

# **INTELLIGENCE ARTIFICIELLE**

# SOMMAIRE

- Les différentes définitions de l'IA
- Types
- Historique
- Domaines d'application
- Éthique

# Introduction

- L'intelligence est souvent associée à la capacité de raisonnement et de réflexion d'une personne.
- Elle s'oppose à l'instinct, qui correspondrait davantage à un réflexe qu'à une pensée élaborée.
- Depuis l'émergence de la robotique et de l'informatique, les chercheurs essaient d'injecter des notions de l'intelligence humaine dans des machines. Étant conçue et fabriquée par l'Homme, on qualifie cette forme d'intelligence d'« artificielle ».

# Première partie

- Les différentes définitions de l'IA
  - Des systèmes qui agissent comme des humains
  - Des systèmes qui pensent comme des humains
  - Des systèmes qui pensent rationnellement
  - Des systèmes qui agissent rationnellement
- Synthèse des approches

# 4 types de définitions

Système qui pense  
comme des **humains**

Système qui pense  
rationnellement

Système qui agit  
comme des **humains**

Système qui agit  
rationnellement

- Deux dimensions :
- Qui est la référence ? Humain ou modèle idéal de rationalité ?
- À l'aune de quoi juge-t-on ? Pensée ou comportement ?

# Types de l'IA

| IA Faible   | IA Forte   |
|---|--|
| <ul style="list-style-type: none"><li>• Les programmes de ce type ne visent pas à évoluer.</li><li>• Un programme ne « pense » pas, il exécute.</li><li>• Très performants dans leur domaine.</li></ul> | <ul style="list-style-type: none"><li>• Approche la plus similaire au comportement humain.</li><li>• Capable d'éprouver une réelle conscience de soi, de ressentir de vrais sentiments et comprendre ce qui la pousse à faire telle ou telle action.</li><li>• Nécessite de savoir apprendre et modifier son comportement.</li></ul> |

# Agir comme les humains : Test de Turing

- Proposé par Alan Turing (1950). Une variante du jeu de l'imitation.
- **Principe :** un humain interagit avec une entité (humain ou machine) par le biais d'un clavier. Un ordinateur réussit le test s'il réussit à tromper le questionneur (30% du temps) et se faire passer pour un humain.

# Agir comme les humains : Test de Turing

- Proposé par Alan Turing (1950). Une variante du jeu de l'imitation.
- **Principe :** un humain interagit avec une entité (humain ou machine) par le biais d'un clavier. Un ordinateur réussit le test s'il réussit à tromper le questionneur (30% du temps) et se faire passer pour un humain.

Nécessite :

- Traitement de la langue naturelle
- Représentation des connaissances
- Raisonnement automatisé
- Apprentissage
- Vision artificielle et robotique pour le test complet.

# Test de Turing

Le test de Turing correspond à une approche pragmatiste : peu importe si la machine raisonne comme un humain, tant qu'elle agit comme tel.

→ Test d'intelligence comportementale.

# Test de Turing

Le test de Turing correspond à une approche pragmatiste : peu importe si la machine raisonne comme un humain, tant qu'elle agit comme tel.

→ Test d'intelligence comportementale.

Prix Loebner ([https://fr.wikipedia.org/wiki/Prix\\_Loebner](https://fr.wikipedia.org/wiki/Prix_Loebner)) : entre 2008 et 2020, compétition internationale récompensant les agents conversationnels satisfaisant le mieux le test de Turing (5 minutes de conversation).

# Test de Turing

Le test de Turing correspond à une approche pragmatiste : peu importe si la machine raisonne comme un humain, tant qu'elle agit comme tel.

→ Test d'intelligence comportementale.

Prix Loebner ([https://fr.wikipedia.org/wiki/Prix\\_Loebner](https://fr.wikipedia.org/wiki/Prix_Loebner)) : entre 2008 et 2020, compétition internationale récompensant les agents conversationnels satisfaisant le mieux le test de Turing (5 minutes de conversation).

Des progrès ont été faits depuis les débuts du prix Loebner, mais aucun agent conversationnel n'est encore infaillible (NB : Misuku, 5 × vainqueur du prix, est désormais embarqué dans le Metavers)

# ChatGPT vs Turing ?

S

J'aimerais savoir si vous êtes capable de passer le test de Turing.



Le test de Turing est un test de la capacité d'une machine à exhiber un comportement intelligent indiscernable de celui d'un être humain. En tant qu'IA basée sur la technologie GPT-3.5, je peux participer à des conversations textuelles et fournir des réponses qui peuvent sembler intelligentes dans de nombreuses situations, mais il est important de noter que le test de Turing est un critère complexe et subjectif pour évaluer l'intelligence artificielle.



Le test de Turing a été proposé par Alan Turing en 1950 et il n'y a pas de consensus clair sur la question de savoir si une machine peut le passer de manière convaincante dans toutes les situations. Cela dépend en grande partie des attentes et des critères spécifiques définis par les personnes qui administrent le test.

