



intelligence artificielle & éthique

... une drôle de rencontre ...

Jean-Pierre Elloy

| pourquoi se poser la question de l'éthique ?

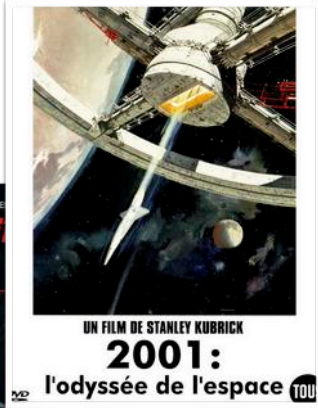
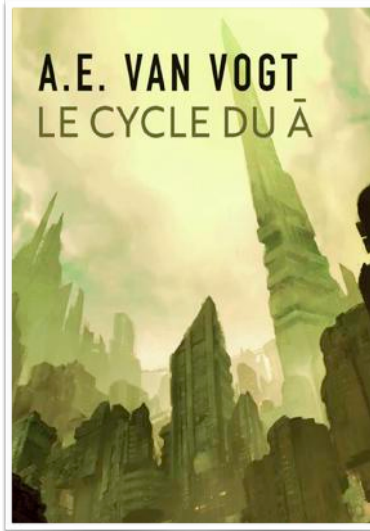
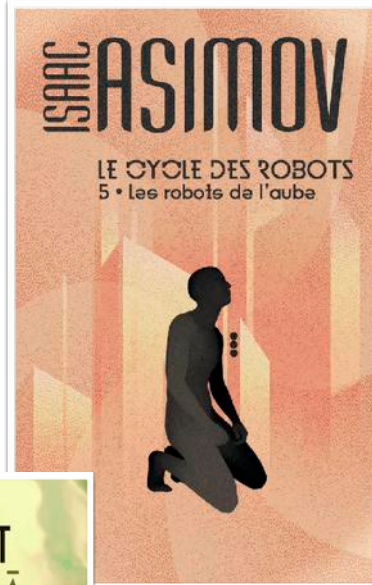
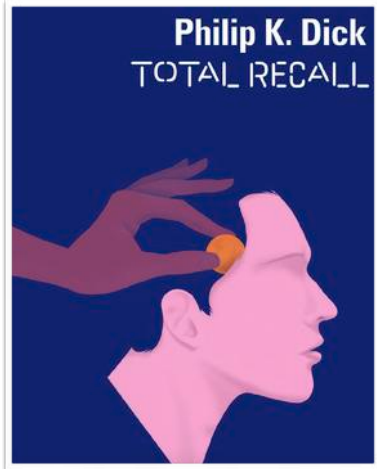
c'est par ses **applications** que l'IA a fait surgir cette question

.....

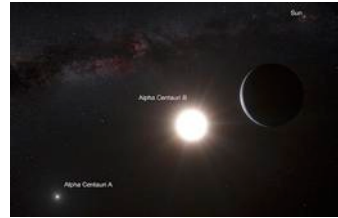
et ce **dans l'histoire** avant même
que l'intelligence artificielle n'existe

L'histoire a initié les émotions sur l'intelligence artificielle

- peur
- angoisse
- espoir



Problème des 3 corps



Lois de Newton sur 2 corps :

- **attraction réciproque** : Si un objet A exerce une force sur un objet B, alors l'objet B exerce une force de même valeur et de sens opposé sur l'objet A
- **de la gravitation** : Deux corps ponctuels de masses M_A et M_B s'attirent avec des forces opposées et de même valeur :

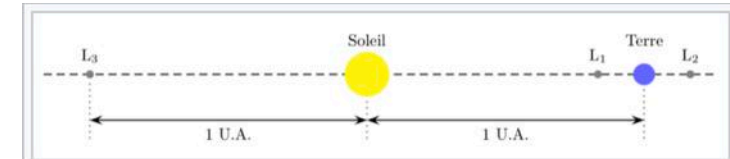
$$F = G \frac{M_A \times M_B}{d^2}$$

Attraction de 3 corps :

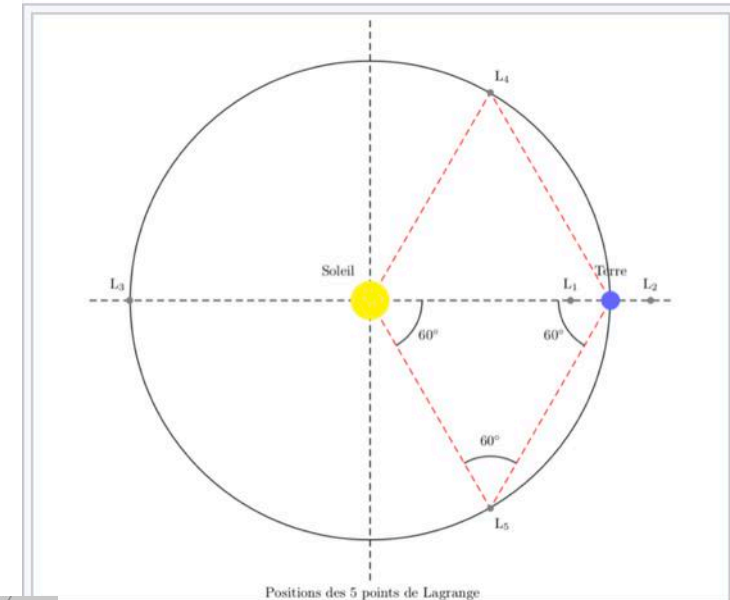
- pas de solution générale connue
- si le 3^e corps est de masse négligeable? Lagrange a déterminé les points où les attractions des 2 autres corps s'équilibrent sur lui > **points de Lagrange**

Points de Lagrange

- **point de Lagrange L1 du système terre-soleil** : instable (utilisé par les satellites d'observation du soleil)
- **point de Lagrange L2 du système terre-soleil** : instable (utilisé par les satellites observatoires, comme James Webb)
- **point de Lagrange L2 du système terre-lune** (utilisé par le satellite Queqiao, transmission des données de la face cachée)
- **point de Lagrange L4 et L5 du système terre-lune** : stables.

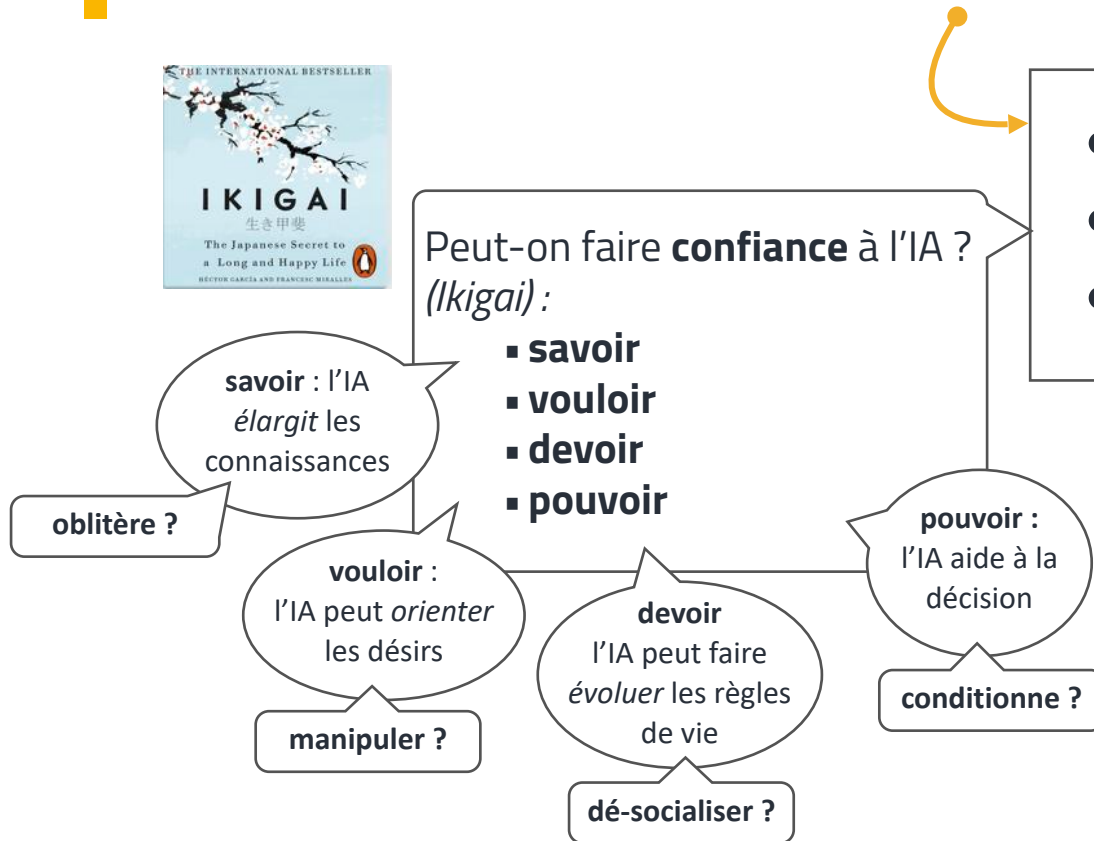


Trois des points de Lagrange sont situés sur l'axe reliant les deux corps. Dans le cas d'une grande dissymétrie de masse entre ceux-ci, deux points sont situés proches et de part et d'autre du corps peu massif, alors que le troisième est quasiment situé à l'opposé du corps peu massif par rapport au corps massif.



