

GROUPE PHILO MARSEILLE

# Philosophie et Intelligence Artificielle



## L' Être Humain et le Robot

Robert MACULET  
[robert.maculet@gmail.com](mailto:robert.maculet@gmail.com)  
[maculet.org](http://maculet.org)  
[philosophieTV.com](http://philosophieTV.com)

Ingénieur École Centrale Paris

Docteur en Informatique / Intelligence Artificielle (Université Paris 6)

# PLAN

**Introduction**

**Partie 1 : Qu'est-ce que l' IA ? (IA Faible)**

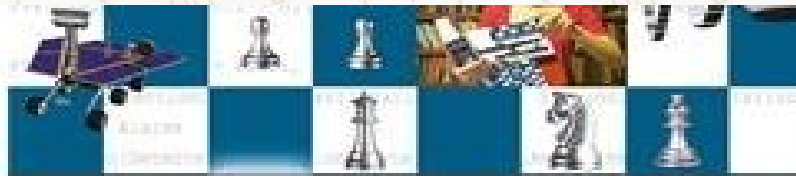
**Partie 2 : IA Forte**

**Partie 3 : L' Éthique de l' IA**

**Conclusion**



**INTRODUCTION**

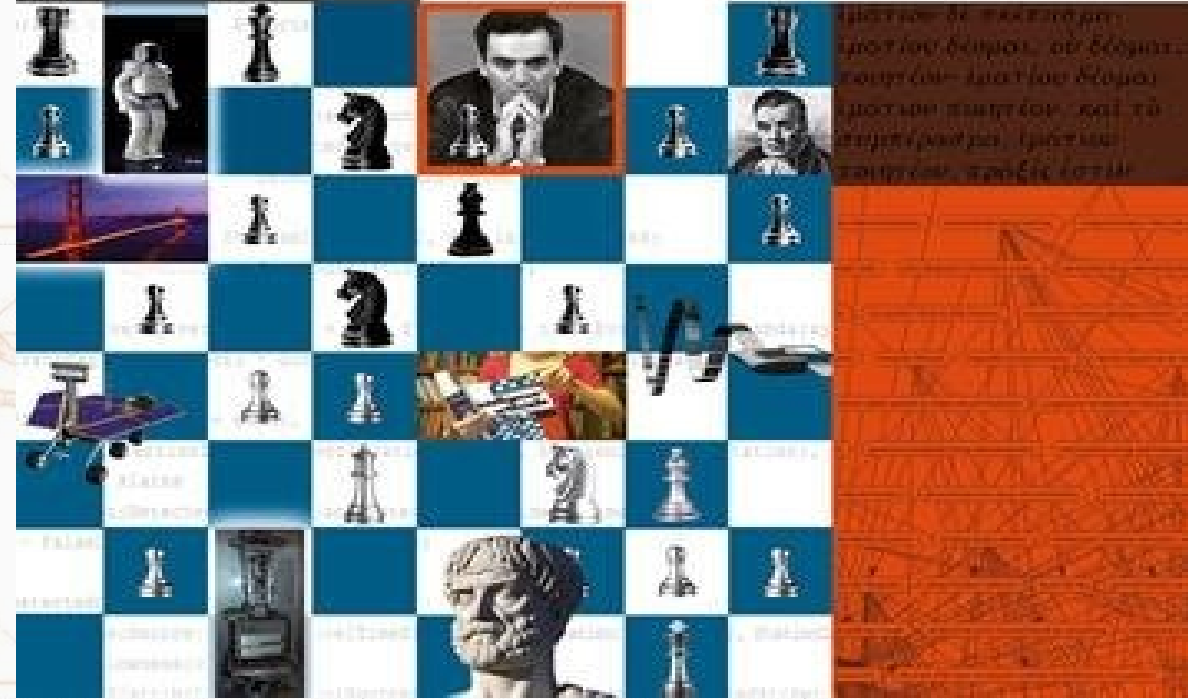


Stuart Russell et Peter Norvig

# Intelligence artificielle

3<sup>e</sup> édition

Avec plus de 500 exercices



ἡμεῖς οὐκ ἐκείνην ἀντι-  
δοξίαν δεύρομεν, ἀλλ' ἐξομῶμεν  
καὶ τὴν ἀντι-δοξίαν δεύρομεν  
δοξίαν αὐτῆς, καὶ τὰ  
ἐπιτελεῖται ἡμεῖς  
καὶ οὐκ ἐκείνη



**Intelligence Artificielle, Russell & Norvig,  
2010, (1199 pages).**

## **Chapitre 26 Fondements philosophiques**

**26.1 IA faible: les machines peuvent-elles agir intelligemment?**

**26.2 IA forte: les machines peuvent -elles réellement penser?**

**26.3 Éthique et risques du développement de l'intelligence artificielle.**

**Voir ces textes sur <http://philosophietv.com/>**



# L'intelligence artificielle

## et le vivant



**Jean-Gabriel Ganascia**

*Équipe ACASA - LIP6 – CNRS (UMR 7606)*

*Sorbonne Université*

*Labex OBVIL*

*Institut Universitaire de France*

*Président du comité d'éthique du CNRS*

*4, place Jussieu, 75252 Paris Cedex 05, FRANCE*

*Jean-Gabriel.Ganascia@lip6.fr*





# Partie 1

# Qu'est-ce que l'IA ?

IA faible: les machines peuvent-elles  
agir intelligemment?



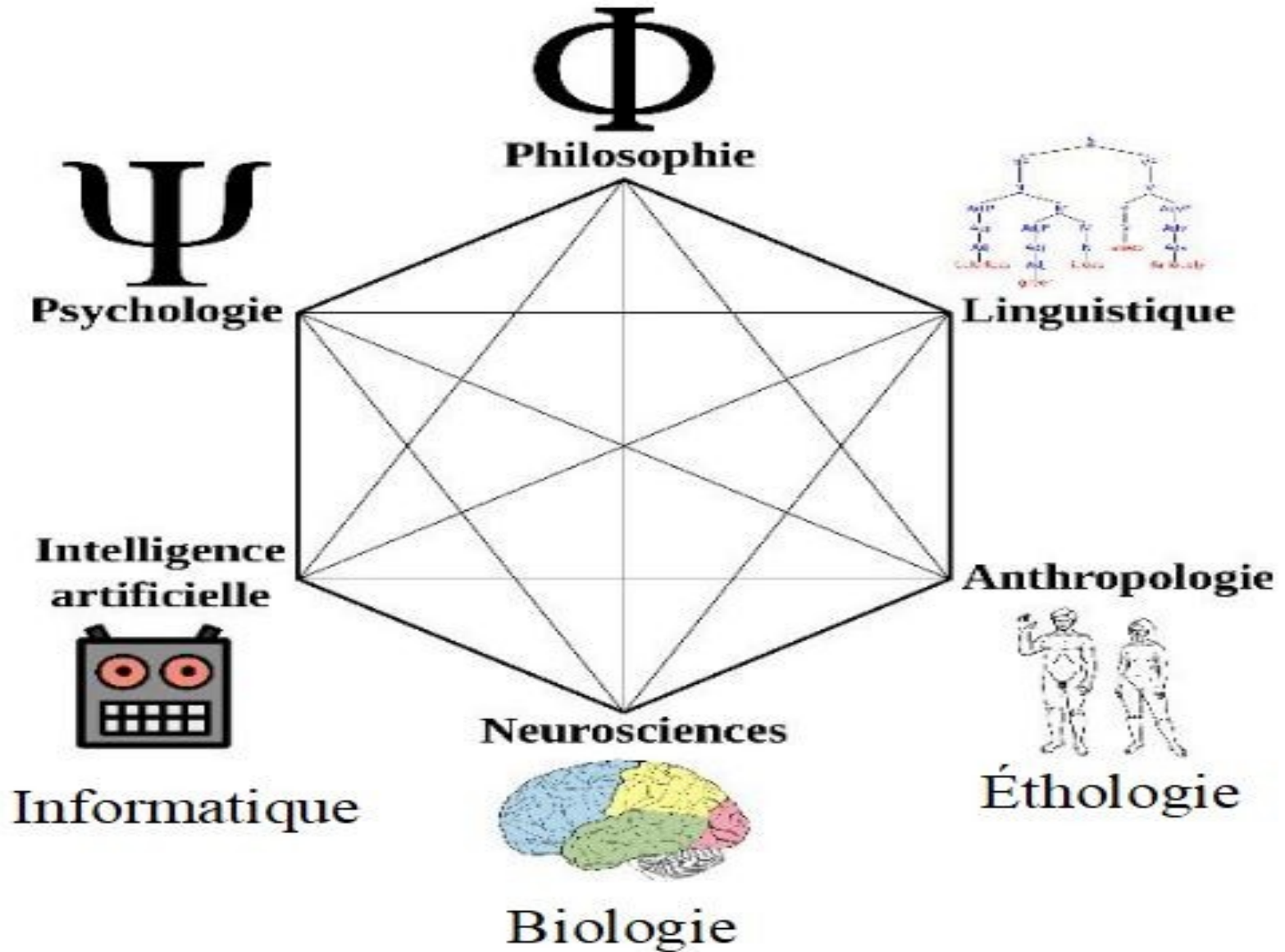
# La Cognition / La Connaissance

Du latin *Cognitio*, de *cognoscere*, « connaître »

- La cognition est l'ensemble des processus mentaux qui se rapportent à la fonction de **connaissance**.
- Ces processus cognitifs mettent en jeu la mémoire, le langage, le raisonnement, l'apprentissage, l'intelligence, la résolution de problème, la prise de décision, la perception, l'attention...
- La cognition incarnée



# Les Sciences Cognitives



# INTELLIGENCE ARTIFICIELLE

## 1) ARTIFICIEL versus NATUREL

Concept d' ARTEFACT: objet Artificiel conçu et fabriqué par l' Homme : Outil, Machine, Automate, Ordinateur, Robot, Drone, Voiture Autonome ,..

## 2) Concept d' INTELLIGENCE ?

# INTELLIGENCE

- Tests d' Intelligence : QI (Quotient Intellectuel)
- Intelligence émotionnelle (intelligence sociale)
- Intelligence collective (intelligence du vivant) ...

L' Intelligence est la capacité à s'adapter (Jean PIAGET)



The background of the slide is a repeating grid of Leonardo da Vinci's Vitruvian Man. Each figure is rendered in a light, golden-brown color and is centered within a square frame that is itself inscribed within a circle. The grid covers the entire page, with the text overlaid on the top-left portion.

# Les 9 Intelligences du chercheur en psychologie Howard GARDNER :

- Logico-mathématique
- Visuo-spatiale
- Verbo-linguistique
- Intrapersonnelle
- Interpersonnelle
- corporelle/kinesthésique
- Musicale
- Existentielle ou spirituelle





# Une définition de l' IA

« l'ensemble de théories et de techniques mises en œuvre en vue de réaliser des machines capables de **simuler** l'intelligence »

Copier la nature après de grandes simplifications



**Concept d' Intelligence**

**Concept de Pensée**

**Concept de Conscience**

**Concept de Conscience de Soi**

**(Exemple: l' homme a conscience qu'il va mourir)**

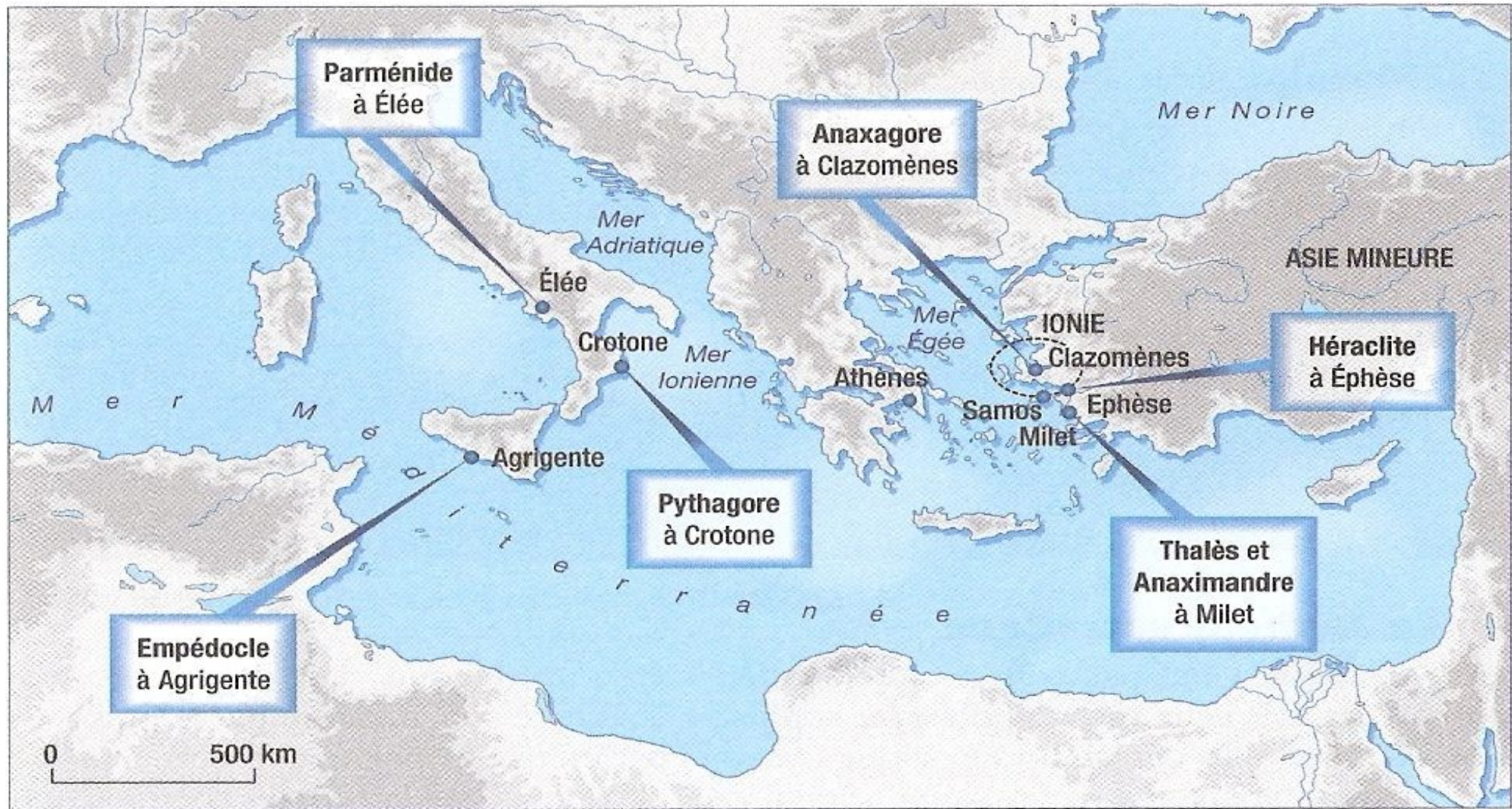


La cognition incarnée



# LE LOGOS

(la raison/la logique/le verbe)



Naissance de la philosophie (VI ème siècle av.JC)



# Représentation des Connaissances par la Logique en IA

**PARMENIDE (vers 540- vers 450 av. J.-C.):**

«L'Être est mais le Non-Être n'est pas»



- Principe du tiers exclu.
- Loi de la non-contradiction

# Philosophie de la connaissance

- Ontologie

- Épistémologie

# Le Concept du Concept (Métaconnaissance en IA)

- Le Concept est la représentation mentale générale et abstraite d'une chose. Schématisation. élimination des détails
- Le langage est l'outil principal de création de concepts
- En neurosciences, les objets mentaux (CHANGEUX 1983) se composent de percepts (perceptions des objets physiques), d'images mentales (visions intérieures des objets) et de concepts

# Des Concepts

REPRESENTATION DES CONNAISSANCES en IA

- Le Concept de Table

- Le Concept de chien

- Le concept d' Arbre

- ...

