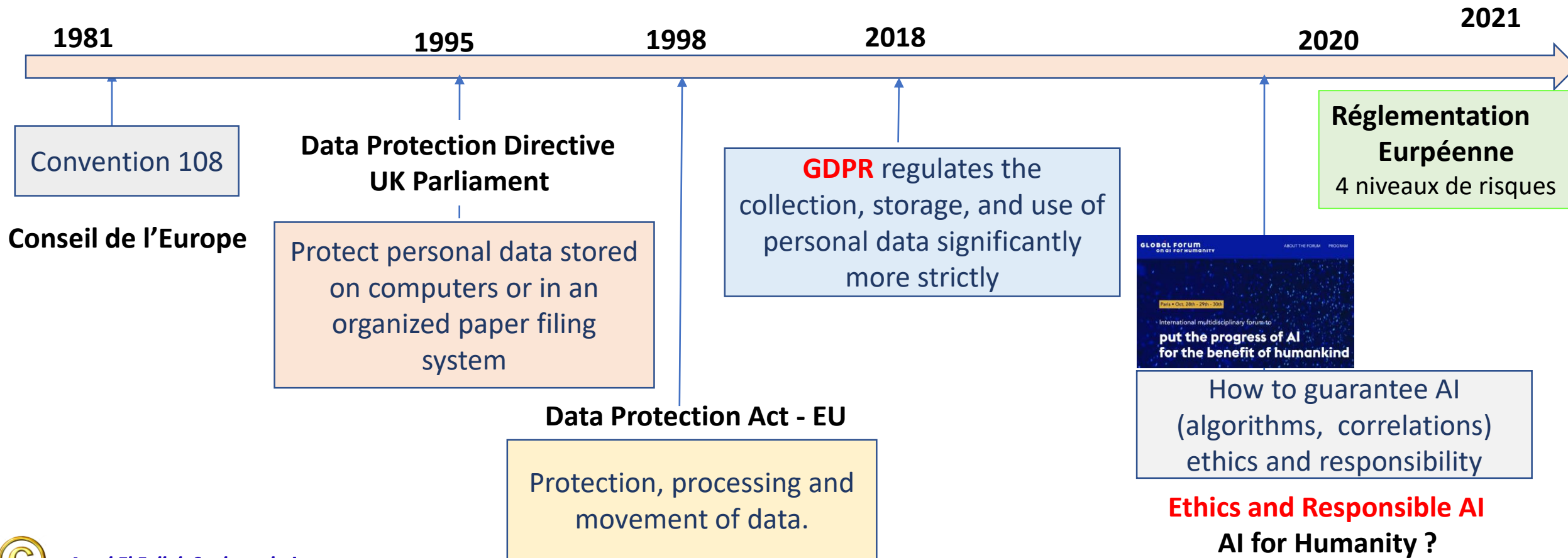
**Profiling, Cambridge Analytica**



# 3. Vie privée et protection des données 2/2

Vie privée : le **droit d'une personne** de garder secrètes **ses affaires personnelles et ses relations**

Travaux fondateurs d'Alan Westin Privacy and Freedom (1967)