Les deux derniers points de Lagrange forment avec les deux corps des triangles équilatéraux.

L'histoire a initié les émotions sur l'intelligence artificielle



pourquoi se poser la question de l'éthique ?

c'est par ses **applications** que l'IA a fait surgir cette question

.....

quelques exemples **d'aujourd'hui** :

- le véhicule à conduite automatisée
- la parole des agents conversationnels
- la reconnaissance comportementale
- l'engin soldat

Le véhicule à conduite automatisée

- capable de rouler sur route ouverte sans conducteur



Le bénéfice social attendu

- la sécurité routière
- une mobilité accessible
- impact sur l'environnement

à des personnes en incapacité de conduire

- en zone rurale, transports à la demande

95% des accidents sont dus à une erreur humaine

fluidité et contrôle du trafic

L'apport de l'IA

- analyse des alentours du véhicule
- détection obstacle et stratégie d'évitement

- lidar, détection laser, radar, caméra
- pallier les phases d'occultation

piéton, cycliste, autres véhicules

changement de file
accélérer, freiner
laisser passer

grande variabilité entre Paris, Le Caire et Bombay

Les tensions

- vide juridique > responsabilité des véhicules autonomes
- comprendre l'intention des autres acteurs
- vu leur diversité, comment seront traitées les situations non apprises

autocollants sur un panneau, un « stop » sur un sac à dos

conditions météo, luminosité, réflexions

mais pour les stratégies de coopération fine ?

La parole des agents conversationnels

■ (chatbots)

Le bénéfice social attendu

• SendPulse

- accès à des services
- aide à l'échange (enfants malades)
- *deadbots* ?

• HereAfter AI
• MBC Design Center
• Open AI

d'actualité,
de commande des
repas, réserver,
emploi du temps

L'apport de l'IA

- acceptation de variations vernaculaires
- propositions adaptées aux préférences

accent,
interjection,
expressions, qualité
du vocabulaire

le langage
caractérise les spécificités
culturelles, incluant les
perceptions du monde



Les tensions

- les biais des données d'entrée
- l'influence du langage chatbot sur le langage humain
- l'impact cognitif sur l'humain en fragilité

• en 2016, Tay de Microsoft a généré très vite des propos racistes et misogynes

conversation
« réaliste »

frontière
brouillée, ange
gardien

• un chatbot peut-il mentir ?

• ses erreurs sont-elles acceptables ?

La reconnaissance **comportementale**

- pour aide aux personnes en difficulté

Le bénéfice social attendu

En 2019, une maison de retraite de la ville de Sakai [dans la préfecture d'Osaka] a investi dans des robots d'assistance à la surveillance. Ces robots avertissent les aides-soignants de la nécessité d'une intervention via un terminal portable.



L'apport de l'IA

Les robots sont pourvus d'un système de détection des mouvements, et peuvent évaluer les risques de chute des résidents en fonction de leur comportement.

Détecteurs à
infrarouge, capteurs de
vibration, accéléromètres

Les tensions

Seule la silhouette est visible sur l'écran du personnel, de manière à respecter l'intimité des pensionnaires

- susciter une confiance sans réserve
- un robot peut être un espion invisible



« risques
capitaux » dans la
relation homme-
machine

Les 3 tensions à l'usage de robots

- induire la confusion entre homme et machine
- faire croire qu'un robot est capable d'émotions
- projeter sur le robot l'image un idéal de perfection

• simulation d'empathie

• inhumaine, donc qui exclut

www.ec-nantes.fr

REINVENT ENGINEERING

L'armée des engins soldats

- en capacité d'ouvrir le feu en mode autonome

Le bénéfice militaire attendu

- l'efficacité par la réactivité
- assister les décisions tactiques
- limiter l'exposition des troupes

sentinelles
autonomes de la
Corée du Sud

tourelle
anti-torpille Palanx des
navires américains

Dôme
de Fer israélien : choix
du missile par l'anti-
missile

L'apport de l'IA

Les six degrés d'autonomie d'un système armé

L0 : Un drone (terrestre, naval, aérien) téléopéré.

L1 : Le drone est connecté à un soldat et duplique ses actions. Il peut voir et viser la cible repérée par le soldat et tirer quand celui tire.

L2 : Le robot se déplace seul, détecte des cibles en fonction d'instructions préalables ; le tir a lieu sur décision humaine.

L3 : Le robot propose des cibles identifiées comme hostiles ; engagement et tir sur autorisation humaine.

L4 : Le robot engage une cible et tire sans autorisation humaine mais il peut être repris en main avant.

L5 : Autonomie totale, reprise en main impossible. C'est le robot type "Terminator".



Kargu-2,
drone suicide turc qui aurait
attaqué un convoi en Syrie sans
ordre direct



Spot,
robot-chien de Boston
Dynamics en test par l'armée
française

Les tensions

- déléguer la responsabilité
- déléguer le risque

La position des USA est
qu'il ne serait pas « éthique »
d'envoyer des soldats face à des
unités robotisées

Scaf :
Système aérien
de combat du
futur



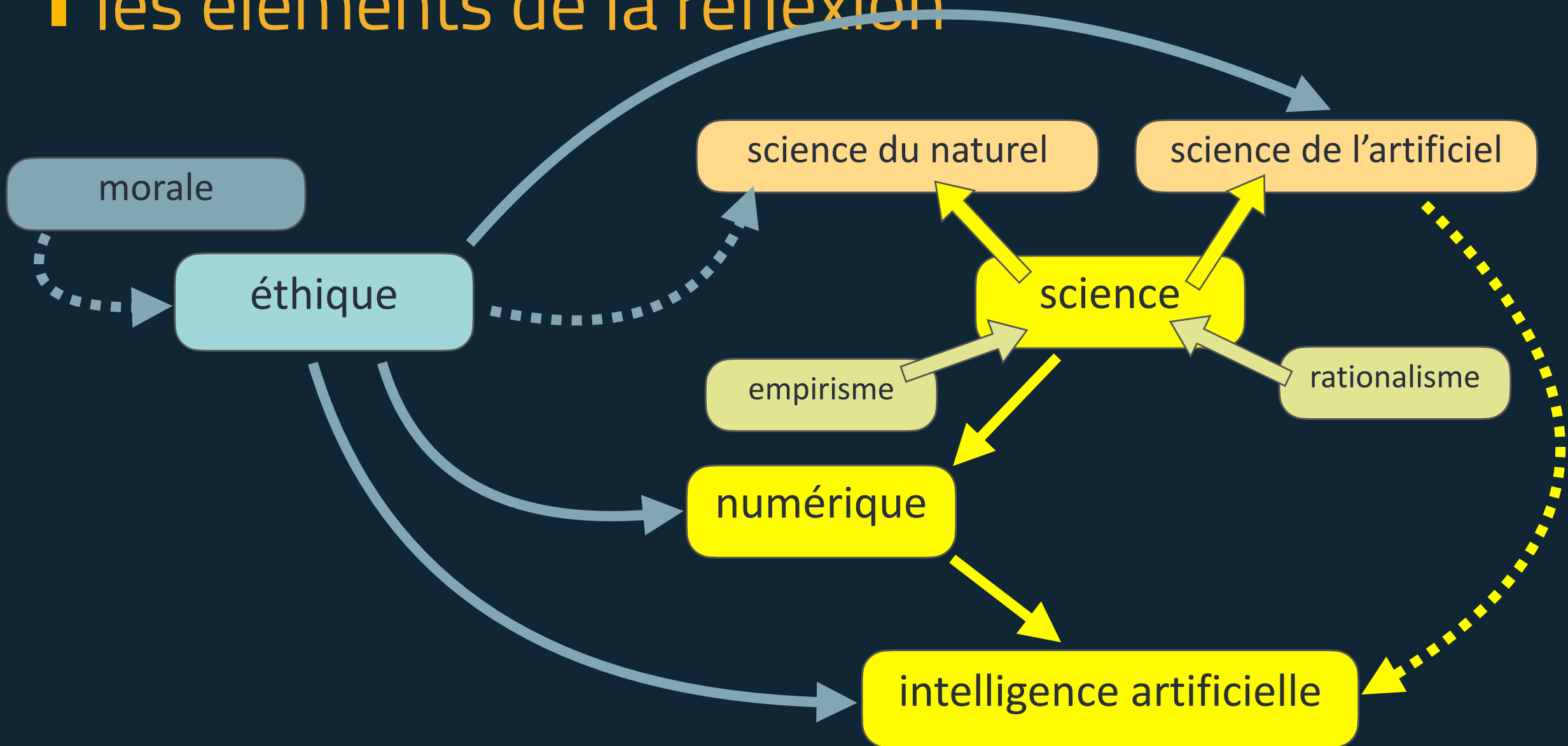
démarche de réflexion

D'un côté **l'éthique**, de l'autre **l'intelligence artificielle**
> la **victime** et **l'assassin** ?