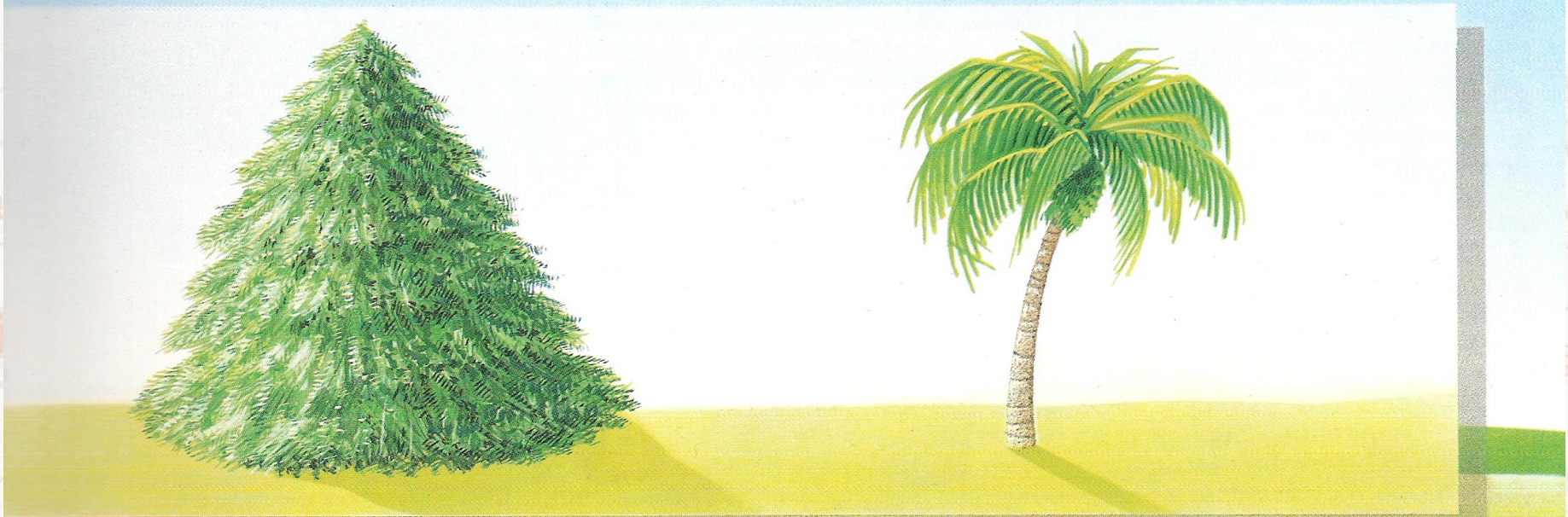




Un système qui stockerait l'image de gauche comme modèle d'arbre aurait les plus grandes difficultés à le reconnaître en vue aérienne (*ci-dessous*), partiellement caché (*ci-contre à droite*) ou modifié (le même arbre en hiver, à *l'extrême droite*). De plus, même si pins et palmiers sont des arbres (*en bas à droite*), leur silhouette, leur forme tridimensionnelle, leurs couleurs et leurs textures s'écartent considérablement de celles du prototype.



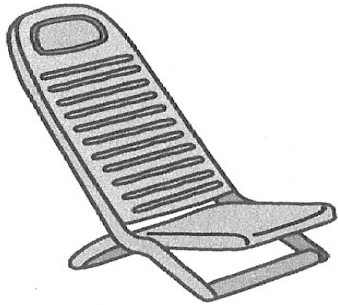




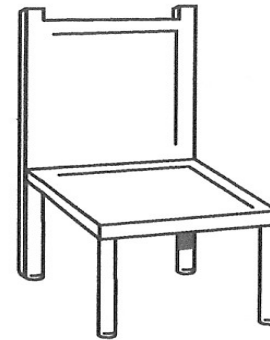


# Le Concept de Chaise

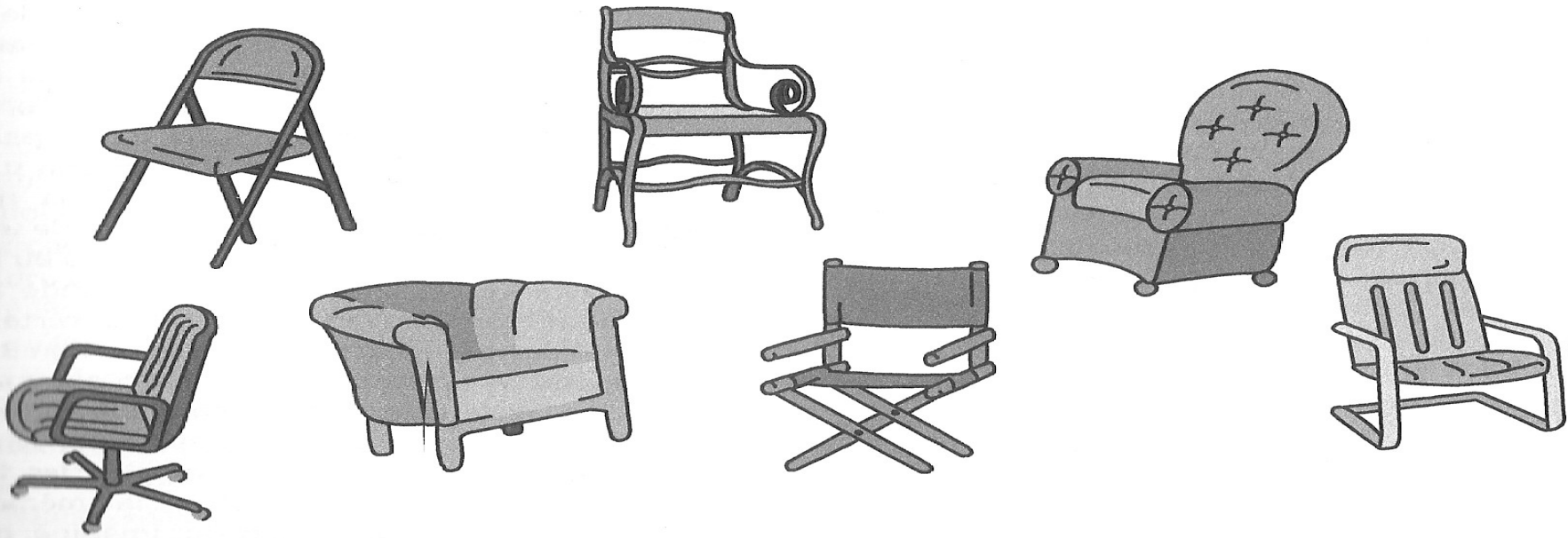
A.



B.



C.



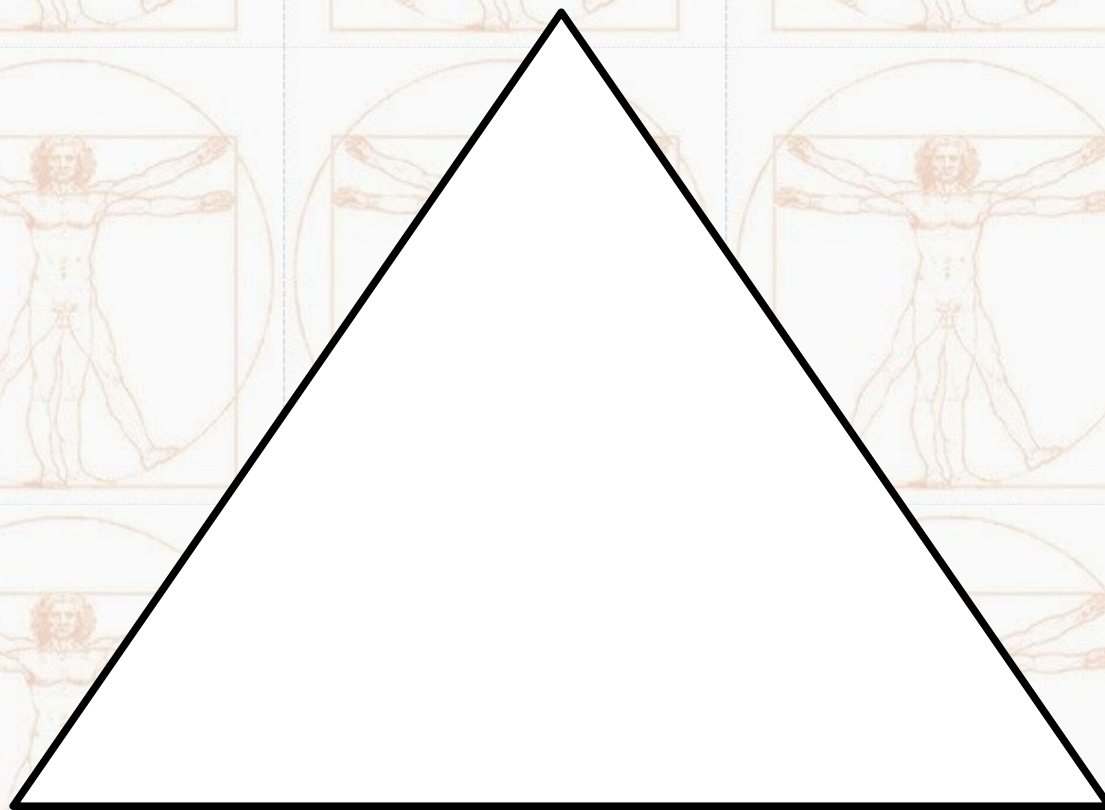


# Le Triangle du Sens

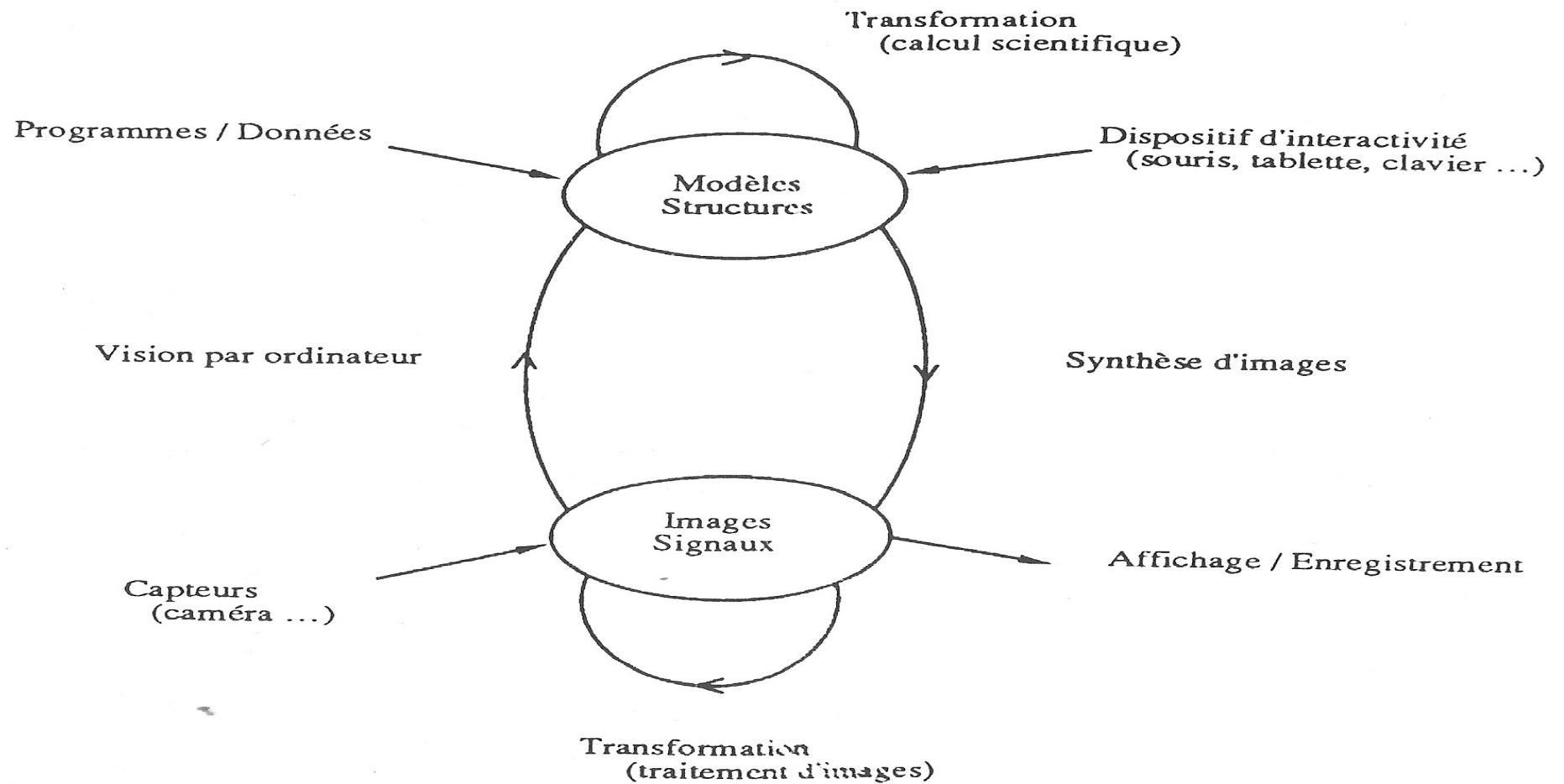
CONCEPT

MOT

CHOSE



# Le Modèle



# Le modèle d'un Agent Intelligent

38

Chapitre 2. Agents intelligents

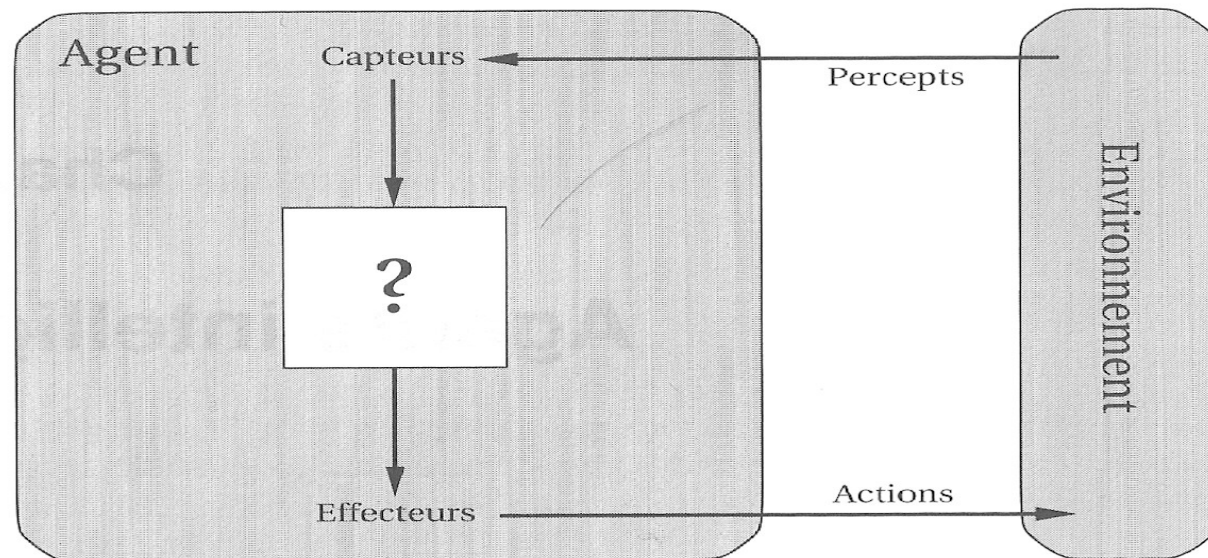
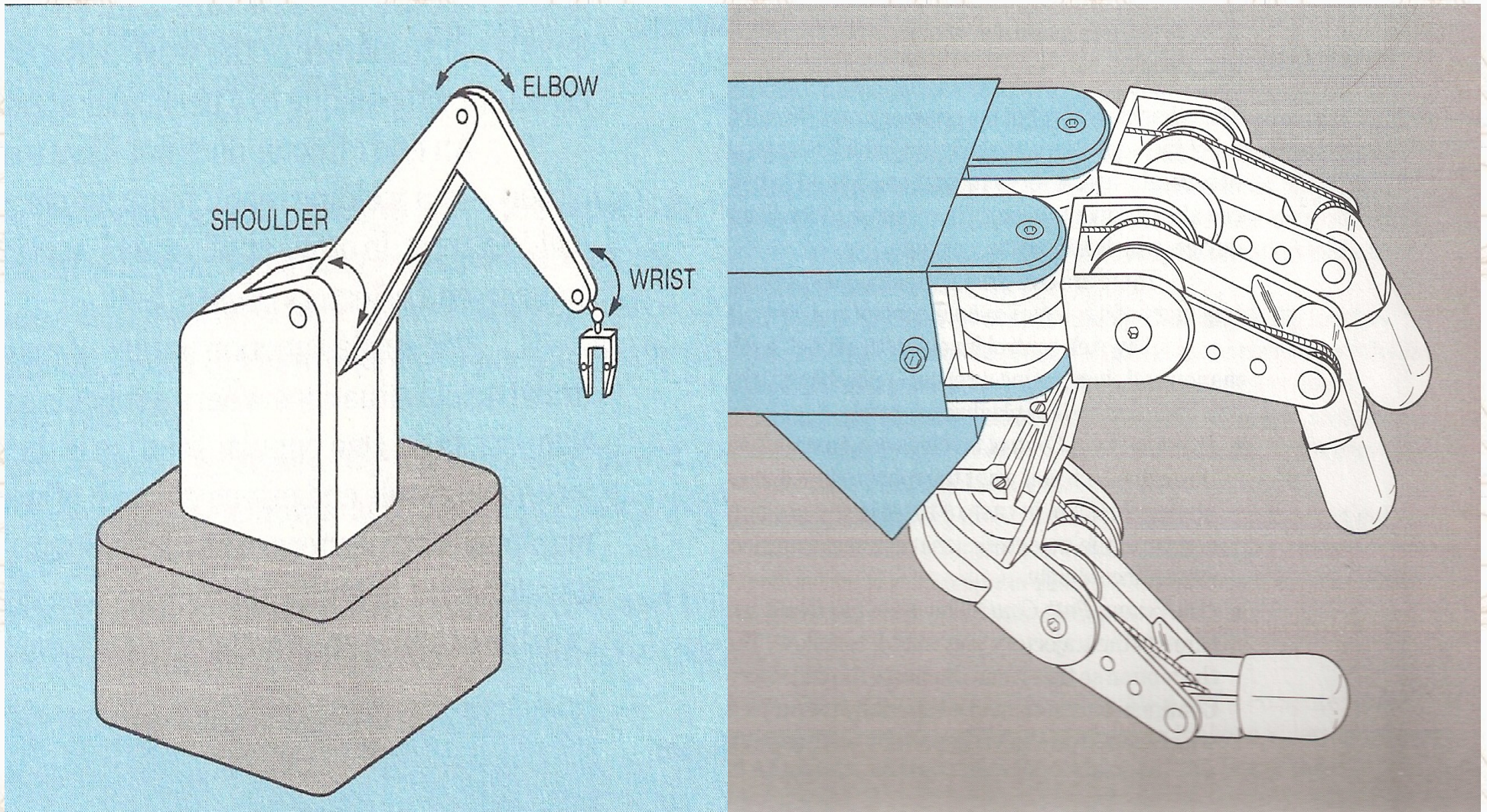


Figure 2.1 : Les agents interagissent avec leurs environnements au moyen de capteurs et d'effecteurs.

interagit avec son environnement par des capteurs et des effecteurs



# Les effecteurs





# PLATON (427-347 av.J.-C.)

Théorie des Idées (eidos en grec)

- Idée de Justice
- Idée de Cercle...