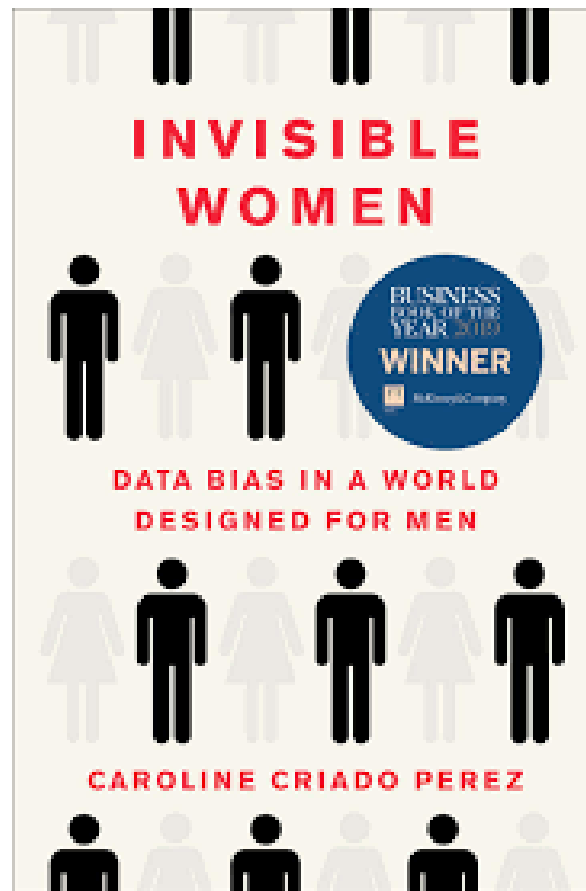
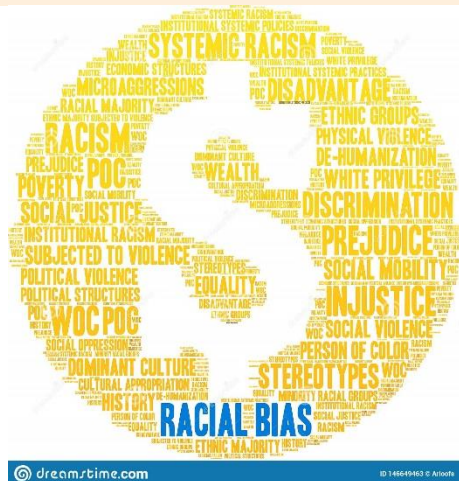
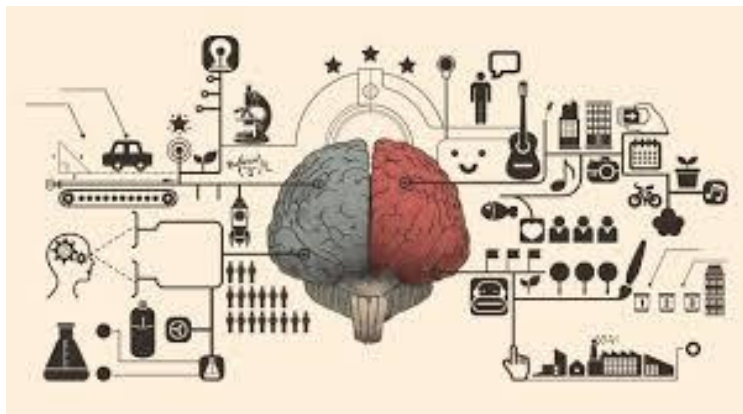
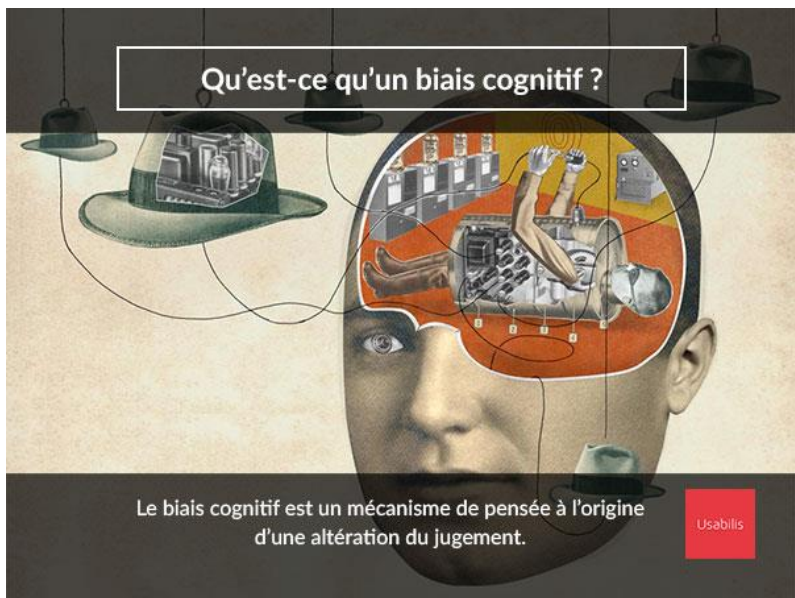