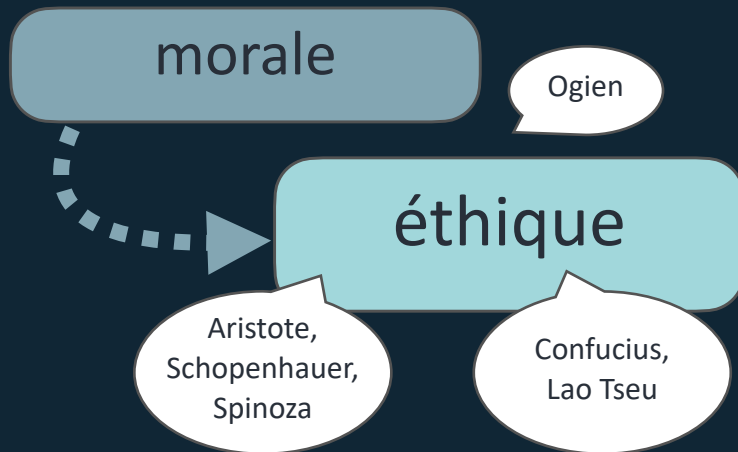
Qui sont-ils ?

- Qu'est-ce que l'éthique ?
- Qu'est-ce qu'est l'intelligence artificielle ?

les éléments de la réflexion



réflexion sur l'éthique



Principes éthiques occidentaux

Aristote

- 384 > 322 avant notre ère
- penseur occidental le plus complet
- **Ethique à Nicomaque**



L'éthique est un **champ de la science** qui doit permettre à l'homme de **vivre une vie meilleure**

Une éthique de soi

- La vertu éthique est toujours un **équilibre** entre deux excès
- L'éthique doit toujours être en recherche du **Bien** (*épanouissement*)
- Le but d'Aristote n'est pas de savoir ce qu'est la vertu éthique, mais de montrer **comment être vertueux**
- Il distingue les vertus **intellectuelles** (*produits de la formation reçue*) et les vertus **morales** (*produits de l'habitude*)

Spinoza

- 1632 > 1677
- cartésianisme (Leibniz et Descartes)
- **L'Éthique selon l'ordre géométrique**



Spinoza introduit les sciences de l'homme

- L'intelligence est le fondement de la vie, l'homme est de comprendre avec son intelligence
- Les hommes poursuivent des fins qui leur sont utiles, **comme "bien" ce qui leur est utile**
- Quel **type de connaissance humaine** mérite notre confiance ?

Pour Spinoza, trois méthodes pour obtenir des **connaissances universelles**

- par opinion et ouï-dire (**empirisme**)
- par raisonnement (**rationalisme**)
- par science intuitive (**essence des choses**)

Schopenhauer

- 1788 > 1860
- un des plus grands "influenceurs"
- **Le fondement de la morale**



Notre époque a normalisé un "**bonheur**" qui se réduit à un **plaisir illusoire**. Comme la propagande mesure ce bonheur à l'aune du succès et de la compétition, l'être humain retombe encore et toujours dans la souffrance de **l'insatisfaction**, qui nous pousse à remettre en cause ce qui est pourtant suffisant et juste. La prétention du "**savoir-être**" affichée sur les réseaux sociaux **a banni les sentiments négatifs**, donc ceux qui rappellent la douleur, la nostalgie et l'absurdité de l'existence. Ce refus **déstructure**.

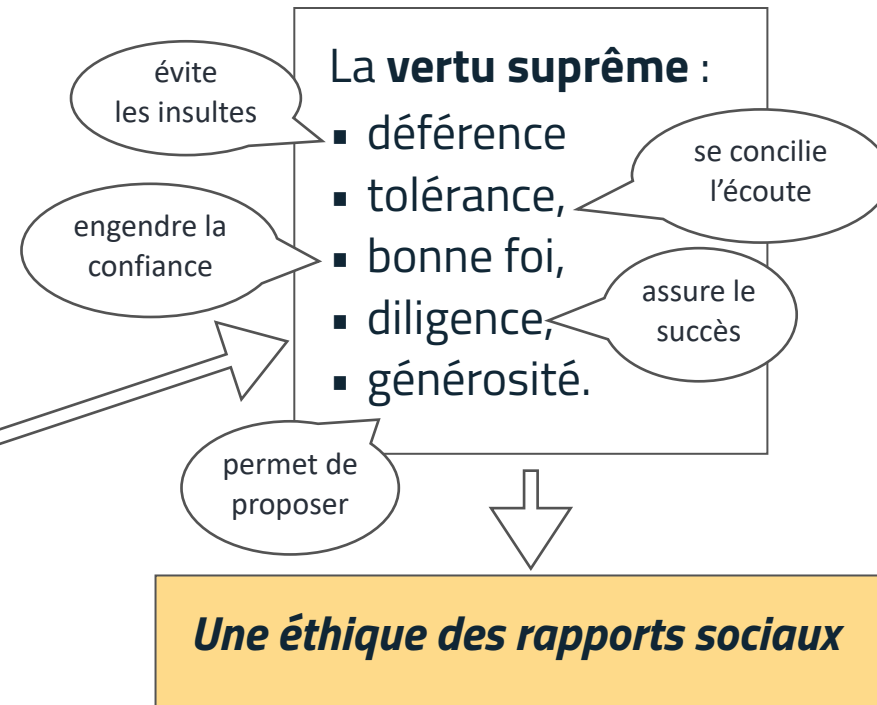
Principes éthiques d'ailleurs

Confucius

- 551 > 479 avant notre ère
- concilie des principes moraux et les faiblesses des humains
- Les **Entretiens** constituent les enseignements de Confucius
- ils ont été rassemblés pendant 2 générations de disciples

Lao Tseu

- contemporain de Confucius ?
- père du Taoïsme
- Le **Livre de la voie et de la vertu** son texte majeur 道德经



L'homme d'une **vertu supérieure** :

- * S'il fait des largesses, il excelle à montrer de l'humanité.
- * S'il parle, il excelle à pratiquer la vérité.
- * S'il gouverne, il excelle à procurer la paix.
- * S'il agit, il excelle à montrer de la capacité.
- * S'il se meut, il excelle à se conformer aux temps.

Une éthique individuelle en société

L'éthique aujourd'hui

éthique
sociale générale

Ethique et morale

- **éthique** > origine ἦθος = mœurs
- **morale** > origine *mores* = mœurs

Différentes interprétations :

- **éthique** > principes *universels*
- **morale** > principes *contextuels*
- **éthique** > rapports de soi *à soi*
- **morale** > rapports de soi *aux autres*

Les deux se confondent souvent



Ethique des anciens

Ethique maximaliste

> rechercher une vie **vertueuse**

- **Aristote** > tout un art de vivre, y compris en société
- **Kant** > devoirs moraux envers autrui et **soi-même**

- gâcher un don
- le suicide
- perdre son temps
- la toxicomanie

Ethique minimaliste

> se réduit à "ne pas nuire à **autrui**"

- **John Stuart Mill** > pas de crime sans *victime*

- autre que soi-même
- entités abstraites
- sans contrainte

Ethique moderne

Si l'éthique des anciens est celle des vertus à avoir soi-même, l'éthique moderne la formule en **devoirs** (vis-à-vis des autres) ... même si c'est au détriment de soi-même ?