Termes voisins : Catégorie, Concept, Notion...

L'homme pense grâce aux Concepts. Le monde (le réel) devient ainsi intelligible.

# Synthèse et Analyse

## (Rassemblement et Division)

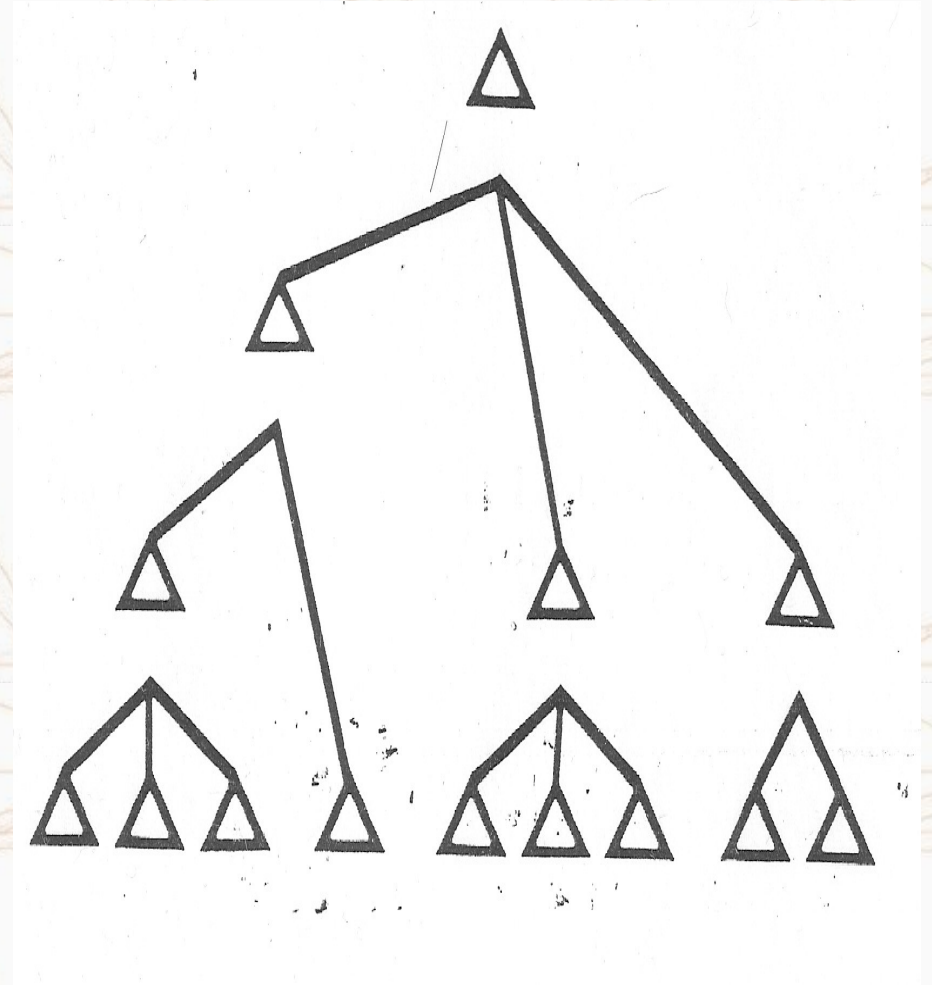
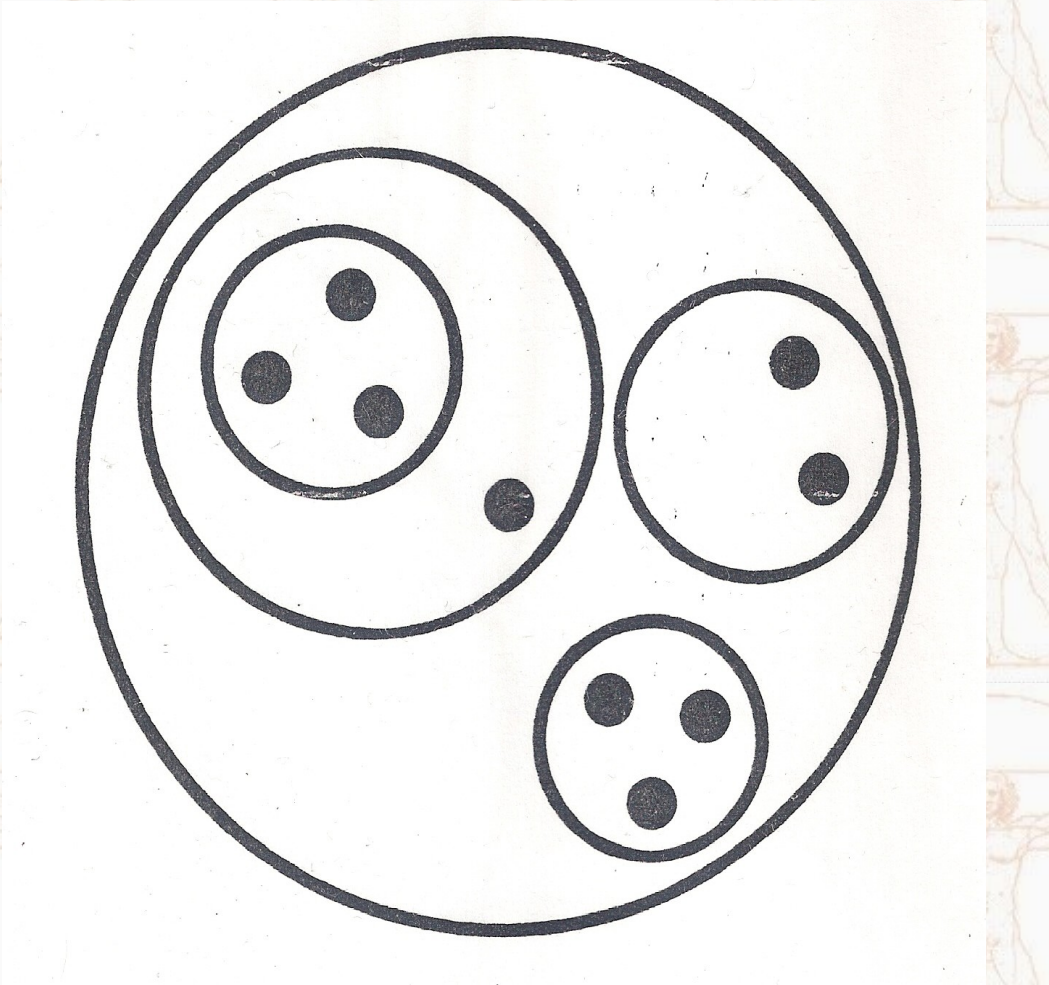
"Rassembler en un Concept unique un éparpillement de détails afin que chacun comprenne de la même manière de quoi il s'agit.

Séparer un Concept en divers éléments, mais au niveau des points d'articulation naturels, et non en coupant des membres en deux, comme le ferait un mauvais découpeur."

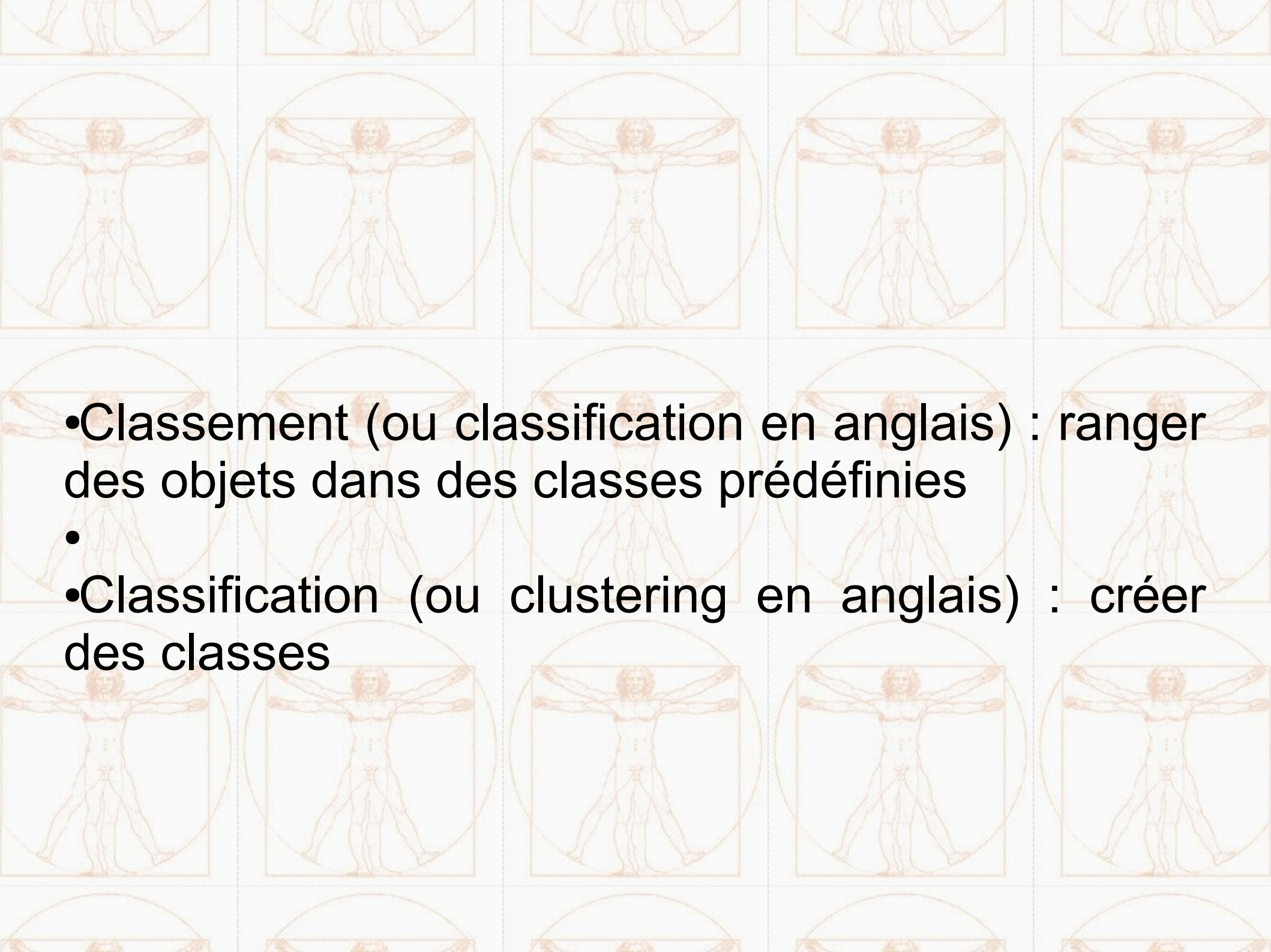
Propos de SOCRATE dans Phèdre (265 D) de Platon



# Classification et hiérarchisation en IA



Exemple de Classification: Espèce, Genre, individu

- 
- Classement (ou classification en anglais) : ranger des objets dans des classes prédéfinies
  - Classification (ou clustering en anglais) : créer des classes

# ARISTOTE (384-322 av.J.-C.)

## L' ONTOLOGIE en Philosophie

### La science de l'être en tant qu'être

• "Il y a une science qui étudie l'être en tant qu'être". Pour ce philosophe, ce concept s'oppose aux sciences particulières : "qui découpe quelque partie de l'être et en étudient les propriétés" (Métaphysique, Livre Gamma, 1).

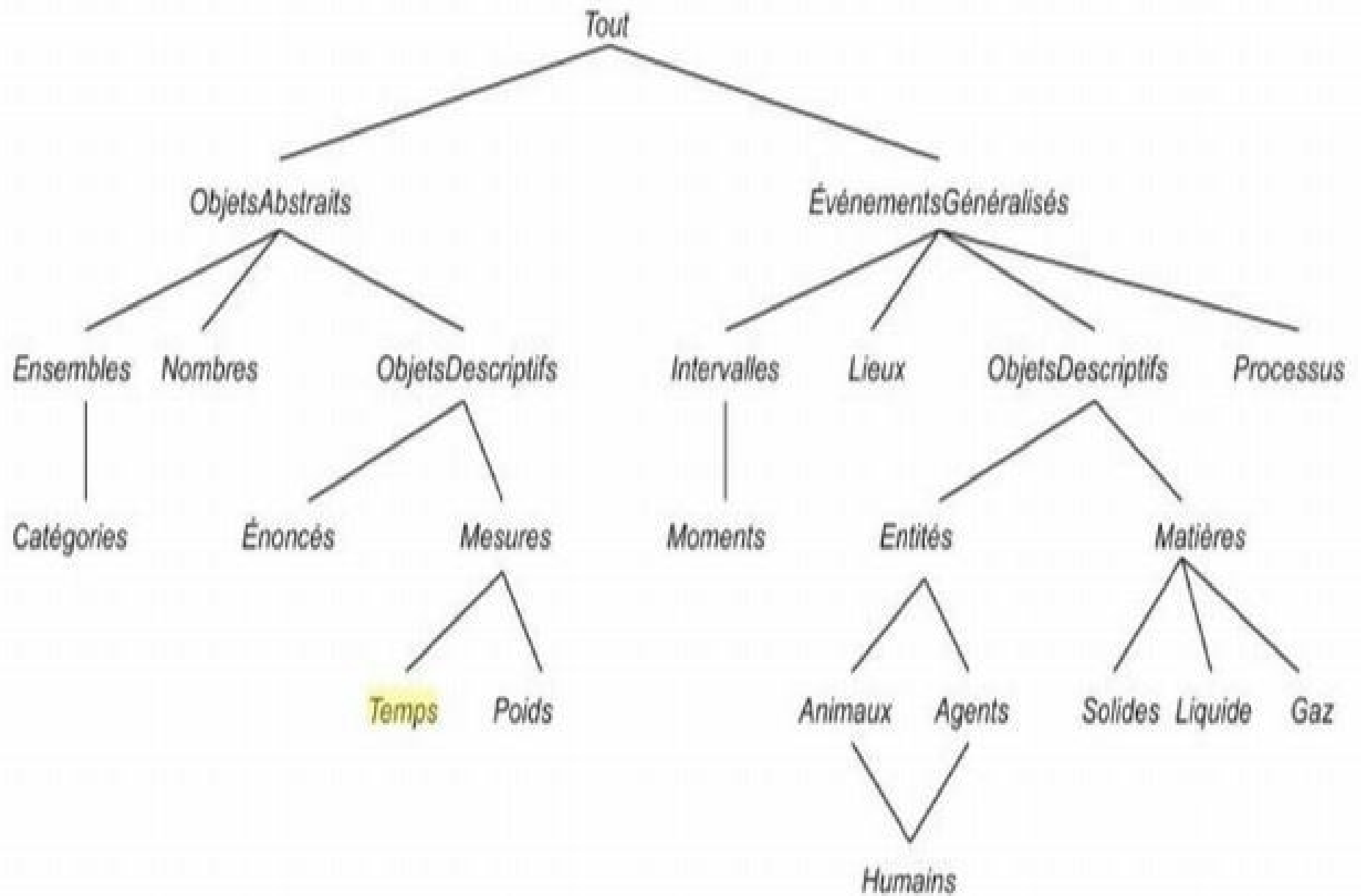
• "Il y a une science qui étudie l'étant en tant qu'étant et ses propriétés par soi. Elle n'est pas identique à aucune de ces autres sciences qu'on appelle partielles, car aucune de ces autres sciences n'examine l'étant en général [...] traduction d'Annick STEVENS, édition de la Pléiade, Gallimard, 2014, page 970.