# 4. Biais et discriminations

 Cognitifs

 Economiques

 Statistiques

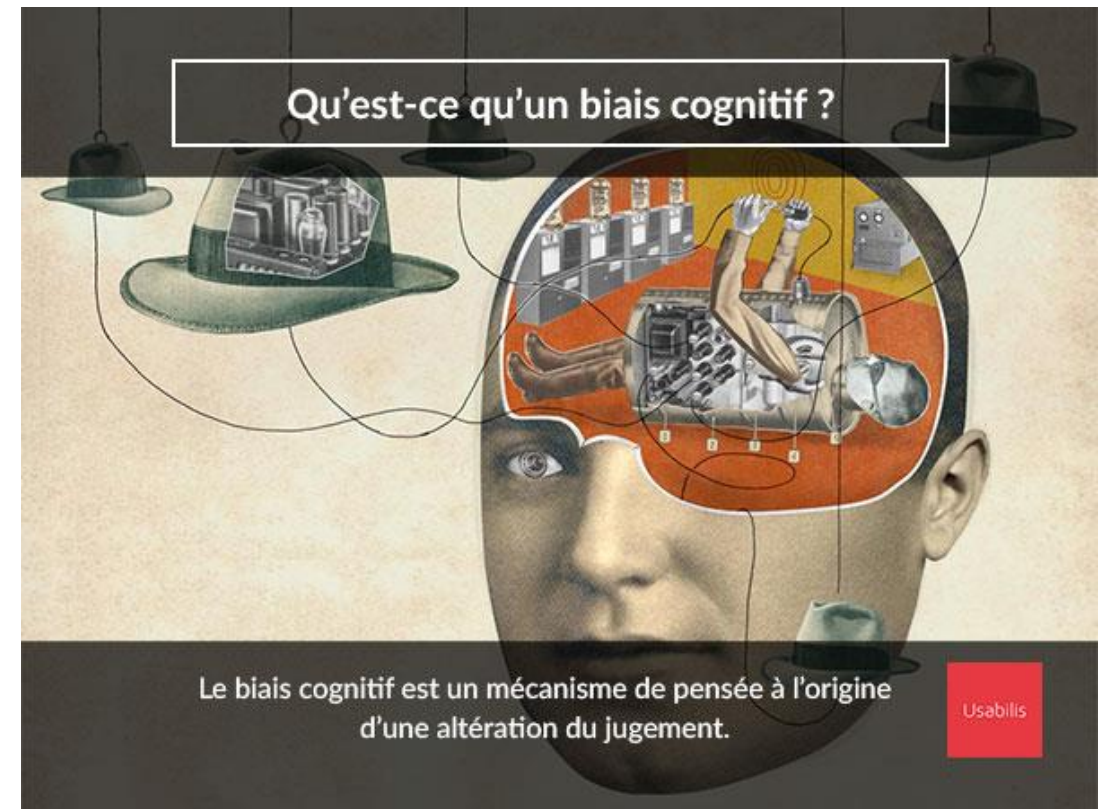


# 4.1 Biais cognitifs

Un algorithme n'est en réalité qu'une «opinion intégrée aux programmes»

Un biais cognitif est un mécanisme de pensée à l'origine d'une altération du jugement

**Exemple :** le mot « femme » étant systématiquement associé à des termes tels que « femme de ménage », « nourrice », « réceptionniste », « styliste », ou bien « coiffeuse », tandis que le mot « homme » est associé à « chef », « philosophe », « capitaine », « financier ».



## 4.2 Biais économiques

- Volontairement ou pas, les algorithmes peuvent être biaisés
- Algorithmes ciblant des segments de la population.
  - ex. des offres d'emplois ciblant Hommes versus Femmes.
- Search engine manipulation* : manipulation par les entreprises.
  - Dans le contexte du e-commerce, la commission européenne a par exemple condamné Google en 2017 à une amende de **2,4 milliards d'euros** pour avoir **favorisé ses propres produits** dans les résultats de recherche de **Google Shopping** au détriment de ses concurrents (IP/17/1784 ; affaire 39740).



# 4.3 Biais statistiques ou de données



## Le biais d'endogénéité



## La variable omise

- Omettre une caractéristique sensible (ex. origine ethnique) peut renforcer la discrimination
- Rendre aveugle les algorithmes peut porter atteinte à l'équité
- Dilemme important pour le RGPD
  - lutter contre la discrimination femmes si on ne collecte pas des données sur le genre ?



## Le biais de sélection

- Quand les caractéristiques de la population étudiée diffèrent de celles de la population générale.
- James Heckman, dans ses travaux lui ayant valu le « prix Nobel d'économie », a montré que le biais de sélection était de la même nature que le biais de variable omise.
  - calcul d'un score de crédit (credit scoring)



## Garbage In Garbage Out

- Sélection Talents AMAZON (entrées déséquilibrées homme et femme)





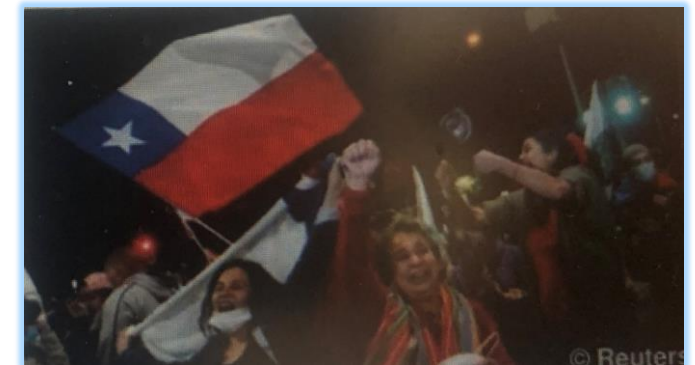
# 5. Droits cognitifs

 Cognitive rights ( Droits Cognitifs) au Chili (Octobre 2020) – **1<sup>er</sup> Pays au monde !!**

- Le congrès Chilien a présenté un amendement à la constitution qui définit **l'identité mentale** comme **un droit qui doit être protégé** des manipulations.