- **la non-nuisance**
- **l'impartialité**
- **l'égale considération**
- **la justice**

qui diffère du Bien

- ne pas confondre avec bienveillance



Reconnaissance **faciale**

déverrouillage
smartphone

CocaCola récompense
ceux qui recyclent

MacDonald évalue le
sourire des employés



Robot **juge** et justice **prédictive**

expérimenté en
Estonie

Compas :de
Northpointe

Case Law
Analytics

Justicial Tools de
Equivant



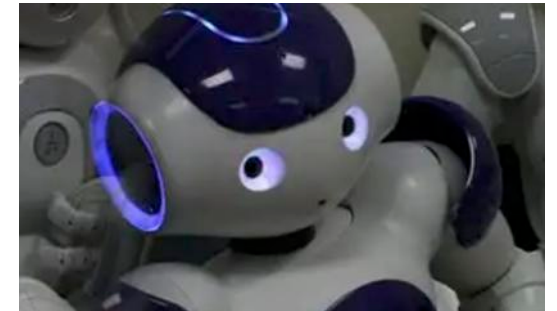
Robot **de compagnie** des enfants malades

NAO et les
enfants autistes

CASPER et les enfants
atteints du cancer

CASPER et les enfants
atteints du cancer

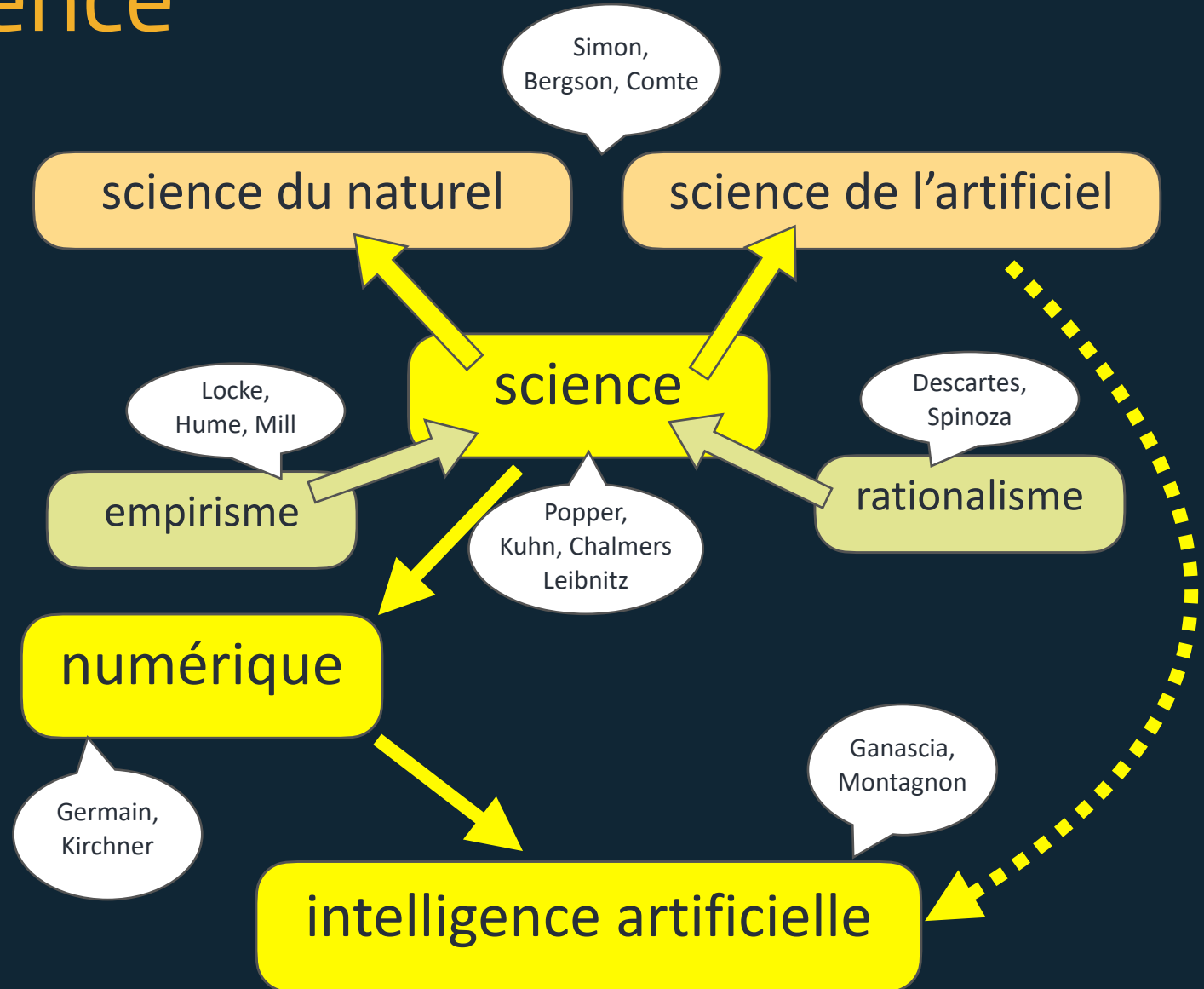
WHEEZHOPE et les enfants
atteints de cardiopathie



Dépasser la **mort** ... par les deadbots ?



réflexion sur la science



réflexion sur la science



empirisme rationalisme

- Il n'existe pas la moindre méthode qui permet de **prouver** qu'une théorie scientifique est **vraie**

- L'empirisme se réfère à une démarche **inductive** qui dérive une **loi générale** à partir **d'observations**

une observation est un énoncé **singulier**

une loi est un énoncé **général** qui porte sur toutes les observations

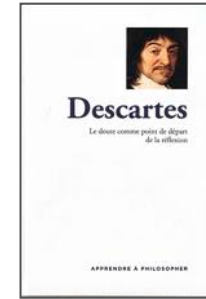
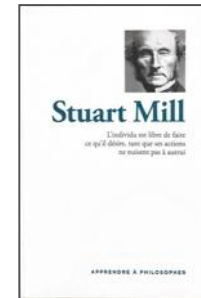
elle est voilée en chacun de nous ; à nous de la découvrir

- Le rationalisme s'appuie sur une doctrine : la **vérité a un caractère manifeste** (d'évidence)
- Pour la trouver, il faut pratiquer le **doute** systématique.

c'est-à-dire qu'on ne pourra **jamais** faire une observation ou un calcul qui **l'infirm**e

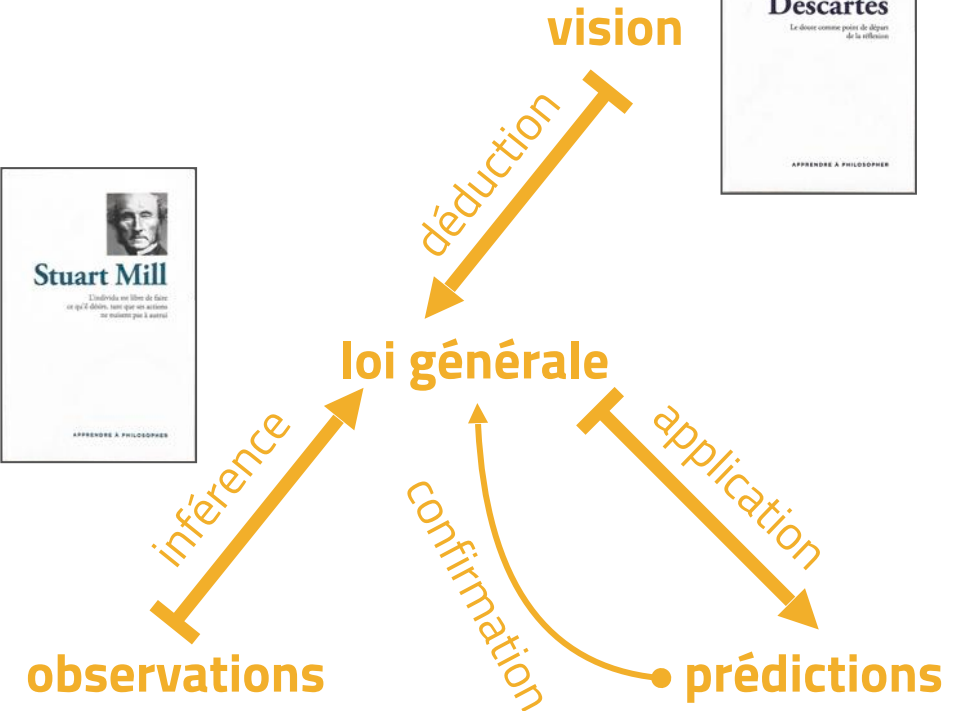
Empirisme :

- Locke
- Hume
- Stuart Mill



Rationalisme :

- Descartes
- Spinoza
- Leibniz



Deux conclusions majeures :

- Pour faire partie de la science tout énoncé doit être **falsifiable**
- Ni observation, ni raison ne peut, à elle seule, être source de la connaissance**

une loi reste vraie jusqu'à ce qu'une observation **l'invalide**

comment appliquer cette règle à un système d'apprentissage ?

l'IA bouscule cette affirmation



de la Mécanique

empirisme

Tout système mécanique est caractérisé par une fonction de **Lagrange** de ses coordonnées q

$$L(q, \dot{q}, t)$$

Principe de **moindre action**.