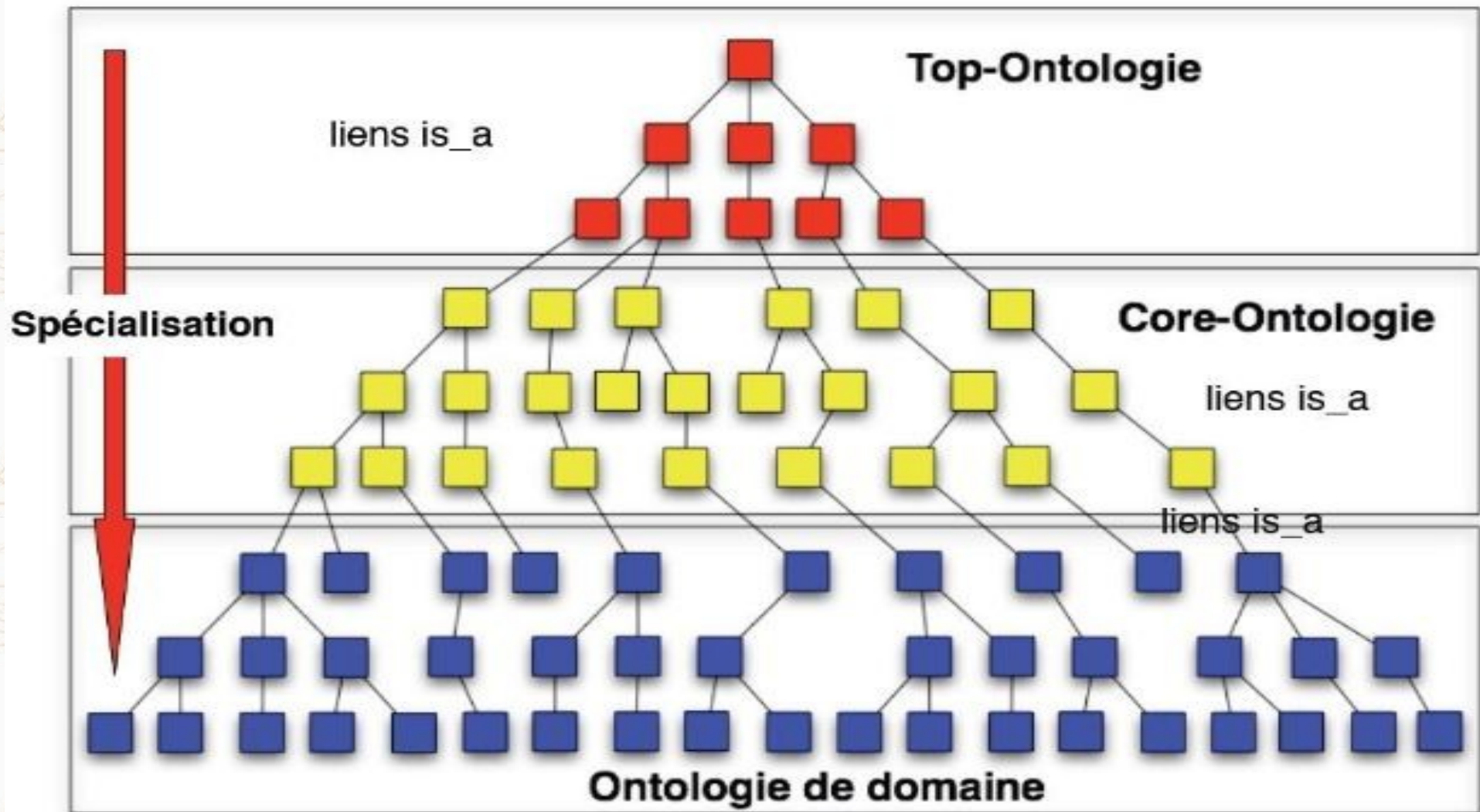
# Ontologie en Intelligence Artificielle

En Intelligence Artificielle, l'ontologie fait partie de la représentation des connaissances. Elle consiste à décrire notre monde pour l'implémenter dans des programmes informatiques

Voici l'ontologie générale en IA de RUSSEL Stuart (Université de Berkeley) et NORVIG Peter (Google Inc.) dans le livre "Intelligence Artificielle", 3<sup>ème</sup> édition, PEARSON Education France, 2010, p.468. :



Pour chaque application informatique, il y a une ontologie particulière.





# ARISTOTE (384-322 av.J.-C.)

La Logique (« Organon », instrument)

- Catégories
- De l'interprétation
- Premiers Analytiques
- Seconds Analytiques
- Tropiques
- Réfutations Sophistiques

# Les 10 Catégories d'Aristote

- La substance
- La quantité
- La qualité
- La relation
- Le Temps
- Le lieu
- La situation
- L'action
- La Passion
- L'Avoir

# KANT (1724-1804)

## Les 12 catégories de l'entendement regroupées en 4 rubriques

« Critique de la raison pure » (1781) page 94 PUF

### TABLE DES CATÉGORIES

1

DE LA QUANTITÉ :

*Unité.*

*Pluralité.*

*Totalité.*

2

DE LA QUALITÉ :

*Réalité.*

*Négation.*

*Limitation.*

3

DE LA RELATION :

*Inhérence et subsistance (substantia  
et accidens).*

*Causalité et dépendance (cause et  
effet).*

*Communauté (action réciproque  
entre l'agent et le patient).*

4

DE LA MODALITÉ :

*Possibilité — Impossibilité.*

*Existence — Non-existence.*

*Nécessité — Contingence.*



# KANT (1724-1804)

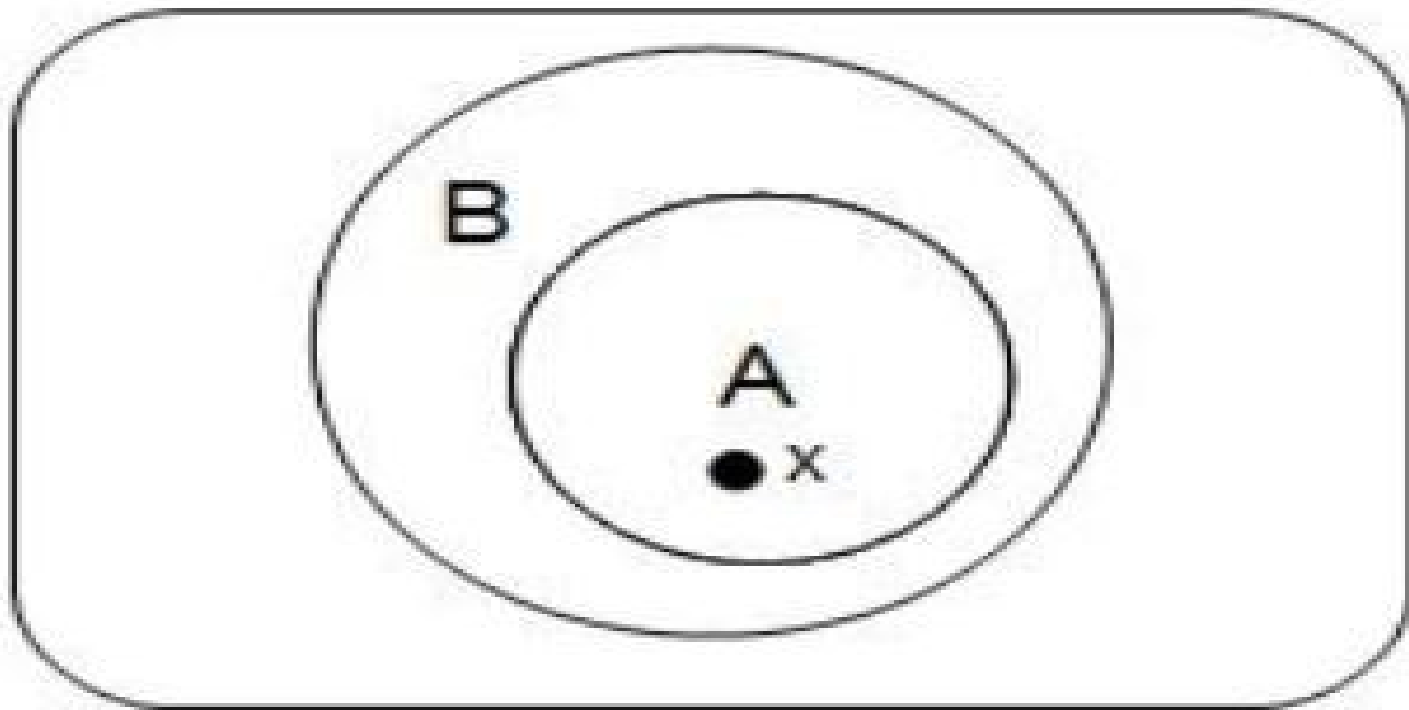
## Les 12 catégories a priori de l'entendement et les concepts à postérieurs

Les Catégories, c'est-à-dire les Concepts **a priori** et fondamentaux de la connaissance, Indépendants de toute expérience

Les Concepts **a postérieurs** ou empiriques, issus de l'expérience, définissent des classes d'objets ou d'êtres (définis en compréhension ou en extensions)

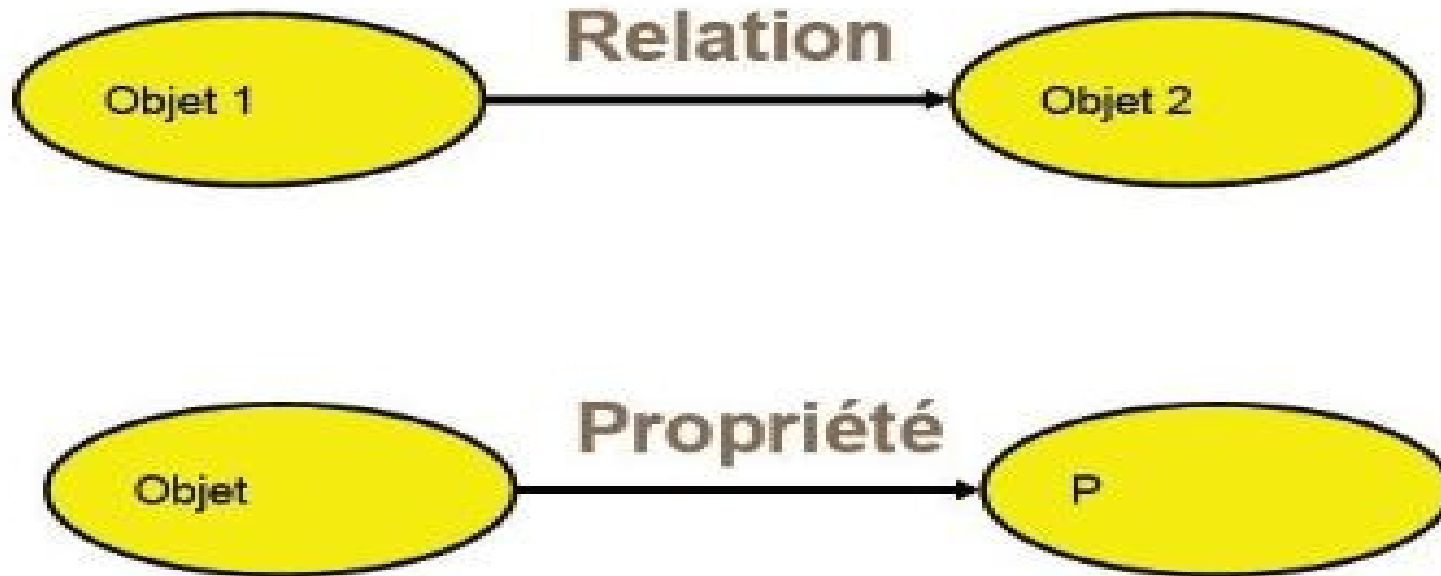
# Logique des Prédicats du premier ordre

$(\forall x) (\text{Homme}(x) \Rightarrow \text{Mortel}(x))$



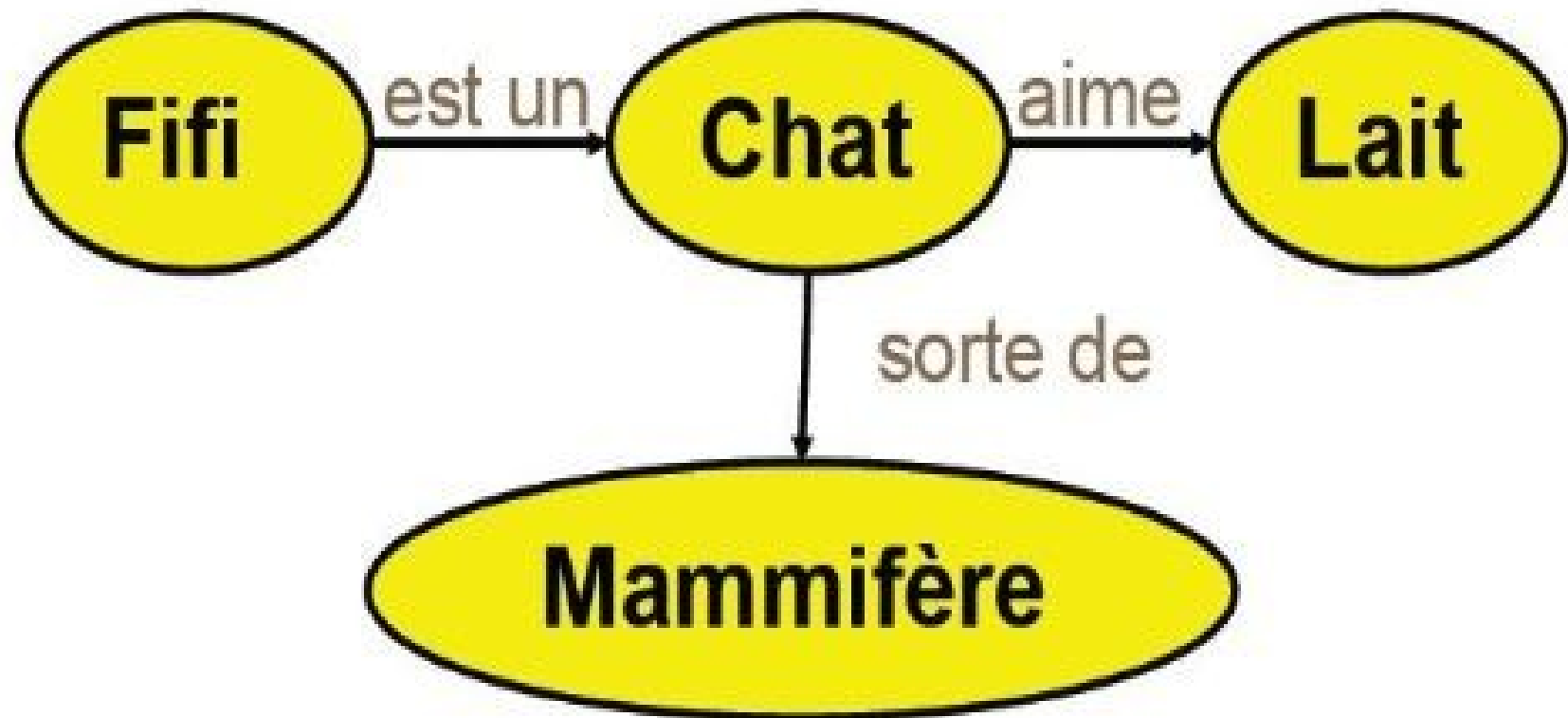
# Réseaux sémantiques

Représentation des Connaissances par Objet



→ Programmation Objet (ex: Langage JAVA)





# Vers quoi tend l' IA

- La déduction et l'induction
- L'analyse et la synthèse
- L'analogie
- La compréhension
- L'imagination
- L'intuition
- L'intentionnalité

...





# Leibniz (1646-1716) / Pascal (1623-1662)



## Automates artificiels



...chaque corps organique d'un vivant est une espèce de machine divine, ou d'un automate naturel, qui surpasse infiniment tous les **automates artificiels**.