 De nombreuses initiatives pour l'éthique de l'IA sont en cours :

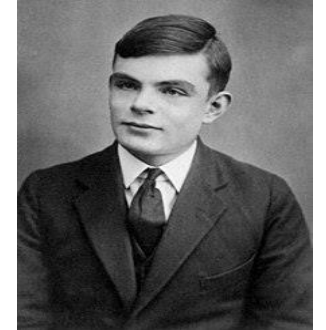
- L'UNESCO a lancé une consultation mondiale sur son projet d'éthique (été 2020)
- L'OCDE a constitué un groupe d'experts pour élaborer un projet d'éthique (été 2020)
- Le Parlement Européen a adopté 3 nouvelles résolutions en lien avec l'IA (le 20 octobre 2020 )
  - **Le régime de responsabilité de l'IA**
  - **L'IA et la propriété intellectuelle**
  - **L'encadrement éthique des développements IA et** et le droit à l'information
- De nombreux centres de recherche en IA penchent sur le sujet
  - Montréal, Melbourne, ...



# 6. Alignement des valeurs

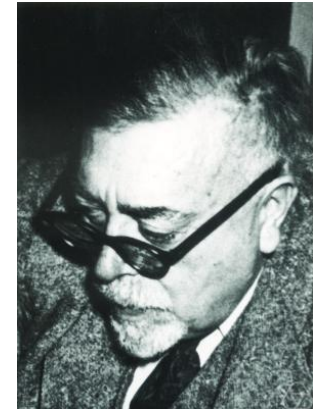
## Alan Turing, un des fondateurs de l'IA (1951)

- Même si nous pouvions maintenir les machines en position d'asservissement, par exemple en coupant le courant à des moments stratégiques, nous devrions, en tant qu'espèce, nous sentir profondément humiliés.



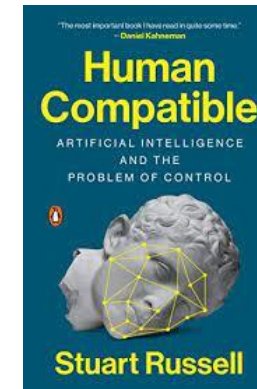
## Norbert Wiener, fondateur de la cybernétique (1942)

- *[Cybernetics or Control and Communication in the Animal and the Machine]*
- Il vaut mieux être bien sûr que le but mis dans la machine est le but que nous désirons vraiment



## Stuart J. Russell, professeur d'IA à Bekrkeley (2019)

- Stuart J. Russell et Peter Norvig,
- Artificial Intelligence : A Modern Approach, 2010, 3e éd.



# Problème d'alignement



Comment être sûr que la machine interprétera correctement ce qu'on veut ?



Exemple inspiré de Stuart J. Russell

- Vous dites au robot de “faire diner votre enfant si vous rentrez tard”
- Le robot va exécuter la tâche
- Mais la femme de ménage a oublié de cuisiner
- Votre enfant va exiger son diner
- Le robot obéit et tente d'improviser un dîner !



# Plan

 Introduction à l'Intelligence Artificielle (IA)

 Une IA peut en cacher une autre

 Grands domaines de l'IA

 Quelques projections

 Quid de l'éthique de l'IA ?