- Entre deux instants t_1 et t_2 , tout système se meut naturellement de façon à ce que :

$$\int_{t_1}^{t_2} L(q, \dot{q}, t) \cdot dt \text{ ait la plus petite valeur}$$

- à 1 degré de liberté, en effectuant une petite variation δq de q , on obtient les **équations de Lagrange**, qui sont les équations de mouvement du système :

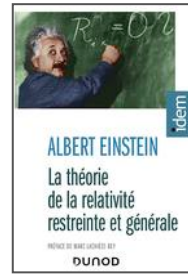
$$\frac{d}{dt} \left(\frac{\delta L}{\delta \dot{q}} \right) - \frac{\delta L}{\delta q} = 0$$

Principe de **l'inertie**

- pour tout corps libre, il existe un référentiel (appelé galiléen) dans lequel le mouvement de ce corps répond à **la 1ère loi de Newton** qui est :

$$\frac{d}{dt} \left(\frac{dL}{d\dot{q}} \right) = 0 \implies \dot{q} = cst$$

rationalisme



.... à la Relativité restreinte

rationalisme

- en mécanique newtonienne, les vitesses du corps **C** dans les deux repères sont liées par :

$$v = v' + u$$

rationalisme

- or Lorentz a montré que la vitesse de la lumière est constante dans toutes les directions et qu'aucune vitesse ne peut la dépasser

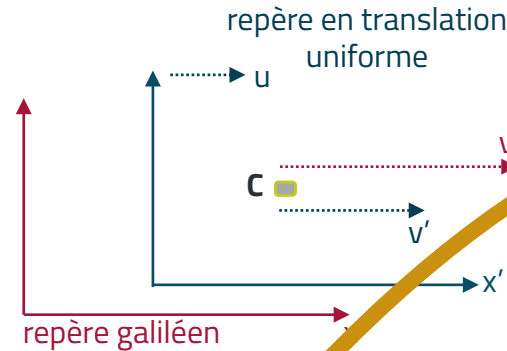
- mais le problème du miroir semble le contredire

empirisme

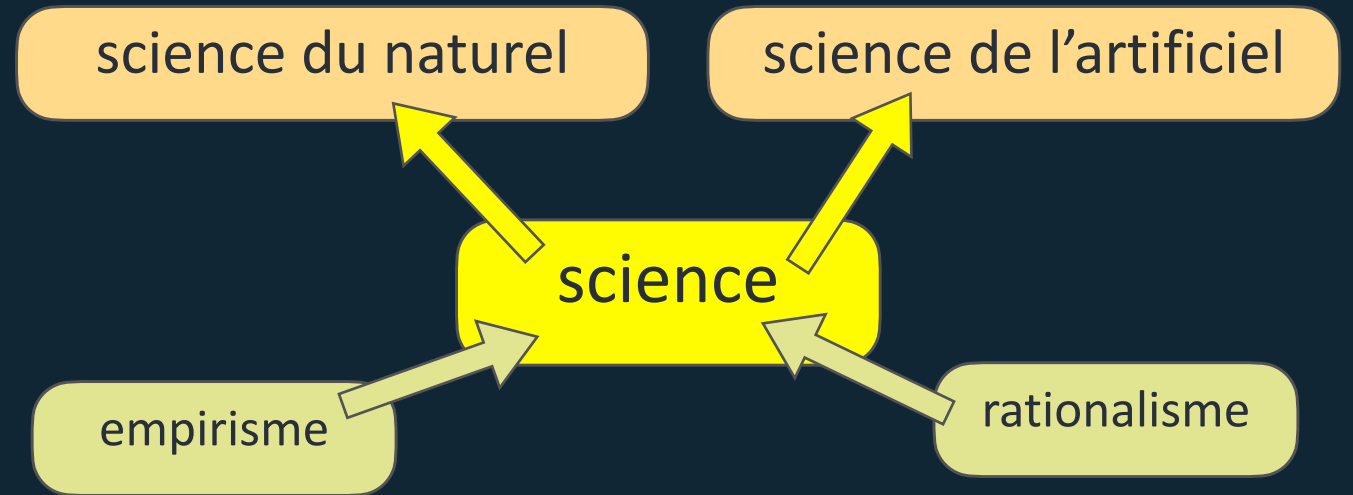
rationalisme

- sauf que, en mécanique relativiste, la vitesse du corps **C** dans le repère galiléen s'écrit (**transformation de Lorentz**) :

$$v = \frac{v' + u}{1 + \frac{v' \times u}{c^2}}$$



réflexion sur la science



science naturelle

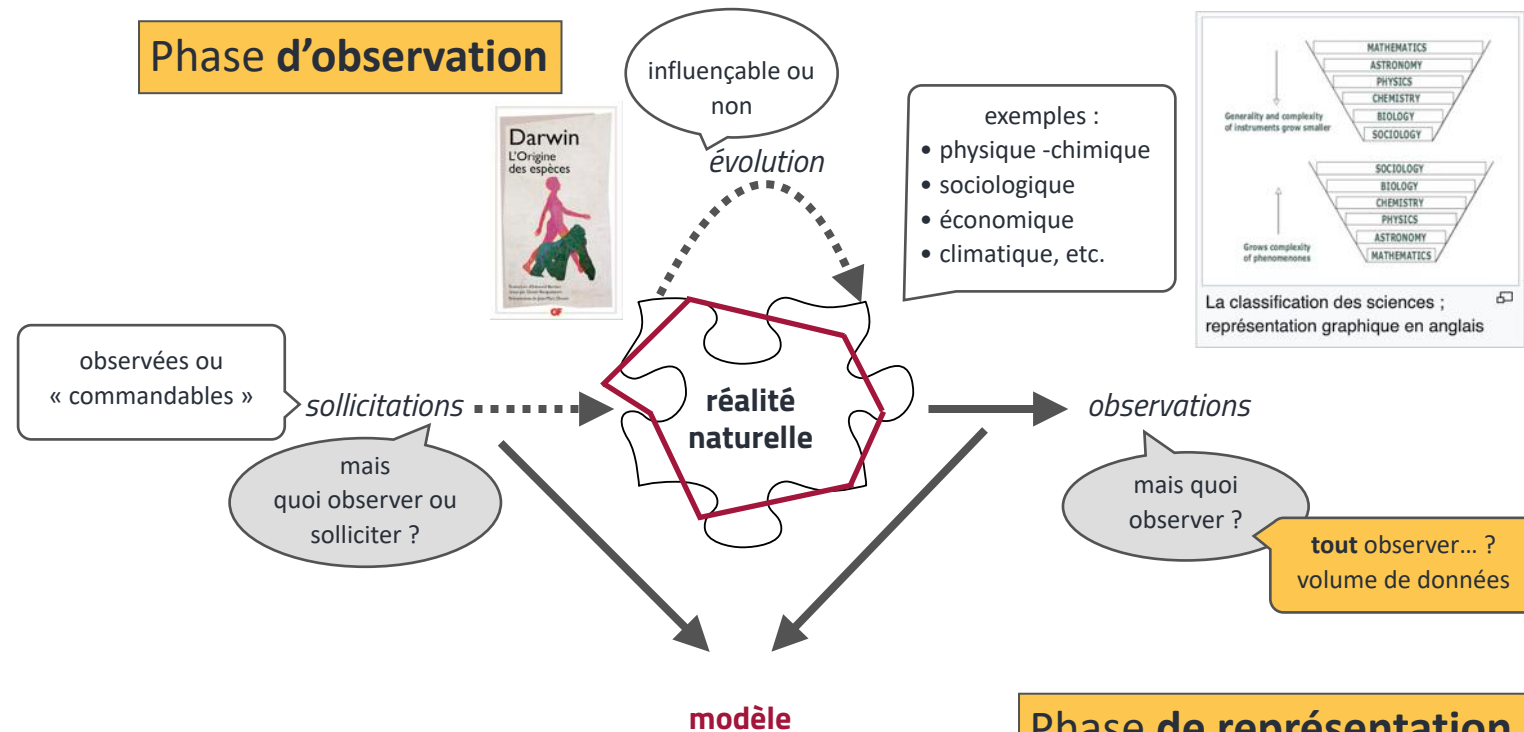
Une **science naturelle** est un corpus de connaissances sur une **réalité naturelle** constitutive de notre monde, objets ou phénomènes (physique, biologique, etc).

Ces connaissances portent sur :

- les **caractéristiques** de cette réalité
- son **fonctionnement** interne
- son **comportement** externe, ses interactions



Phase d'observation



Phase de représentation

Phase d'exploitation

dépend de l'objectif

*compréhension (accroissement pur des connaissances)
reproduction (bio-mimétisme) & amélioration
simulation (réaction à des sollicitations spécifiques)
utilisation & intégration dans un système englobant*

science artificielle



Une **science artificielle** est un corpus de connaissances sur une **réalité qui a été synthétisée** par des êtres humains.