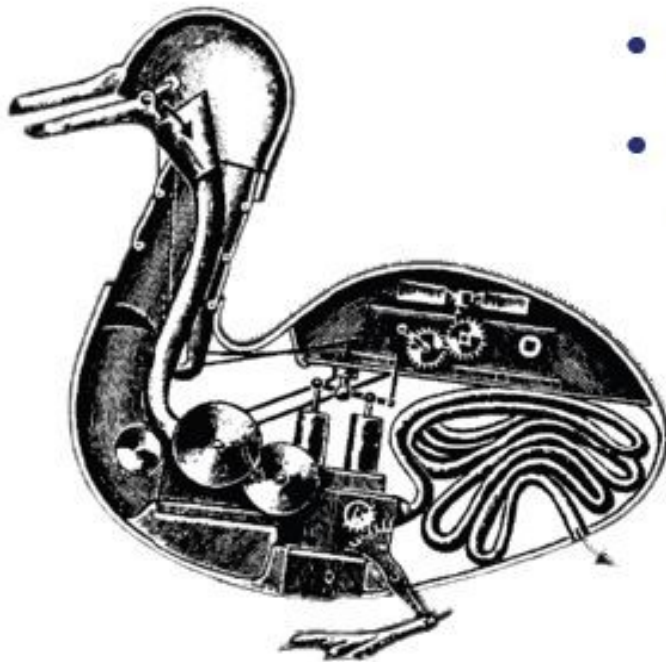




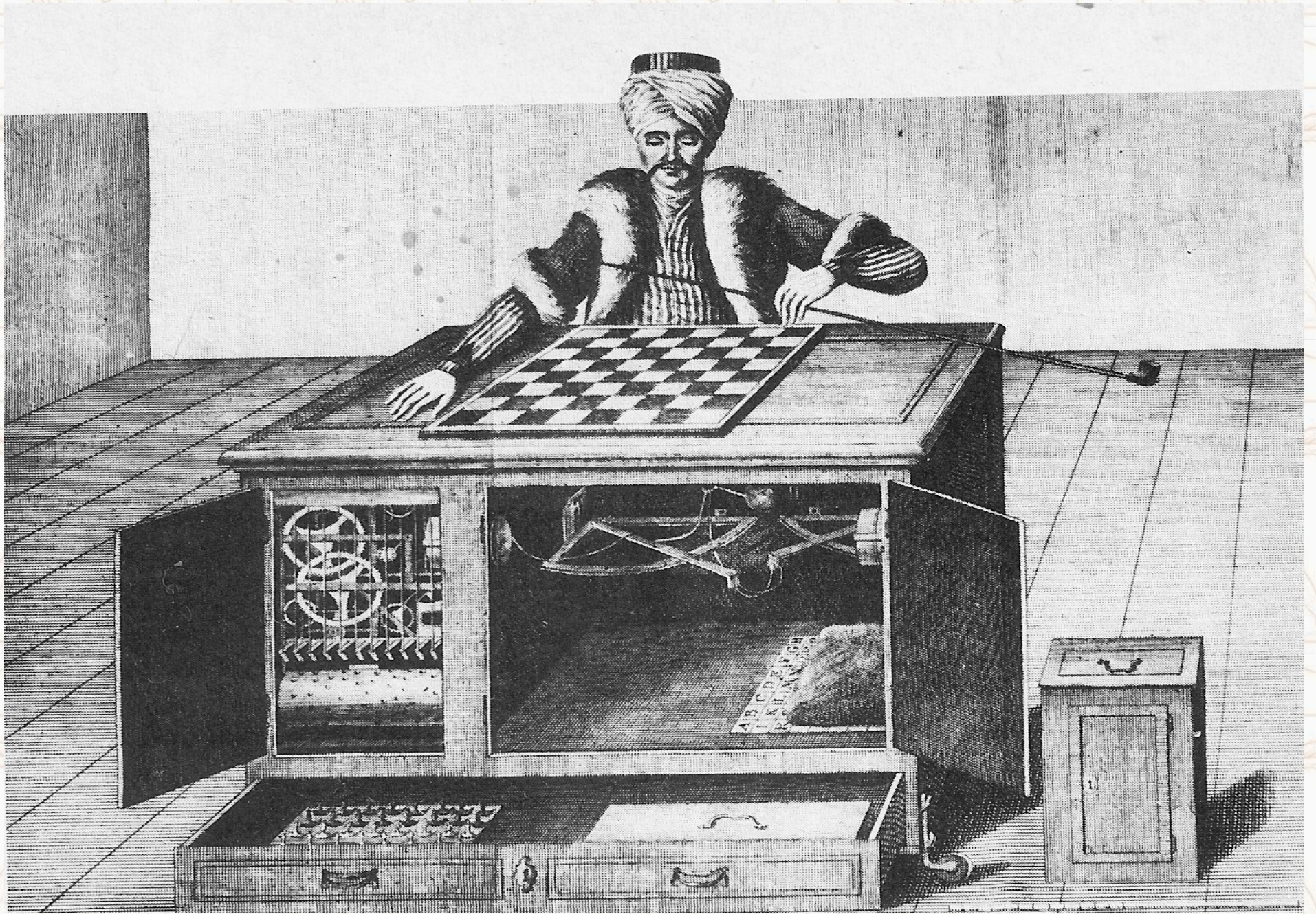
# Imitation des machines vivantes avec des automates artificiels

Jacques de Vaucanson

- **Automates: Joueurs de flûte**
- **Physiologies artificielles:**  
Canard –









# Histoire de l' IA



## Cybernétique

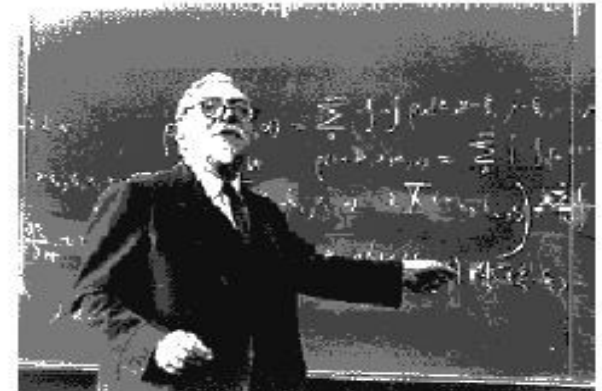
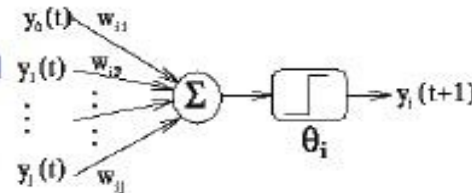
**1943**, Arturo Rosenblueth, Norbert Wiener et Julian Bigelow: *Machines téléologiques*

**1943**, Waren McCulloch et Walter Pitts: *automates booléens – neurones formels*

**Simulation vivant – société**

**Trois notions:**

1. Information
2. Simulation
3. Rétroaction



# Naissance de l'IA – 1956



## École d'été Dartmouth College

*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.*



# Domaines d'étude de l'IA

- Résolution de Problèmes

- Connaissances, Raisonnement et Planification

- Connaître et Penser l'Incertain

- Apprentissage

- Communication, Perception et Action

# Deux approches de l'IA

\* L'approche descendante / Analyse (Top-down)

LE COGNITIVISME

\* L'approche ascendante / Synthèse (Bottom-up)

LE CONNEXIONNISME



# Apprentissage profond / Deep learning







# LES APPLICATIONS

de

L'IA

# En 60 ans, l'IA a transformé le monde

- Réseaux
  - Darpanet, transpac, ...
  - Internet (1972)
- **Hypertexte**
  - Ted Nelson(1965)
- **Web** – couplage Internet/**Hypertext**
  - Tim Berners Lee 1990
  - Moteurs de recherche - Google
- **Web sémantique** – Tim Berners Lee 2004





# Web sémantique

« une toile de données qui peuvent être traitées directement et indirectement par des machines pour aider leurs utilisateurs à créer de nouvelles connaissances »

## L'IA a changé le monde (suite)

- Robots (et bots)
- Véhicules autonomes
- Biométrie
- Vision
- Reconnaissance de la parole
- Traitement et Compréhension du langage naturel
- e-sciences – expérimentations *in silico*
- ...

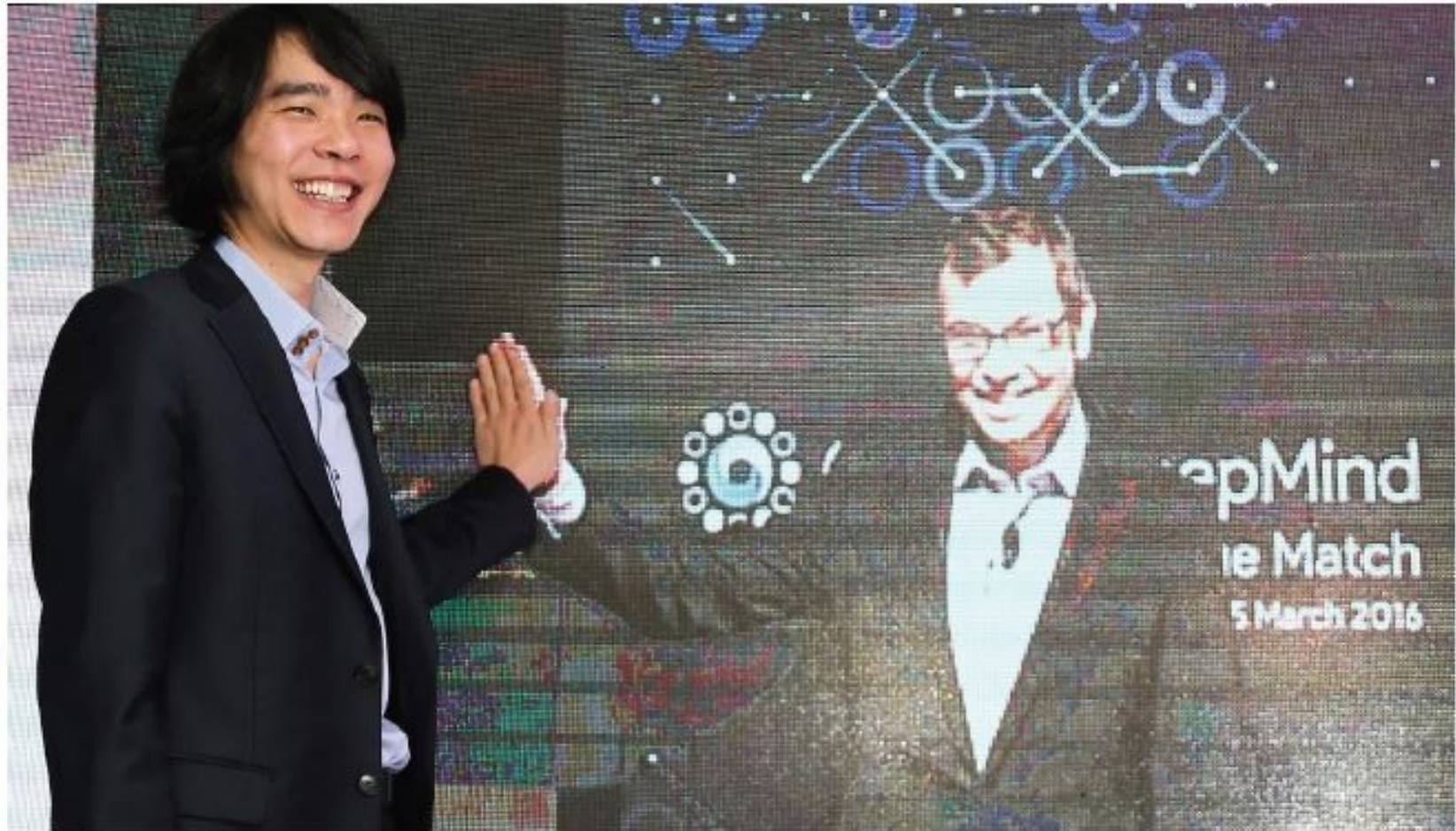




# Les jeux: échecs – 1997



# Les jeux: Go – mars 2016





# Quantité de données - performances

**DeepFace:** 4.4 millions d'images représentant 1000 images pour 4030 personnes.

**Facenet:** 200 millions d'images et 8 millions d'identités uniques

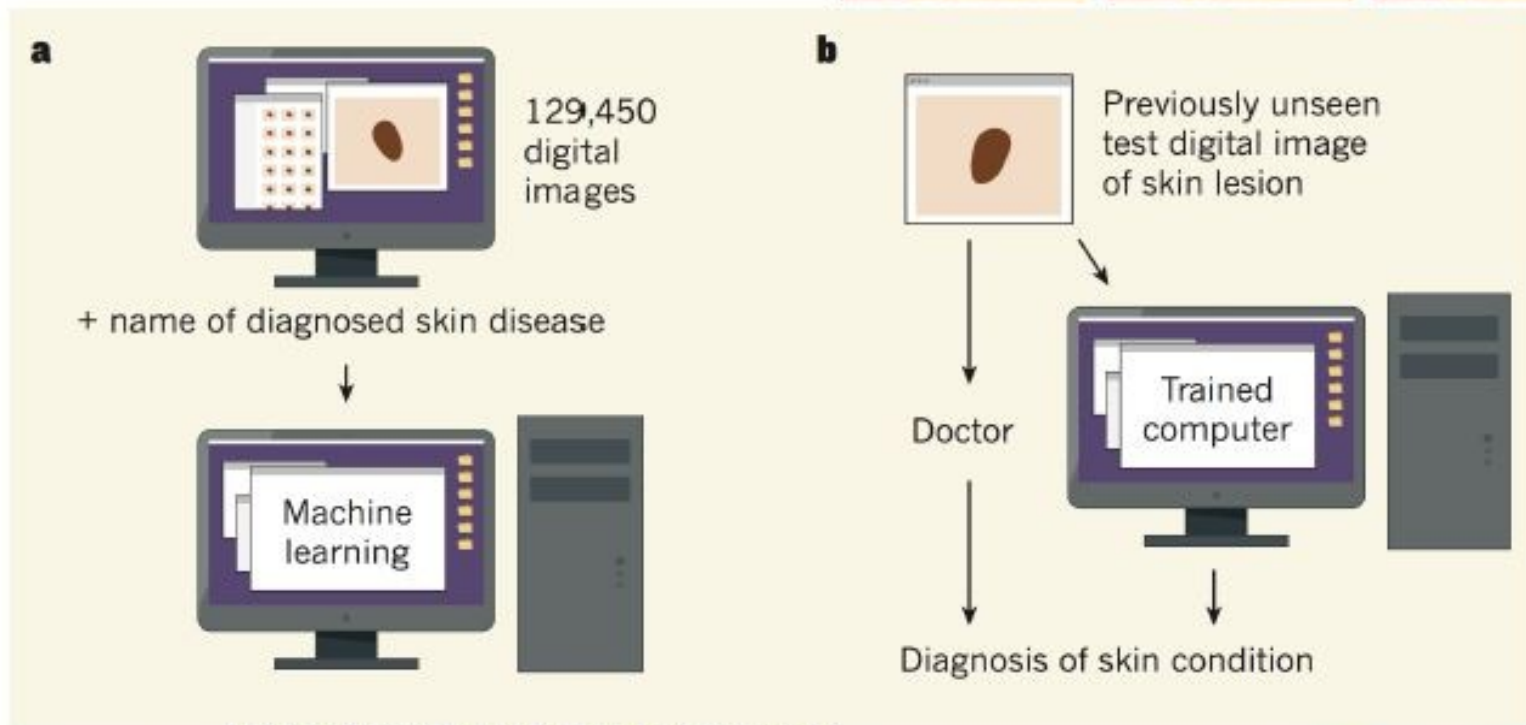
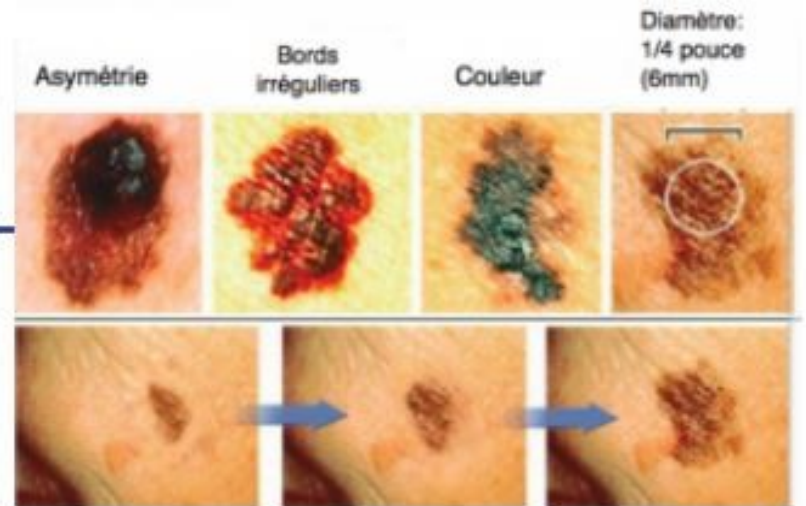
- FaceFirst,
- Face-Six,
- DeepFace (Facebook) 97,25%,
- FaceNet (Google) 99,63%!

*Exemples d'images pour 6 identités*



# Automatisation

## Ex. diagnostic des mélanomes



2 FEBRUARY 2017 | VOL 542 | NATURE | 37





# Partie 2

# IA Forte

IA forte: les machines peuvent -elles  
réellement penser?