Mon but est d'assister et de fournir des informations utiles, mais ma performance peut varier en fonction de la nature des questions et des conversations. Je ne suis pas capable de reproduire une intelligence humaine à tous égards, et il est généralement assez facile de déterminer que je suis une machine si vous creusez suffisamment en posant des questions complexes ou en observant mes réponses de manière critique.

# Les objections au test de Turing

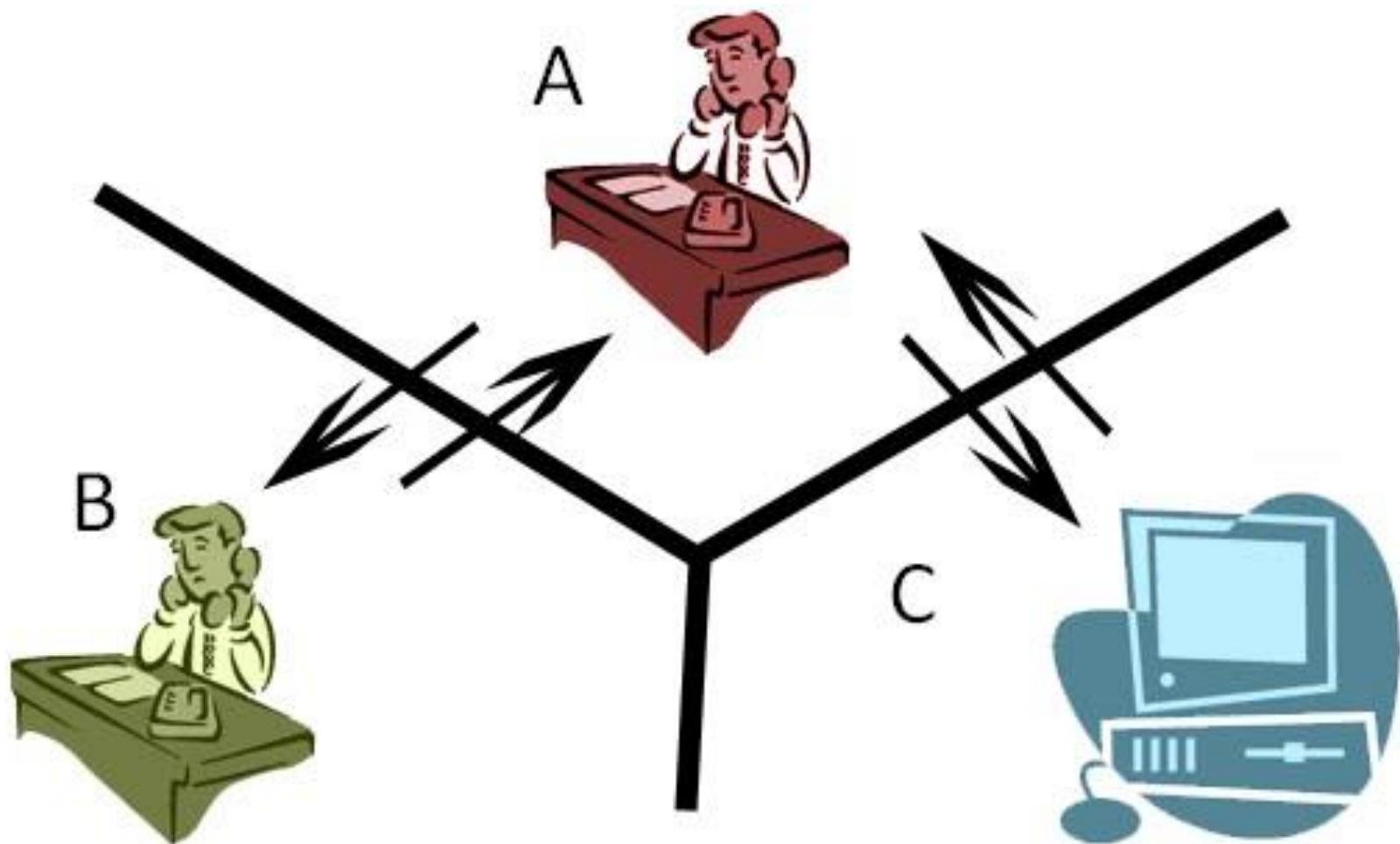
Il existe de nombreuses objections à l'utilisation du test de Turing comme test d'intelligence (certaines formulées par Turing lui-même) :

- Dualité de l'humain et transcendance
- Conscience et expérience de l'émotion
- → Chambre chinoise
- Stupidité artificielle
- ...

13 /  
73

Nous en verrons quelques-unes dans cette présentation (mais pas toutes).

# Test de Turing



# Des systèmes qui pensent comme des humains

Des systèmes qui pensent comme des humains

Des systèmes qui pensent rationnellement

Des systèmes qui agissent comme des humains

Des systèmes qui agissent rationnellement

# Penser comme les humains :

## Approche cognitive

- Requiert des théories scientifiques de l'activité interne du cerveau.
- Implémenter les théories et comparer avec les humains.
- Il faut prédire et tester le comportement de sujets humains (sciences cognitives)
- Les sciences cognitives combinent modèles informatiques de l'IA et techniques expérimentales de psychologie pour élaborer des théories précises et vérifiables du fonctionnement de l'esprit.