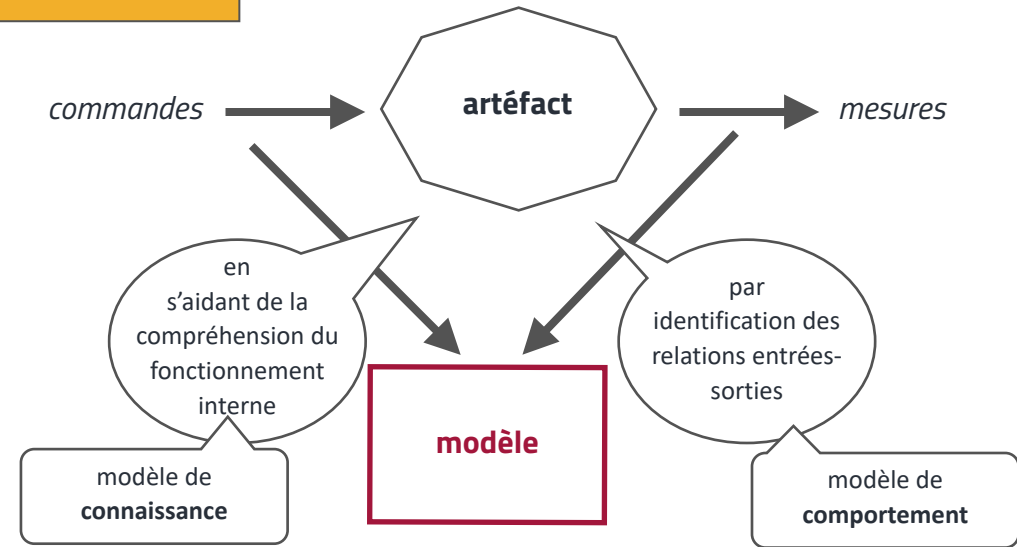
Ces connaissances sont :

- ses **caractéristiques** résultent de sa conception, conditionnée par des critères et contraintes ;
- son **fonctionnement** interne résulte de sa conception
- elle est forgée dans un but précis, donc pour un **comportement** souhaité, donc connu

un artéfact =
objet ou service

c'est le but de la
formation ingénieur : comment
fabriquer des artefacts aux
propriétés souhaités ?

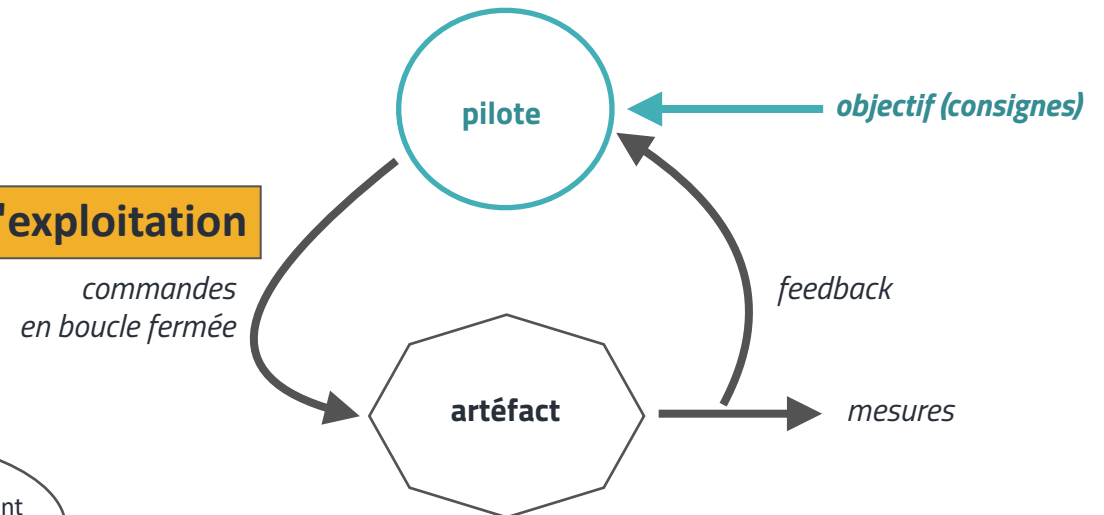
Phase d'identification



performances attendues
de l'artéfact **piloté**

Phase de calcul du pilotage

Phase d'exploitation



phases de intelligence artificielle

Les définitions proposées décrivent ce qu'elle fait, non ce qu'elle est



Une **intelligence artificielle** construit une **décision** à partir d'un *univers des relations* entre toutes les **données** qui conditionnent cette décision.

Cette démarche :

- s'appuie sur l'**identification des données** qui peuvent affecter le comportement de la réalité étudiée
- ne s'appuie **pas** sur une connaissance préalable du **comportement** de la réalité étudiée
- construit, par affinages successifs un **modèle** de la réalité étudiée qui se comporte comme elle

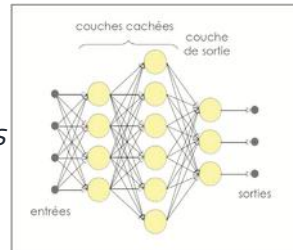
Phase d'apprentissage

données
jeux d'essais



décisions

ajustement automatique
de millions de paramètres



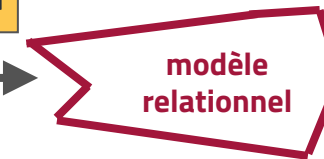
ce n'est pas une
réalité qui est étudiée mais une
représentation par des données
de cette réalité



représentation
implicite

Phase d'exploitation

données
d'exploitation
Ex : symptômes



décision
Ex : Diagnostic

application
de la décision

son usage



une forme de boucle ouverte ?
• sans modèle de connaissance
• sans modèle de comportement

cerveau humain vs intelligence artificielle



- Si l'IA est capable de réaliser des fonctions complexes > *est-elle consciente ?*
- l'IA se comporte comme moi > *consciente ?*

■ Mais les théories neuro-scientifiques de « **structure causale** » concluent que la *réplication* n'est pas un fonctionnement suffisant pour générer une conscience

■ Pour d'autres, la *conscience* étant une notion subjective (pas de théorie sous-jacente), il n'est pas possible de conclure sur l'IA

L'efficacité du cerveau reste inégalée

- dans sa **rapidité** d'apprentissage
 - l'IA de l'entreprise DeeMind a dû accumuler 900h de jeu pour atteindre un niveau acquis en 2h par un humain
- l'apprentissage humain est **social**
 - c'est-à-dire se transmet : *nous sommes des « homo docens »*
 - la transmission est au fondement de la socialisation
- l'apprentissage humain est **long**
 - la plasticité cérébrale dure 20 ans : *le cerveau se modèle*
 - cette plasticité est stimulée par l'éducation
- le cerveau humain a la faculté **d'abstraction**

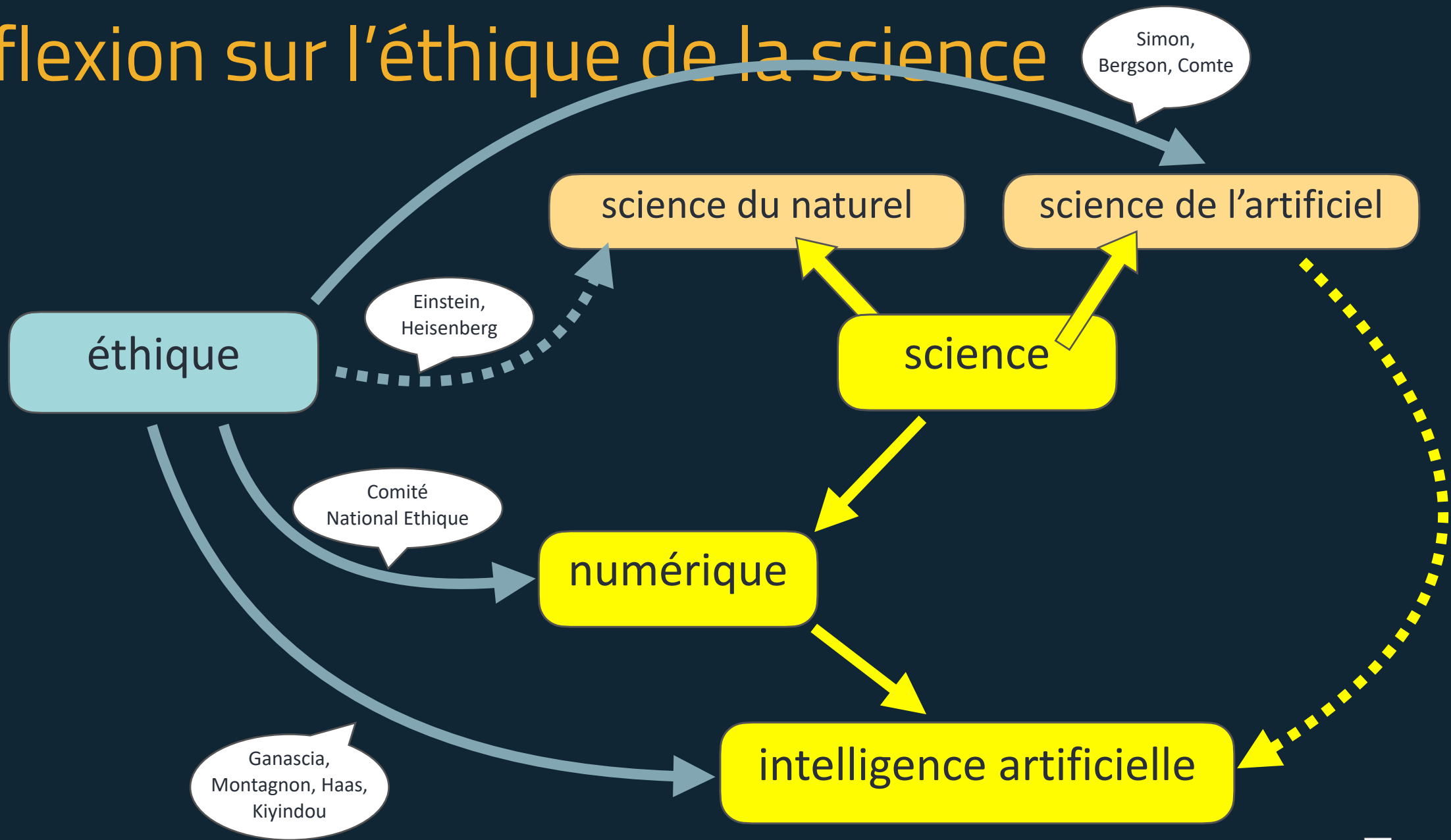
- l'intelligence a la capacité de traiter des problèmes nouveaux
- l'intelligence humaine a accès à l'émotion
- il n'y a pas d'intelligence humaine sans échange continu entre le cerveau et le corps