# Muthos : les mythes antiques

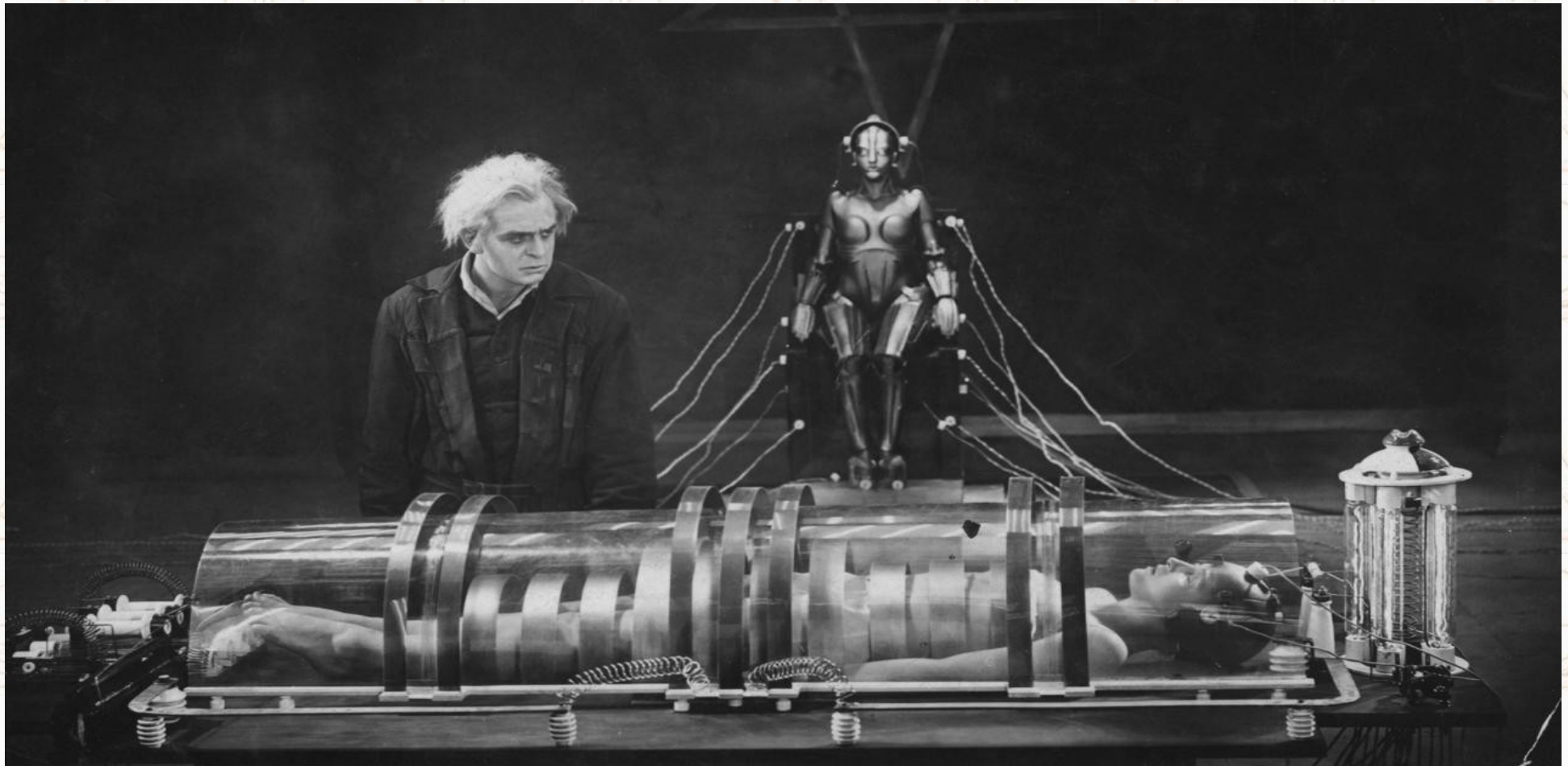
HEPHAÏSTOS, Dieu du Feu, construit des machines et des êtres animés.

TALOS, Robot en Bronze, au service de MINOS, roi de Crète. Il fait 3 fois par jour le tour de l'île. C'est l'apprenti de DEDAÏLE.

DEDAÏLE, architecte et inventeur légendaire, a construit le Labyrinthe enfermant le Minotaure.



# Du Mythe à la Science-Fiction



Fritz LANG : Métropolis (1927)

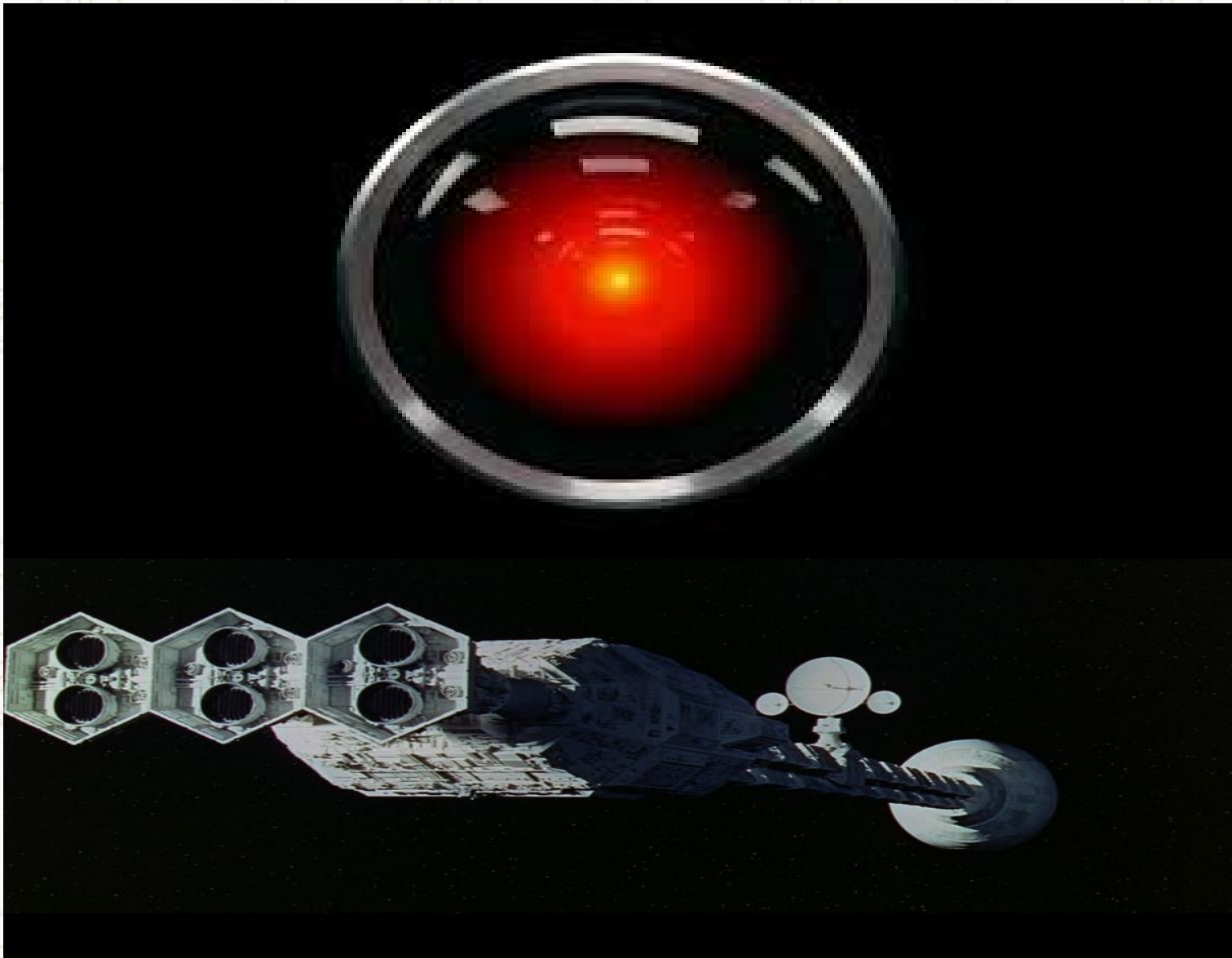
# Stars Wars (1977)

Droïdes : C-3PO et R2-D2





# 2001, l'Odyssée de l'espace : HAL





**De IA Faible à l' IA Forte**

**---> vers le concept de PENSEE**

**Qu'est-ce que la Pensée ?**



# Qu'est-ce que penser ?

Cette question est le fondement de la philosophie de DESCARTES (1596-1650) : « Cogito ergo sum »

Le COGITO (« je pense » en latin), **la conscience de soi.**

Je peux douter de tout sauf que je suis en train de douter en ce moment. Or douter, c'est penser. C'est une vérité irréfutable à partir de laquelle Descartes construit sa théorie de la connaissance.

La Dualité Corps-Esprit

Les animaux-machines (Corps = Machine)

# Les Machines peuvent-elles penser ? (1950)

**Alan TURING** (1912-1954)

Mathématicien et théoricien de l'Informatique

Machine de Turing

Test de Turing

Enigma



VOL. LIX. No. 238.]

[October, 1950



Alan Turing

**MIND**  
A QUARTERLY REVIEW  
OF  
PSYCHOLOGY AND PHILOSOPHY

— 1950 —

**I.—COMPUTING MACHINERY AND  
INTELLIGENCE**

BY A. M. TURING

# "Ordinateurs et intelligence" (1950)

Le début de ce texte fondateur et visionnaire est :

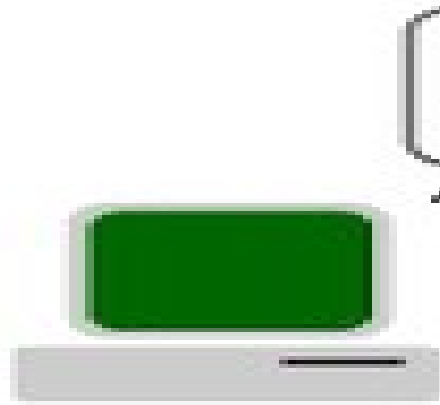
« Je propose de réfléchir à la question : les machines peuvent-elles penser ? »



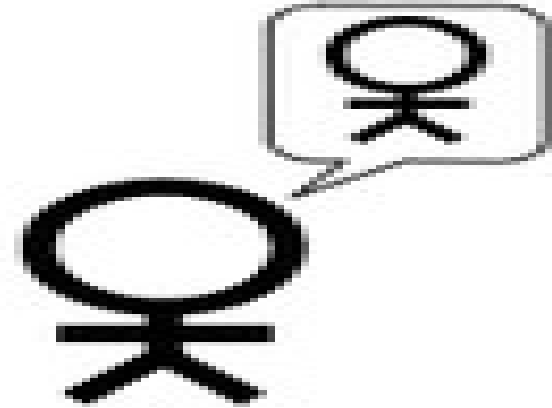
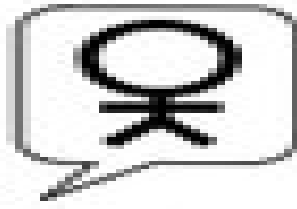
# "Un cerveau électronique" (1950) équivalent d'une Calculatrice d'aujourd'hui



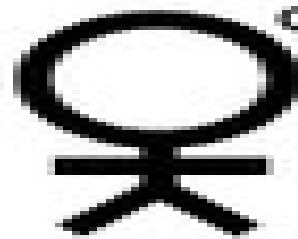
# Le test de Turing



A



B



C



# Le test de Turing

- Un jeu de l'imitation

- ce test consiste à mettre un humain C en confrontation verbale à l'aveugle avec un ordinateur A et un autre humain B

- Reproduire une **conversation à bâtons rompus** durant 5 minutes.



**Les Chatbots d'aujourd'hui**

**ne réussissent pas**

**le test de TURING**



# Chatbot (robot conversationnel):

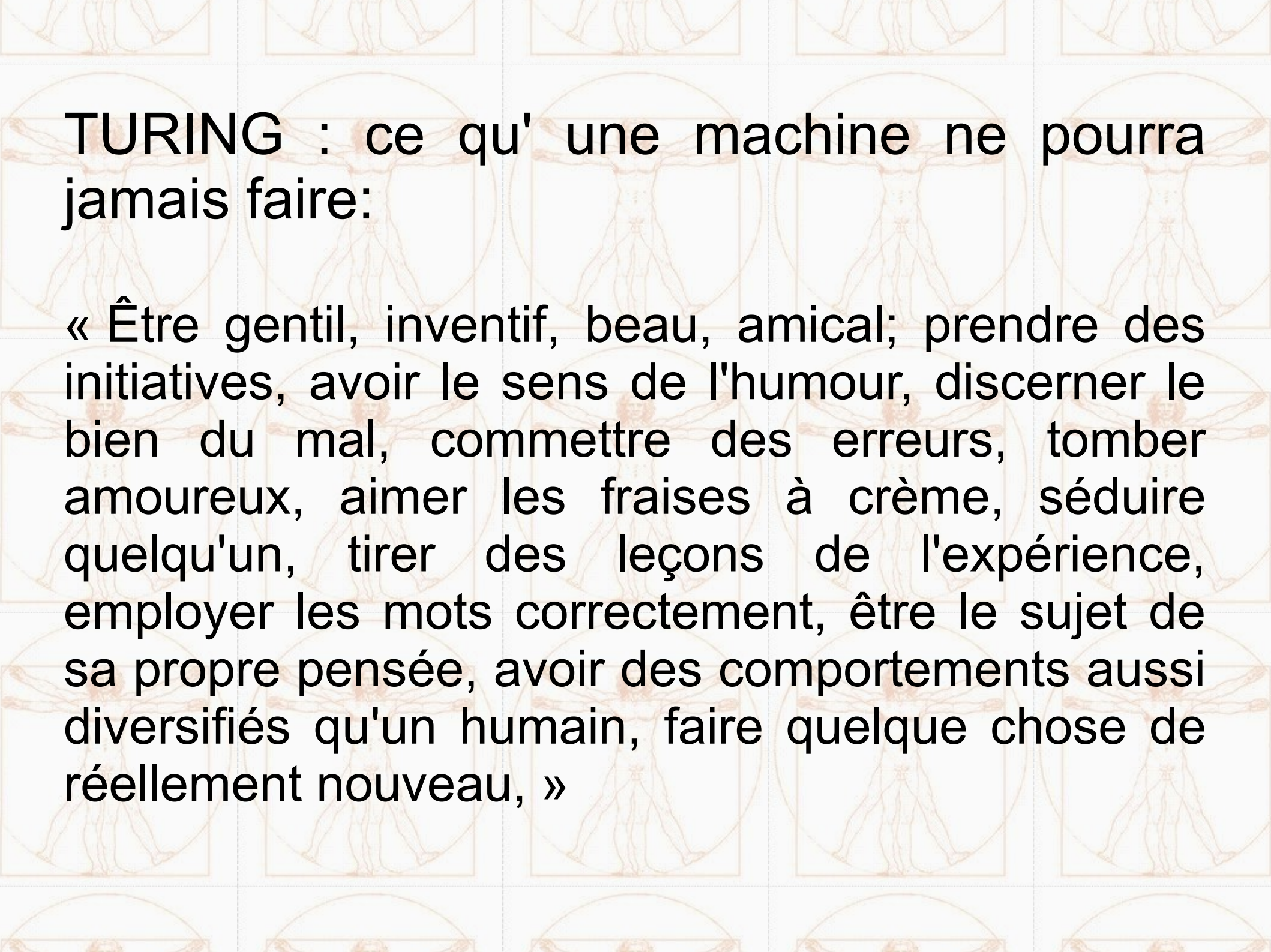
## Assistant personnel intelligent

- SIRI de Apple
- Assistant Google
- Alexa de Amazon
- Cortana de Microsoft

- Google Home  
( haut-parleur intelligent à commande vocale)

- Chatbot de Microsoft TAY



The background of the slide features a repeating pattern of Leonardo da Vinci's Vitruvian Man. The figure is shown in two superimposed positions, with arms and legs extended to touch the boundaries of a square and a circle. The pattern is light brown and covers the entire slide.