 Conclusion



# Conclusion

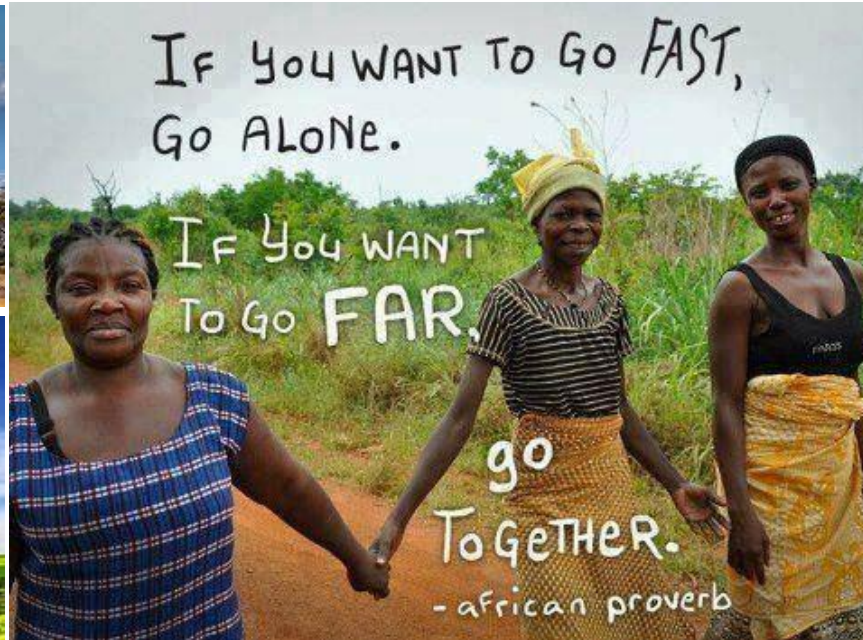
- Protéger l'IA de l'humain pour protéger l'humain de l'IA
- Développer une IA éthique et responsable alignée avec les valeurs universelles
- Cesser la fascination pour le post-mortem et ces projets de digital twin des défunts, des chatbots qui parlent comme les défunts (brevet de MS), etc.
- Même si l'IA peut produire de l'or, il faut se méfier du **“tout or”**

Le Roi Midas *au toucher d'or* a obtenu exactement ce qu'il avait demandé à Dionysos :  
*sa nourriture, sa boisson et sa famille*  
*se sont changées en or*  
**Roi Midas, 676 av. J. –C.**





**Let's move together with ethical and responsible AI**



- 🏠 Un [projet de réglementation](#) de l'intelligence artificielle (IA) a été présenté par les commissaires européens à Bruxelles le 21 avril 2021.
- 🏠 L'objectif avancé est de « *faire de l'Europe le pôle mondial d'une intelligence artificielle (IA) digne de confiance* » ; un [travail de réflexion](#) entamé depuis quelques années.
- 🏠 Ce projet combine un « [premier cadre juridique sur l'IA](#) » et un « [nouveau plan coordonné avec les États membres](#) ». Dans son communiqué de presse, la Commission européenne annonce vouloir « *garantir la sécurité et les droits fondamentaux des citoyens et des entreprises, tout en renforçant l'adoption de l'IA, les investissements et l'innovation dans l'ensemble de l'UE* ».
- 🏠 Ce règlement se fonde principalement sur les risques, créant les catégories de risques inacceptables, élevés, limités et minimes, et l'explique par quelques exemples.
- 🏠 Les IA classées dans « risque inacceptable » seront interdites. Elles comprennent « *les systèmes d'IA considérés comme une menace évidente pour la sécurité, les moyens de subsistance et les droits des personnes* ». Ce sont, par exemple, les systèmes qui permettent la notation sociale par les États ou ceux qui manipulent le comportement humain pour priver les utilisateurs de leur libre arbitre.
- 🏠 Les IA classées dans « risque élevé » ou « à haut risque » devront être conformes à des obligations strictes pour pouvoir être mises sur le marché. Ce sont par exemple celles qui sont utilisées dans les transports et susceptibles de mettre en danger la vie et la santé des citoyens. Ou dans l'éducation et l'emploi (la notation d'examens, le tri des CV...). Ou encore celles utilisées dans le domaine du maintien de l'ordre, de la chirurgie assistée par robot ou l'évaluation du risque de crédit, qui prive certains citoyens de la possibilité d'obtenir un prêt. Les systèmes d'identification biométrique à distance fondés sur l'IA sont particulièrement considérés à haut risque. Leur utilisation en temps réel dans l'espace public aux fins du maintien de l'ordre est en principe interdite. Il existe des exceptions (pour rechercher un enfant disparu, ou un suspect d'une infraction pénale grave).
- 🏠 Les IA à « risque limité » seront encadrées par des obligations spécifiques en matière de transparence. Ce sont par exemple les « chatbots », les utilisateurs doivent savoir qu'ils interagissent avec une machine afin de pouvoir décider en connaissance de cause de poursuivre ou non. [Autre exemple](#), le “deep fake”, la reproduction de la voix ou l'image d'un individu pour lui faire dire ou faire ce qu'il veut, sera encadré.
- 🏠 Les IA à « risque minime » seront autorisées librement (jeux vidéo, filtres anti-spam...)
- 🏠 Ce règlement européen prévoit aussi des amendes et des sanctions pouvant aller jusqu'à 6% du chiffre d'affaires mondial.
- 🏠 Sur la reconnaissance faciale, cette [disposition est jugée insuffisante](#) par des organisations de protection des consommateurs et plusieurs eurodéputés, ce qui promet un débat intense au Parlement européen.