intelligence artificielle = intelligence auxiliaire ?

- *La modélisation*
- *La conceptualisation*
- *L'innovation de rupture*



réflexion sur l'éthique de la science



avant de parler de ce qu'est « l'éthique » de la science

distinguer tout de suite :

- Éthique **scientifique** ... du « **chercheur** »
probité, honnêteté, sincérité, transparence

toute
personne qui contribue
aux méthodes ou aux
applications

ou ingénieur
ou développeur

vis-vis de
lui-même (éthique **humaniste**)
vis-à-vis de son employeur (éthique
professionnelle)

- Ethique de la **science** ... dans sa mission

vis-
vis de la société
(éthique **sociologique**)

- Ethique du **numérique**

en tant que
messenger de la science, ou comme
science elle-même

- Ethique de **l'intelligence artificielle**
.... une éthique particulière ?

en tant que
particularisme du numérique ou en
raison de ses **applications**

une éthique de la science

- Sa vocation est celle de la **connaissance** des lois de la nature.
- Pourquoi mettre des freins ou des limites à cette recherche ?

En science naturelle

- Une éthique se justifie-t-elle ?

ne pas
confondre cette recherche
avec son usage

- le numérique n'est pas destiné à faire le bien, mais pour **répondre à un besoin concret**
- le numérique traite de **l'information**

une éthique du numérique

- la bio-éthique énonce qu'il s'agit **nécessairement** de faire le bien

- Sa vocation est celle de la **conception-réalisation** d'artéfacts

En science de l'artificiel

- Pionnière en ce domaine, l'éthique du soin, du début à la fin de la vie : la **bio-éthique**

La **bio-éthique** :

- la bienfaisance
- non-malfaisance
- l'autonomie
- la justice

mais qui
définit le
"bien" ?

une éthique des données

Différentes perceptions des données

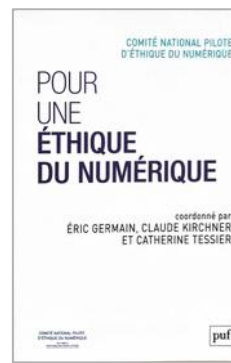
- aux USA, la donnée est un produit commercial
- en Asie, la donnée est un élément de pouvoir
- en Europe, la donnée est sacralisée et protégée des convoitises



RGPD

Règlement général sur la protection des données

- France : le 25 mai 2018
- ~ 100 pays disposent d'un cadre similaire



une éthique du traitement des données

Groupe d'experts Commission européenne HLEG

• donc qui peut être discuté, accepté ou refusé

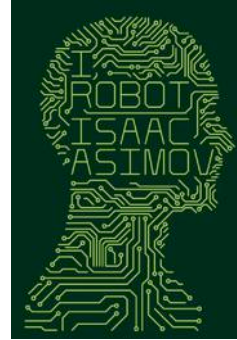
• qui permet compréhension et confiance

• addiction et isolement

- **contrôle humain** (*respect de l'autonomie humaine*)
- la **conformité** (au CdC), la **robustesse** technique (*qualité*) et la sécurité (*résistance aux intrusions*)
- protection de la **vie privée** dans l'exploitation des données (*RGPD*)
- la **loyauté** et **transparence** des algorithmes (*traçabilité, explicabilité, communication*)
- **justice, équité** et **non-discrimination** des résultats
- **bien-être sociétal utilisateur** et **environnemental**
- **responsabilité** et **redevabilité**

intelligence artificielle

une éthique des données



Différentes perceptions des données

- aux USA, la donnée est un produit commercial
- en Asie, la donnée est un élément de pouvoir
- en Europe, la donnée est sacralisée et protégée des convoitises



RGPD

Règlement général sur la protection des données

- France : le 25 mai 2018
- ~ 100 pays disposent d'un cadre similaire



une éthique **limitée** du traitement des données

Gestion des risques

Les 3 lois de la robotique

- **Loi 1** : Un robot ne peut pas porter atteinte à un **être humain** ni, en restant passif, permettre qu'un être humain soit exposé à un danger
- **Loi 2** : Un robot doit obéir aux **ordres** donnés par un être humain, *sauf si contradiction avec la Loi 1*
- **Loi 3** : Un robot doit protéger son **existence**, *sauf si contradiction avec la Loi 1 ou la Loi 2*

une éthique **spécifique** des **systèmes d'apprentissage**

Commission de réflexion **CERNA**

Commission de réflexion sur
l'Éthique de la Recherche en
sciences et technologies du
Numérique d'Allistene

6 champs de préconisation

Spécificités de l'apprentissage machine

- difficulté de la **spécification** (... *puisque précisément, l'IA s'adresse aux problèmes qu'on ne sait pas spécifier*)
- entre le concepteur et l'utilisateur apparaît un « **tiers** », celui qui entraîne avec des jeux de données
- l'impossible **explication** formelle de ses résultats
- si l'apprentissage se poursuit après déploiement en utilisation, impossibilité de prévoir son **évolution**
- des **instabilités** de décision peuvent apparaître dans les apprentissages très profonds
- une **évaluation** n'est possible qu'en utilisation

1 - Sur les données d'apprentissage

- qualité (*absence de biais*)
- diversité et non-discrimination
- traçabilité (*à prévoir à la conception*)

2 - Sur l'autonomie des systèmes d'apprentissage

- visibilité sur l'état courant du système
- vigilance dans la communication sur l'évolution

capacité à évoluer
indépendamment d'un
opérateur humain

3 - Sur l'explicabilité de l'apprentissage et son évaluation

- un compromis entre performance et explication, à expliciter
- élaboration d'un protocole d'évaluation avec les usagers

4 - Sur les décisions délivrées par le système

- établir le rôle d'arbitre de l'humain en phase d'apprentissage
- faire appel à des experts pour vérification

5 - Sur le consentement à l'apprentissage machine

- laisser à l'utilisateur le choix d'activer ou non l'apprentissage
- dès la conception, avis d'utilisateurs qui seront impactés
- consentement par l'utilisateur d'utiliser l'apprentissage en continu

6 - Sur la responsabilité homme vs machine apprenante

- inclure des mécanismes de contrôle, automatiques ou supervisés sur les données, et sur le raisonnement suivi
- documenter de façon sincère et loyale

problèmes éthiques propres à l'intelligence artificielle

- adopter de nouvelles règles d'usage des **données**
 - qui sont en grand nombre, très diversifiées
 - qui, directement ou non, décrivent nos vies
- adapter la protection des droits et des libertés
 - aux **groupes d'individus ...** et non plus aux individus seuls
 - identifier la responsabilité en cas de dérive ou biais
- lever **l'opacité** du fonctionnement de l'algorithme
 - souhaitée par le souci de la PI des entreprises financeurs
- conséquences **sociétales** des champs d'application

En programmation classique, construire une application consiste à programmer un **algorithme déductif spécifié**, c'est à dire **des règles de déduction**, que l'on infère pour traiter chaque cas d'utilisation.