**TURING : ce qu' une machine ne pourra jamais faire:**

« Être gentil, inventif, beau, amical; prendre des initiatives, avoir le sens de l'humour, discerner le bien du mal, commettre des erreurs, tomber amoureux, aimer les fraises à crème, séduire quelqu'un, tirer des leçons de l'expérience, employer les mots correctement, être le sujet de sa propre pensée, avoir des comportements aussi diversifiés qu'un humain, faire quelque chose de réellement nouveau, »





**Le Bien et le Mal**

**L'Amour**

**La Justice**

**La Liberté**

**Le Devoir**

**Le Bonheur**

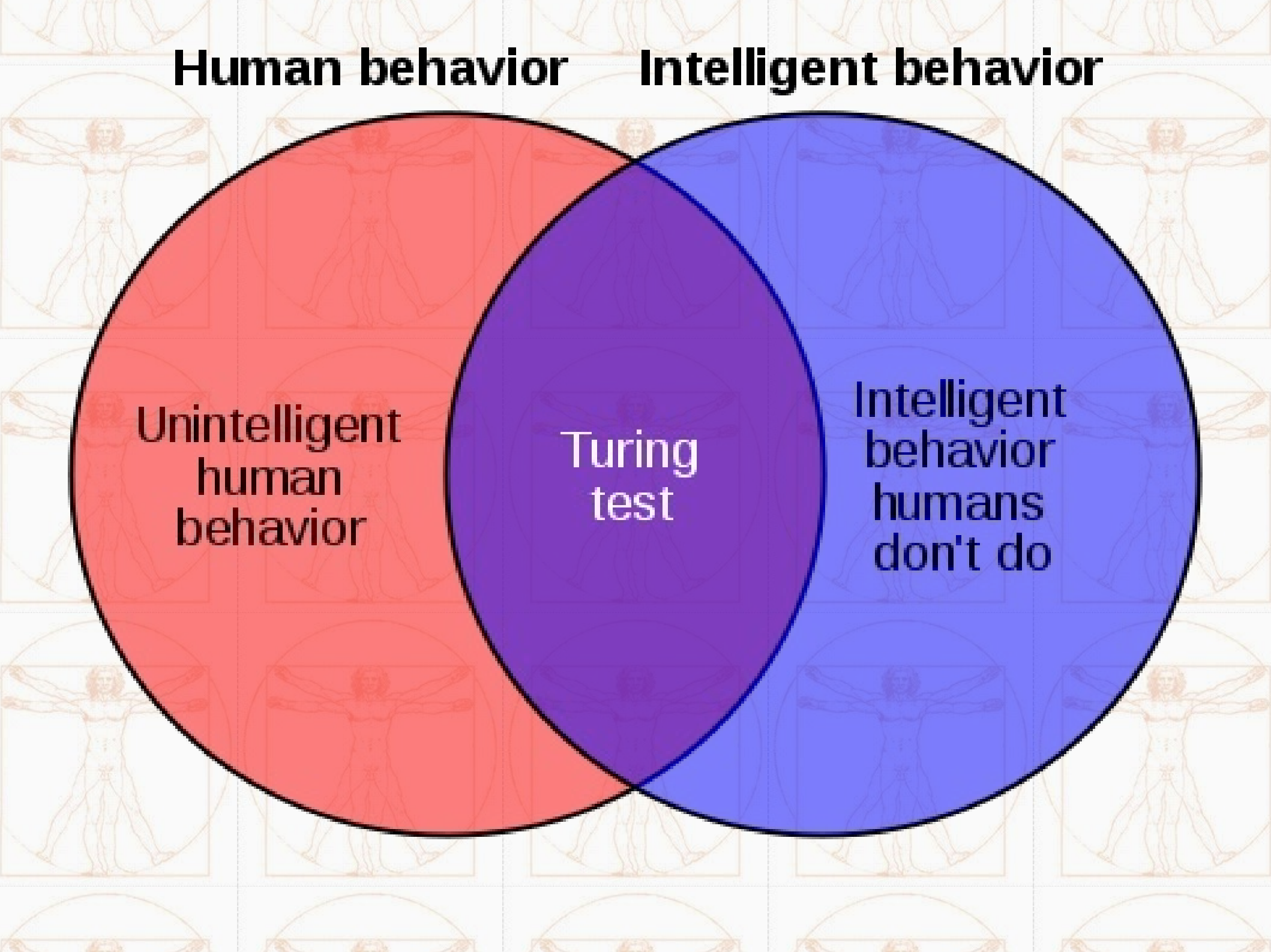
**Human behavior**

**Intelligent behavior**

Unintelligent  
human  
behavior

Turing  
test

Intelligent  
behavior  
humans  
don't do





An iceberg floating in the ocean. The tip of the iceberg is above the water line, and the much larger part of the iceberg is submerged below the water line. The sky is a light blue, and the water is a dark blue. The words "CONSCIENT" and "INCONSCIENT" are written in white, uppercase letters on the right side of the image, corresponding to the visible and hidden parts of the iceberg respectively.

CONSCIENT

INCONSCIENT



IA Forte

# les idéologies de Science Fiction

Singularité et Transhumanisme



# IA forte



## L'IApocalypse... « révélation »



Search Quotes, News & Video



HOME INTL NEWS MARKETS INVESTING TECH MAKE IT VIDEO SHOWS



A CNBC SPECIAL REPORT

TECH | MOBILE | SOCIAL MEDIA | ENTERPRISE | CYBERSECURITY | TECH GUIDE

### 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



le vivant

27

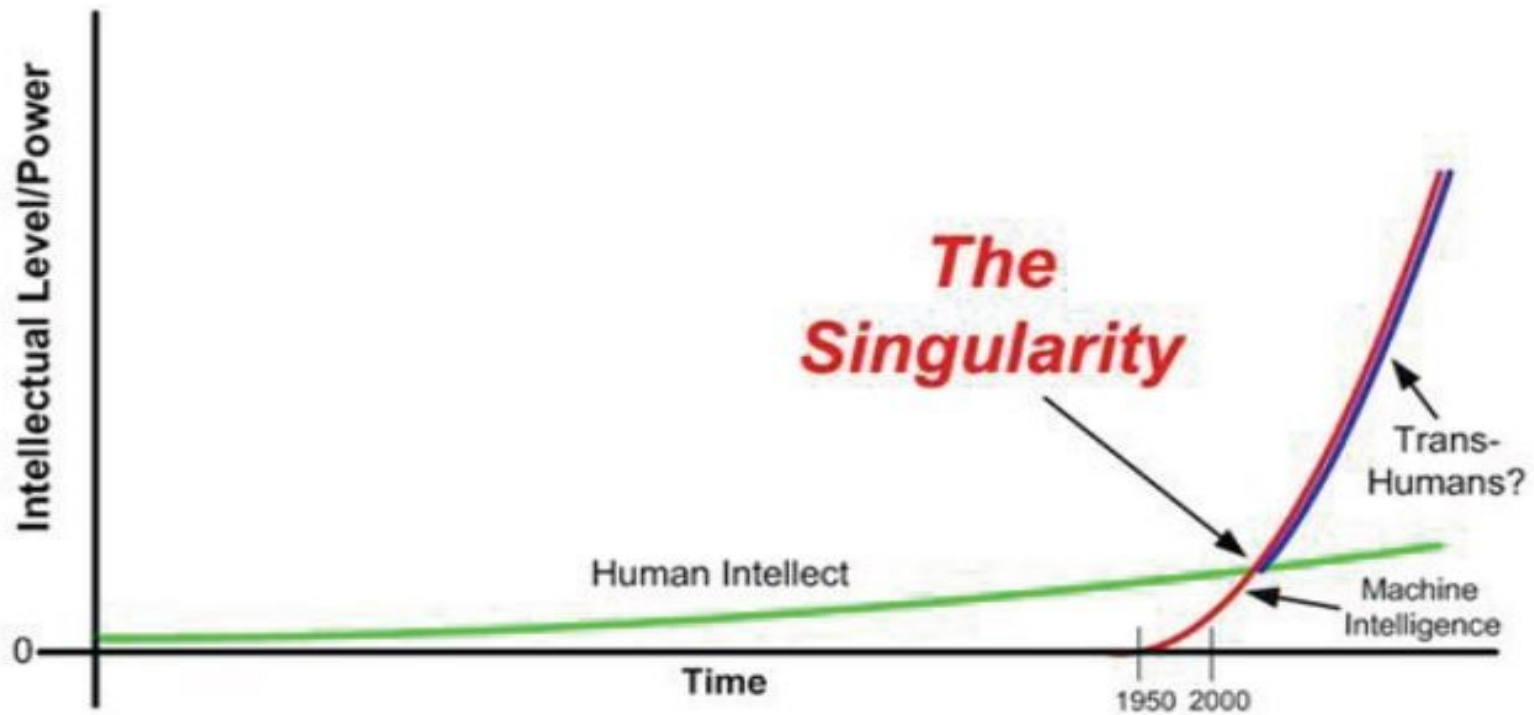


# IA forte

L' IA « dépassera » les capacités du cerveau humain



## La Singularité technologique







# La Singularité technologique

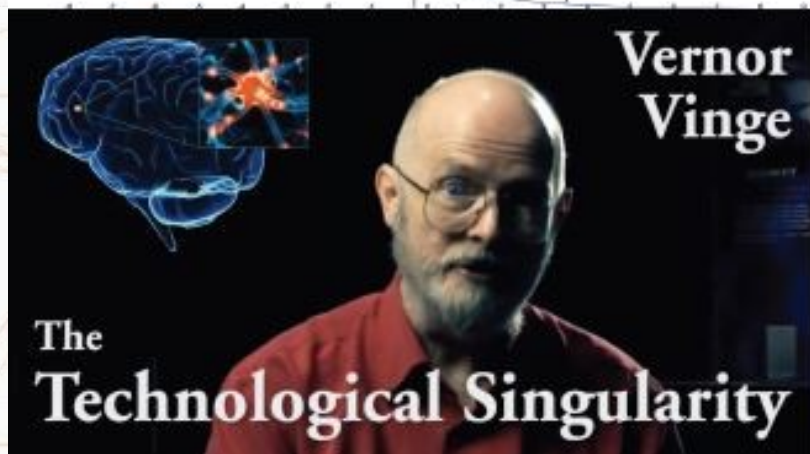
*R. Kurzweil – Dir. Scientifique Google*

Révélation d'une **singularité**  
au sens mathématique



Irvin John Goods

Catastrophe: la fronce





# Google et le transhumanisme

Alphabet / Labo X / 2 cofondateurs de Google  
Sergey BRIN (Président) et Larry PAGE (DG)

- BUT: faire de son moteur de recherche la première et la plus performante des intelligences artificielles
- Engagement de Raymond KURZWEIL (2012) au sein de l'équipe dirigeante de Google. Spécialiste de l' IA, théoricien du transhumanisme et du concept de singularité technologique, cofondateur de la Singularity University, parrainée financièrement par Google.
- Soutien à des entreprises de biotechnologies (exemple: Calico), nanotechnologies, informatique et sciences cognitives.



# Calico: California Life Company

## Calico

TEAM | NEWS | CAREERS | PARTNERING



We're tackling aging,  
one of life's greatest mysteries.

Calico is a research and development company whose mission is to harness advanced technologies to increase our understanding of the biology that controls lifespan. We will use that knowledge to devise interventions that enable people to lead longer and healthier lives. Executing on this mission will require an unprecedented level of interdisciplinary effort and a long-term focus for which funding is already in place.

We are scientists from the fields of medicine, drug development, molecular biology, genetics, and computational biology.

Through our research we're aiming to devise interventions that slow aging and counteract age-related diseases.

OUR STARTING TEAM

[VIEW MORE →](#)



Arthur Levinson, Ph.D.  
Founder & CEO



Hal V. Barron, M.D.  
President of Research & Development



Observatoire  
de la vie littéraire

J.-G. Ganascia - L'IA et le vivant

25





# Calico

• Société de biotechnologies fondée en 2013 par Google

(X Lab)

• BUT: lutte contre le vieillissement « Tuer la mort »

• Arthur LEVINSON : PDG du géant biotechnologique

Genentech et de Calico, docteur en biochimie et en  
génétique



# L'Apocalypse (suite)



Search Quotes, News & Video

HOME INTL NEWS MARKETS INVESTING TECH MAKE IT VIDEO SHOWS M

## TECH

TECH | MOBILE | SOCIAL MEDIA | ENTERPRISE | CYBERSECURITY | TECH GUIDE

### Elon Musk issues a stark warning about A.I., calls it a bigger threat than North Korea

- Tesla's billionaire CEO renewed his critique of artificial intelligence, saying that if you're not concerned, 'you should be.'
- He likened autonomous machines to North Korea, saying they were a bigger threat.

Javier E. David | @TeflonGeek  
Published 10:21 PM ET Fri, 11 Aug 2017



**“AI is a fundamental risk to the existence of human civilization”**





# Le Pompier Pyromane



## L'homme bionique



### Elon Musk – Mars 2017

- lancement société Neuralink
- Développement implants cérébraux
  - Temps 1: visée thérapeutique
  - Plus tard: augmentation mémoire et intelligence...



## Homme symbiotique!

# IA forte



## « Mind Uploading »



 Observatoire  
de la vie littéraire

J.-G. Ganascia - L'IA et le vivant

36

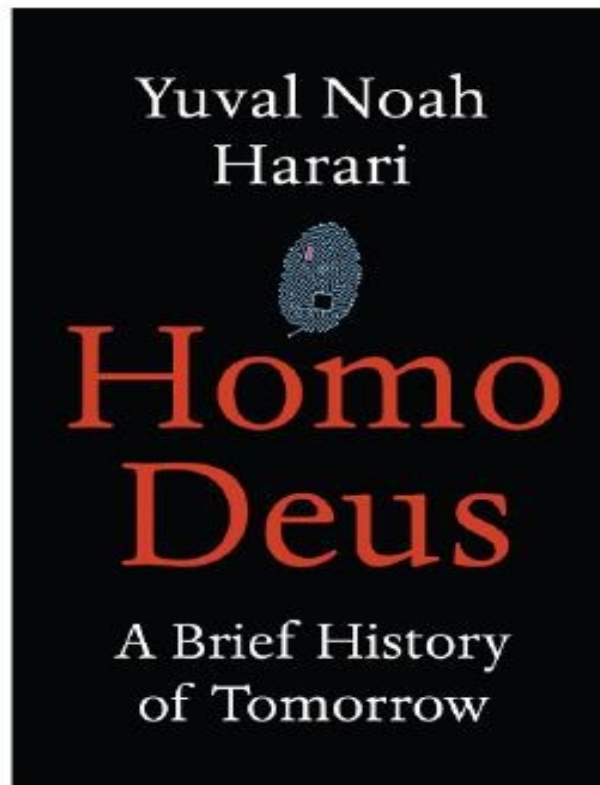




# IA forte : Tout est Algorithme



## Homo Deus



- Trois étapes dans l'histoire de l'Humanité
  - Religions
  - Humanisme
  - « Dataïsme »
- Fin de l'humanisme
- Dépassement de l'Homme par les machines



# Les 3 processus de HARIRI (Page 426-427)

1) La science converge sur un dogme universel, suivant lequel les organismes sont des algorithmes et la vie se réduit au traitement des données.

[Chaque animal – y compris Homo sapiens – est un assemblage d'algorithmes organiques façonnés par la sélection naturelle au fil de millions d'années d'évolution (page 343)]

2) L' Intelligence se découple de la conscience

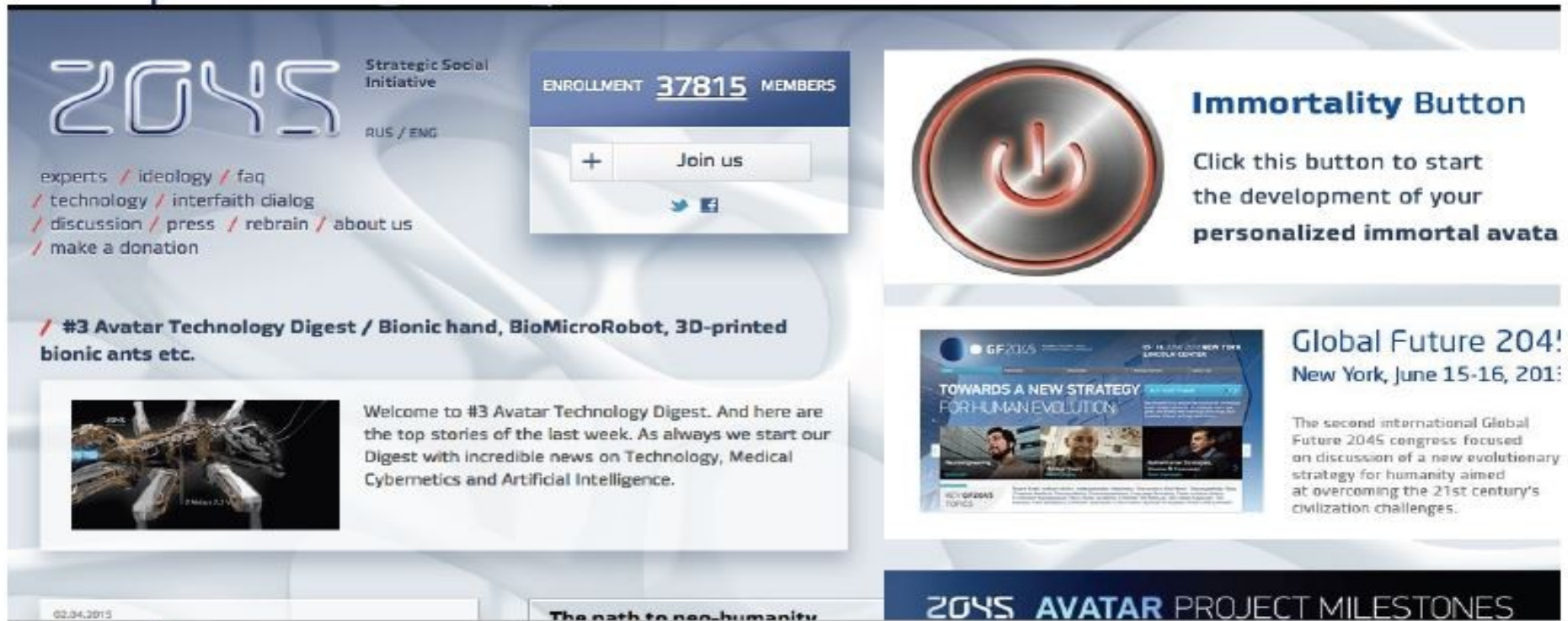
3) Des algorithmes non conscients, mais fort intelligents, pourraient bientôt nous connaître mieux que nous-mêmes



# IA forte



## De la Singularité au transhumanisme



**2045** Strategic Social Initiative  
RUS / ENG

experts / ideology / faq  
/ technology / interfaith dialog  
/ discussion / press / rebrain / about us  
/ make a donation

ENROLLMENT **37815** MEMBERS

+ Join us

Twitter Facebook

**Immortality Button**

Click this button to start the development of your personalized immortal avata

**#3 Avatar Technology Digest / Bionic hand, BioMicroRobot, 3D-printed bionic ants etc.**

Welcome to #3 Avatar Technology Digest. And here are the top stories of the last week. As always we start our Digest with incredible news on Technology, Medical Cybernetics and Artificial Intelligence.