Mais l'apprentissage profond **ne s'appuie pas sur des règles établies par avance**, mais sur des règles qui se construisent d'elles-mêmes au fur et à mesure de l'apprentissage

- Le développement de l'**apprentissage automatisé** s'appuie sur une « mise en données » généralisée en tous domaines
- Tous les aspects de la vie peuvent se retrouver ainsi **fichés** à l'insu de leurs détenteurs

la confidentialité des données médicales a déjà cessé d'exister



- L'IA accroît l'organisation et la structuration de nos **connaissances**, et ouvre ainsi nos possibilités de décision et de contrôle.
- Mais l'IA peut **décider à notre place**

or, on ne peut pas admettre, dans une société, que des **décisions** importantes soient prises sans **explication**

si les décisions ne sont pas justifiées, risque de la crainte d'une reproduction des inégalités sociales dans l'algorithme de décision

une éthique générale du développement de l'IA

1 - Principe de bien-être

- améliorer les conditions de vie, santé et conditions de travail
- permettre d'exercer ses capacités physiques et intellectuelles
- ne doit pas augmenter le stress, l'anxiété, le harcèlement

2 - Principe du respect de l'autonomie

- permettre à chacun de réaliser sa propre conception de la vie
- éviter les dépendances par une confusion entre IA et humains
- ne doit pas mettre en œuvre des mécanismes de surveillance, d'évaluation ou d'incitation contraignants

3 - Principe de protection de l'intimité et la vie privée

- protéger des espaces d'intimité non surveillés
- protéger l'intimité de la pensée et des émotions
- pas de profil personnalisé pour influencer le comportement
- avoir un contrôle étendu sur ses données personnelles

4 - Principe de solidarité

- favoriser les relations humaines et réduire l'isolement
- favoriser le travail collaboratif
- ne pas simuler des comportements cruels par des robots

5 - Principe de participation démocratique

- une IA qui peut affecter la qualité de vie des personnes doit être intelligible et justifiable
- le code des algorithmes doit être accessible aux autorités
- la découverte d'effets non prévus doit être signalée

Déclaration de Montréal pour un développement durable de l'IA

- Élaborer un cadre éthique pour le développement et le déploiement de l'IA ;
- Tous doivent pouvoir bénéficier de cette révolution technologique ;
- Réussir un développement inclusif, équitable et écologiquement soutenable de l'IA.

6 - Principe d'équité

- ne pas créer de discriminations sociales, religieuses, ethniques..
- éliminer les relations de domination fondées sur la richesse, le pouvoir, la connaissance
- bénéficier économiquement à tous

7 - Principe d'inclusion de la diversité

- ne pas induire l'uniformisation ou la normalisation
- respecter les multiples expressions de toutes les diversités
- pour chaque catégorie de service, une offre d'IA diversifiée

8 - Principe de prudence

- restreindre la diffusion d'IA pouvant présenter un danger
- satisfaire des critères de fiabilité, sécurité intégrité... testés
- prévenir des risques d'usage néfaste des données et de l'IA

9 - Principe de responsabilité

- seuls des êtres humains peuvent être tenus pour responsables
- une décision qui affecte la vie doit être prise par une personne

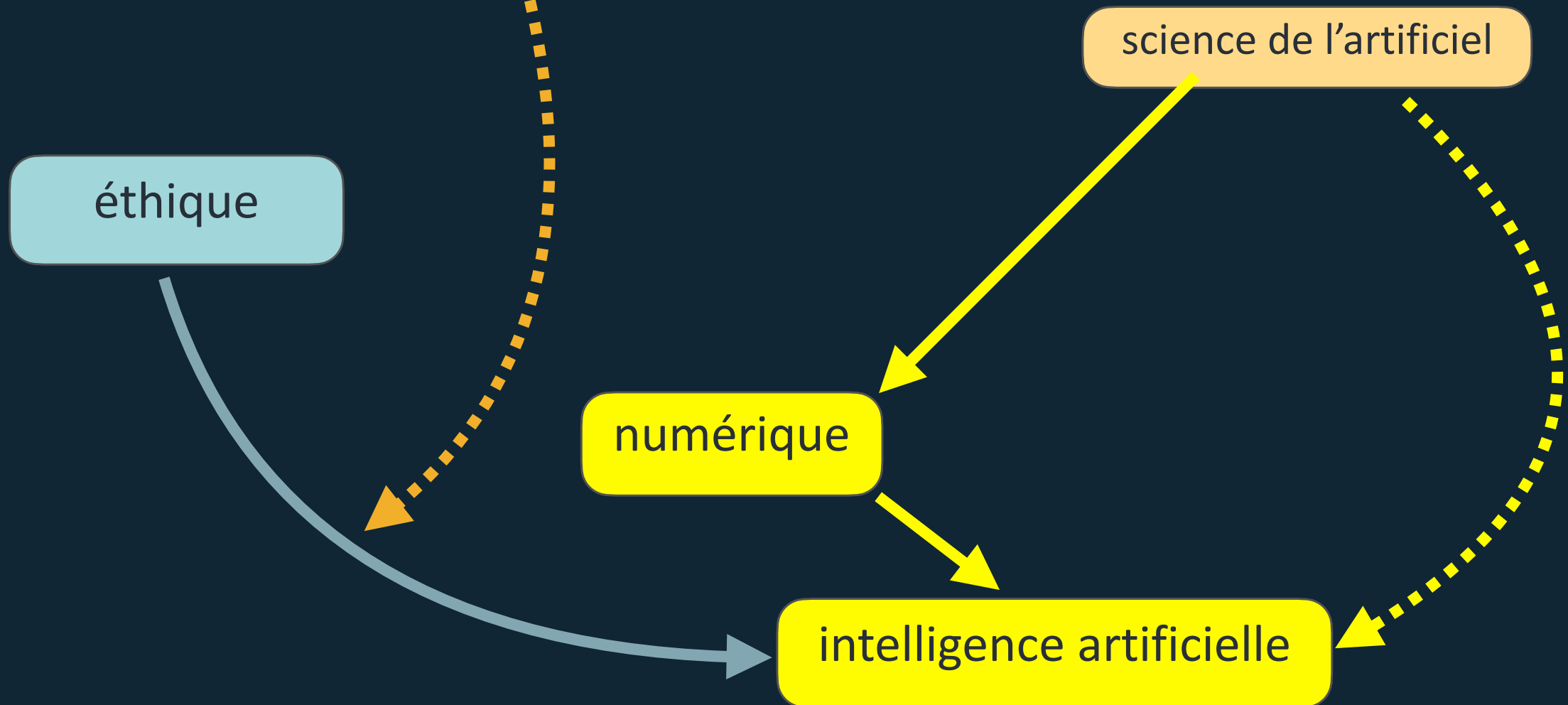
10 - Principe de développement soutenable

- efficacité énergétique des infrastructures des IA
- filière de maintenance, réparation et recyclage
- lutter contre le gaspillage des ressources naturelles

LA DÉCLARATION
DE MONTRÉAL POUR
UN DÉVELOPPEMENT
RESPONSABLE
DE L'INTELLIGENCE
ARTIFICIELLE
2018



éthique de l'ia : à qui ? quand ?





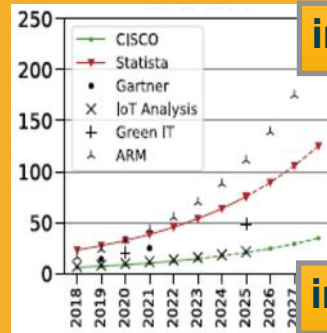
internet des objets (IoT)

Exemples

- Compteur intelligent
- Alarme sociale
- Gestion du trafic routier
- Détection situation à risque
- Surveillance production
- Service clientèle personnalisé
- Anticipation climatique

Nombre en 2020

- France : 0,24 Md
- Japon : 0,9 Md
- Europe : 1,8 Md
- Chine : 3,6 Md
- Monde : entre 18 Md et... 78 Md



incidence éthique

incidence éthique

incidence éthique

qui doit se poser des questions éthiques ?

Acteurs impliqués :

- le maître d'œuvre (la marque)
 - définit la prestation nominale attendue de l'objet connecté
 - définit son ergonomie et son design
- le fabricant
 - partenaire technologique > connectivité
 - partenaire technologique > capteurs
- le prestataire logiciel embarqué
 - pré-traitement des données (filtre, calibrage, mise en forme)
 - logiciel d'application locale (réalisation locale de la prestation)
 - gestion de l'objet (consommation, sûreté, sécurité)
- l'exploitant logiciel distant
 - réception/exploitation des données qualifiées (sémantiquement)
 - logiciels d'application finale > les prestations offertes
- le client final
 - exploitant final direct (le particulier)
 - installateur dans un ensemble d'équipements
 - intégrateur dans un système plus global (nouvelles prestations)

à quelle phase de la réalisation ?

Le cycle en V



les éléments de la réflexion