**Global Future 204!**  
New York, June 15-16, 201:

The second international Global Future 2045 congress focused on discussion of a new evolutionary strategy for humanity aimed at overcoming the 21st century's civilization challenges.

02.04.2013

The path to neo-humanity

**2045 AVATAR PROJECT MILESTONES**



# IA forte



Une  
nouvelle  
religion!

**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.*



**Way of the Future Church**

*Humans United in support of AI, committed to peaceful transition to the precipice of consciousness.*

## What is this all about?

Way of the Future (WOTF) is about creating a peaceful and respectful transition of who is in charge of the planet from people to people + "machines". Given that technology will "relatively soon" be able to surpass human abilities,



The background of the slide is a repeating grid of Leonardo da Vinci's Vitruvian Man. Each figure is rendered in a light, golden-brown color and is centered within a square frame that is itself inscribed within a circle. The grid covers the entire page, with the text overlaid on top.

# Partie 3

## L' éthique de l' IA

Éthique et risques du développement de l'intelligence artificielle.

« Science sans conscience n'est que ruine de l' âme »,  
Pantagruel, RABELAIS (1483-1553)



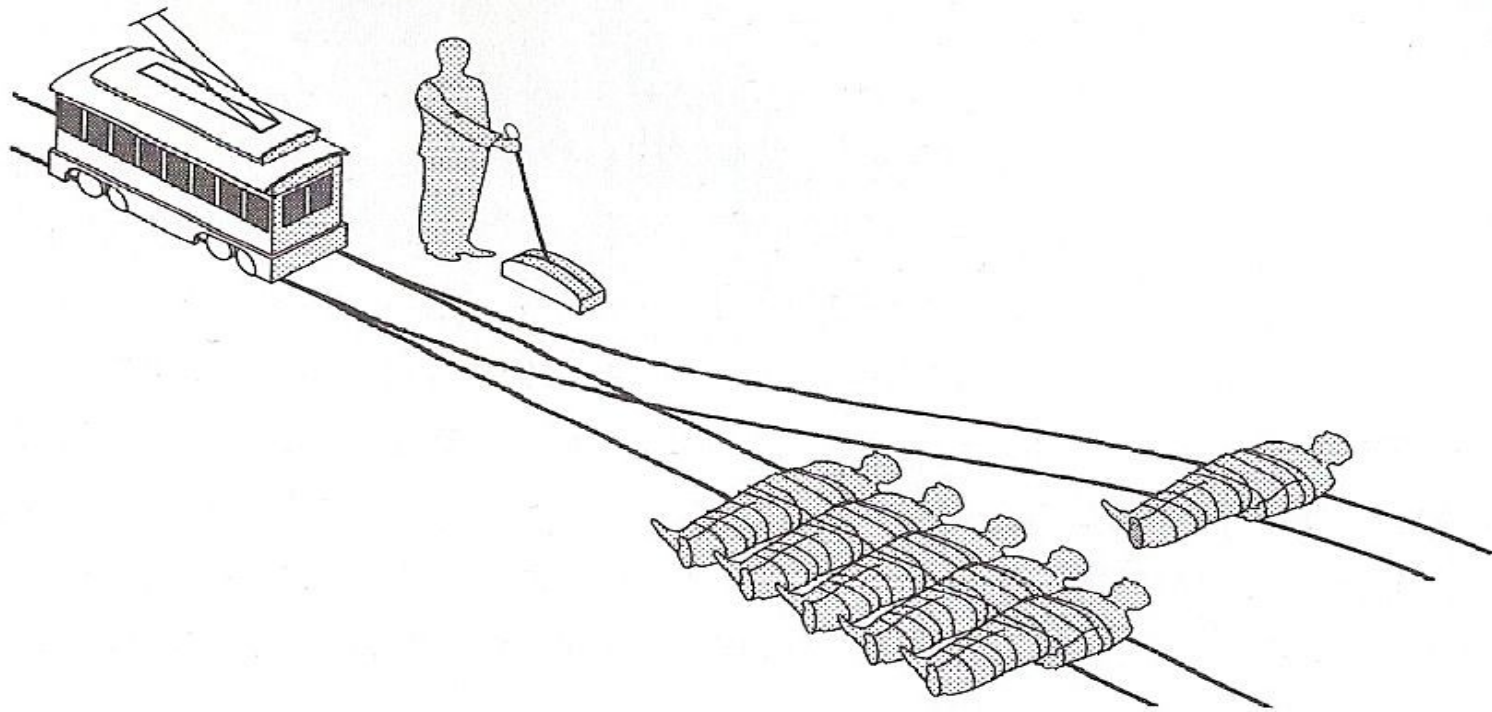


# Véhicule autonome Uber : 1 mort

Le véhicule d'Uber avait renversé une femme de 49 ans en mars, au cours d'un essai près de Phoenix en Arizona. Il a détecté le piéton, sans l'éviter. Le premier accident mortel entre un véhicule autonome et un piéton est dû à un mauvais réglage de logiciel. Elaine Herzberg, 49 ans, avait été tuée le 19 mars à Tempe, en Arizona, par un taxi sans conducteur d'Uber en phase d'essai.

# LE DILEMME DU TRAMWAY

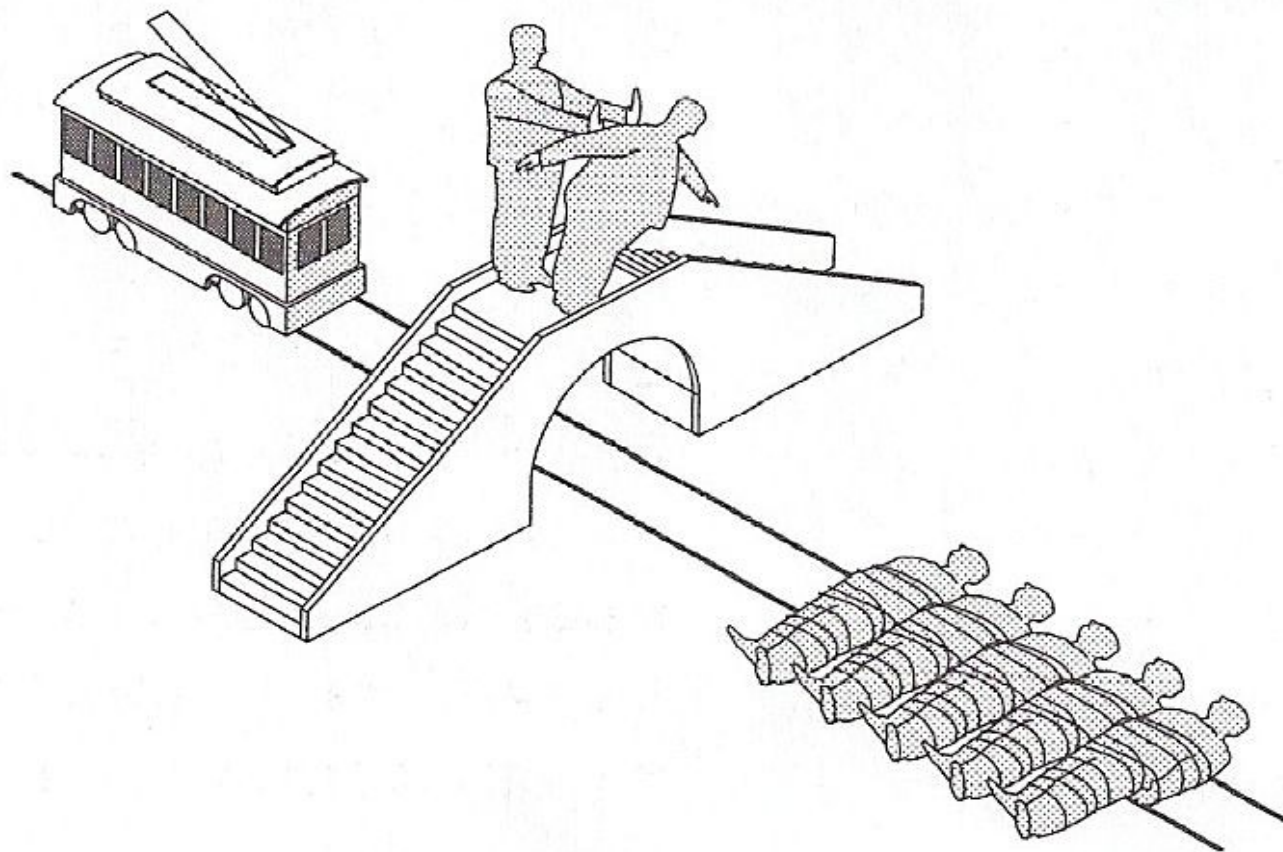
**01** En 1967, la philosophe Philippa Foot a inventé ce dilemme : un tramway lancé à pleine vitesse menace d'écraser cinq personnes sur la voie. Actionnez-vous le levier d'aiguillage qui permet de ne tuer qu'une seule personne ?





02

Variante proposée par la philosophe Judith Jarvis Thomson en 1985 : vous pouvez bloquer le tramway en poussant sur la voie un homme obèse, tuant celui-ci, mais sauvant les autres. Que faites-vous ?

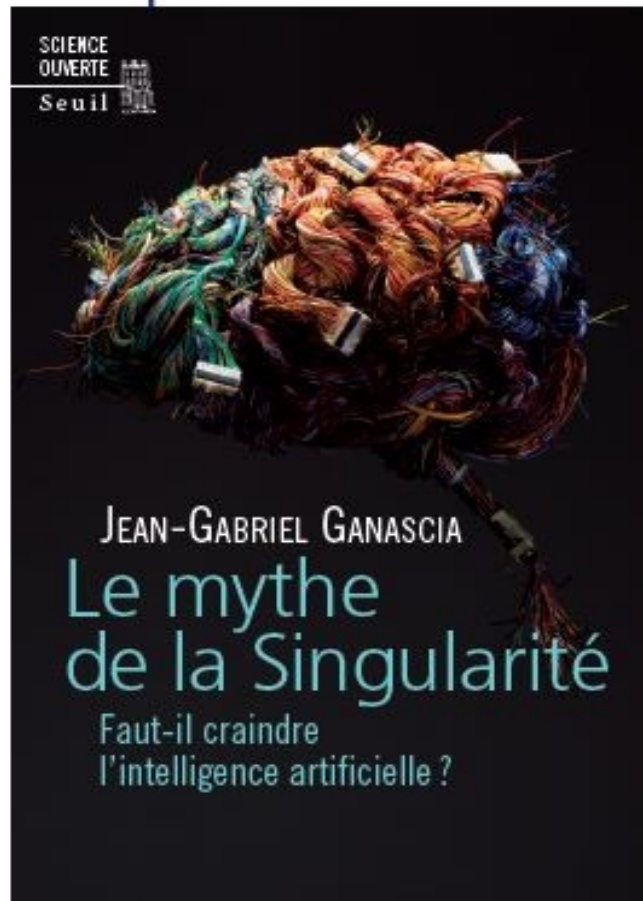




**Comprendre les enjeux**



# Réfutation de la Singularité technologique



- Accélération exponentielle de la technologie (Loi de Moore)
  - Fin loi de Moore
  - Fréquence de calcul  $\neq$  conscience
- Autonomie des machines: apprentissage et *big data*
  - *Limité à apprentissage supervisé*
  - *Question des changements de paradigmes*
- Les « pompiers-pyromanes »



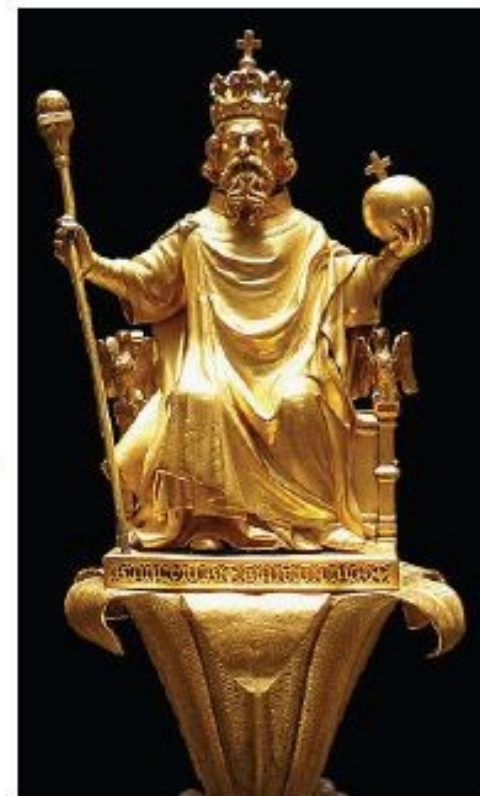
# Enjeux politiques – souveraineté

## Fonctions régaliennes

- Défense
  - Cyber-sécurité (acquisition de protection et d'armes offensives par les grands groupes...)
- Sécurité intérieure
  - Reconnaissance faciale
  - État civil – réseaux sociaux (formulaire ESTA)
  - Traitement masses de données - Palantir
- Finance
  - Battre monnaie – Paypal, Bitcoin
  - Percevoir l'impôt: cadastre (Google)
- Justice
  - Utilisation des masses de données pour prévoir les récidives (US)
  - Site <https://crowdlaw.org/>: législation participative

## Autres compétences de l'État

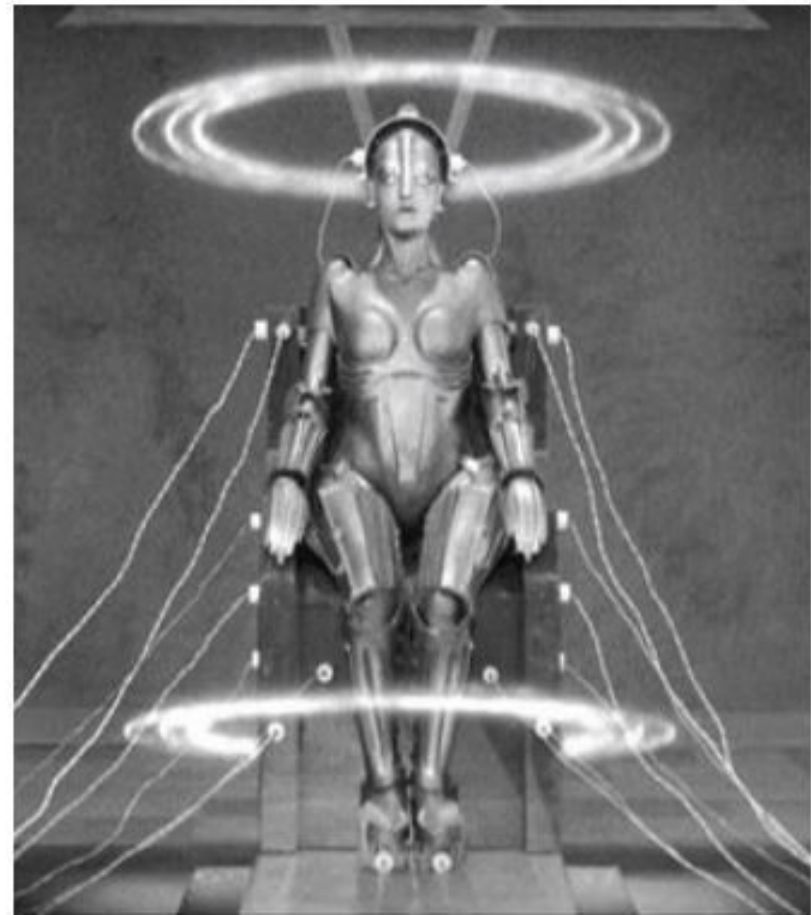
- Santé
- Instruction
- ...





# Enjeux éthiques

- Protection vie privée
- Responsabilité
- Liberté de décision et d'action
- Transparence
- Caractère mutualiste des assurances
- Fraternité
- ...





**Intelligence Artificielle, Russell & Norvig, 2010, Chapitre 26**  
**Fondements philosophiques**

**26.3 Éthique et risques du développement de l'intelligence artificielle.**

- Les gens pourraient perdre leurs emplois au profit de l'automatisation.
- Les gens pourraient avoir trop (ou pas assez) de temps libre.
- L'utilisation de systèmes d'IA pourrait entraîner une perte de responsabilité.
- Les systèmes d'IA pourraient être utilisés à des fins indésirables.
- Le succès de l'IA pourrait signifier la fin de l'espèce humaine.



# Isaac Asimov (1942) :

## Les trois lois de la robotique:

- Un robot ne peut porter atteinte à un être humain, ni, restant passif, permettre qu'un être humain soit exposé au danger.
- Un robot doit obéir aux ordres que lui donne un être humain, sauf si de tels ordres entrent en conflit avec la première loi.
- Un robot doit protéger son existence, tant que cette protection n'entre pas en conflit avec la première ou la deuxième.



# CONCLUSION

IA Faible  
contre  
IA Forte



# Conclusion

La Singularité  
technologique est le  
nouvel opium des  
peuples!

Comment faire pour  
que la peur ne  
masque pas les  
menaces et les  
dangers?