# IA forte et conscience

Argument de la **conscience** : « Une machine ne peut pas être dite intelligente si elle n'est pas consciente de ses actes »

# IA forte et conscience

Argument de la **conscience** : « Une machine ne peut pas être dite intelligente si elle n'est pas consciente de ses actes »

Lié à l'expérience de la **chambre chinoise** [Searle, 1980] pour réfuter le test de Turing comme test d'intelligence.

# IA forte et conscience

Argument de la **conscience** : « Une machine ne peut pas être dite intelligente si elle n'est pas consciente de ses actes »

Lié à l'expérience de la **chambre chinoise** [Searle, 1980] pour réfuter le test de Turing comme test d'intelligence.

Pose la question de la nature de la conscience (dualisme ou matérialisme).

# IA forte et conscience

Argument de la **conscience** : « Une machine ne peut pas être dite intelligente si elle n'est pas consciente de ses actes »

Lié à l'expérience de la **chambre chinoise** [Searle, 1980] pour réfuter le test de Turing comme test d'intelligence.

Pose la question de la nature de la conscience (dualisme ou matérialisme).

Turing élude la question en déclarant que cette question disparaîtra d'elle-même lorsque l'on aura un système capable de gagner au jeu de l'imitation (on postule bien que les autres humains ont une conscience...).

# Approche cognitive

**Objectif :** réalisation de programmes imitant dans leur fonctionnement le raisonnement du cerveau humain.

# Approche cognitive

**Objectif** : réalisation de programmes imitant dans leur fonctionnement le raisonnement du cerveau humain.

**Difficulté** : saisir comment pensent les humains.

# Approche cognitive

**Objectif** : réalisation de programmes imitant dans leur fonctionnement le raisonnement du cerveau humain.

**Difficulté** : saisir comment pensent les humains.

- introspection
- expérimentations en psychologie
- imagerie cérébrale

# Approche cognitive

**Objectif** : réalisation de programmes imitant dans leur fonctionnement le raisonnement du cerveau humain.

**Difficulté** : saisir comment pensent les humains.

- introspection
- expérimentations en psychologie
- imagerie cérébrale

→ Approche par nature pluridisciplinaire.

# IA forte vs IA faible

Objection principale à l'IA forte : le développement d'une machine intelligente passe-t-il nécessairement par une imitation de la pensée humaine ?

cf développement de l'aéronautique...

# Des systèmes qui pensent rationnellement

Des systèmes qui pensent comme des humains

Des systèmes qui pensent rationnellement

Des systèmes qui agissent comme des humains

Des systèmes qui agissent rationnellement

# Penser rationnellement

- Lois de la pensée données par la logique (Aristote, Leibniz, Boole, Frege...).
- Socrate est un homme; tous les hommes sont mortels; donc Socrate est mortel.
- La logique formelle permet d'écrire des énoncés sur les objets dans le monde.
- Lien direct entre les mathématiques et la philosophie vers l'IA moderne.
- Il est difficile de traduire les connaissances et les états du monde réel en des équations logiques (incertitude).

# Le modèle de la logique

Il existe tout une branche de la science qui s'intéresse à la formalisation de la notion de **raisonnement rationnel** : la **logique**.

Codification des procédés des raisonnements irréfutables

# Le modèle de la logique

Il existe tout une branche de la science qui s'intéresse à la formalisation de la notion de **raisonnement rationnel** : la **logique**.

Codification des procédés des raisonnements irréfutables

- Aristote (*syllogismes*)
- Logiciens du XIXème siècle (formalisation mathématique)
- Premiers programmes d'IA (années 1950-60)

# Le modèle de la logique

Il existe tout une branche de la science qui s'intéresse à la formalisation de la notion de **raisonnement rationnel** : la **logique**.

Codification des procédés des raisonnements irréfutables

- Aristote (syllogismes)
- Logiciens du XIXème siècle (formalisation mathématique)
- Premiers programmes d'IA (années 1950-60)

Cette vision idéalisée se heurte à deux obstacles principaux :

- modélisation de la connaissance informelle
- limites de la puissance de calcul

# Des systèmes qui agissent rationnellement

Des systèmes qui pensent comme des humains

Des systèmes qui pensent rationnellement

Des systèmes qui agissent comme des humains

Des systèmes qui agissent rationnellement

# Agir rationnellement

- Agir selon ses croyances pour atteindre des objectifs
- Exemple: *réflexe de cligner les yeux, retirer sa main d'un objet brulant...*
- Approche reposant sur la notion d'agent

**Agent rationnel:** Agent qui agit de manière à atteindre la meilleure solution ou, dans un environnement incertain, la meilleure solution prévisible.

# L'agent rationnel

Objectifs : à la fois plus ambitieux et moins ambitieux que les approches précédentes.

# L'agent rationnel

Objectifs : à la fois plus ambitieux et moins ambitieux que les approches précédentes.

Plus ambitieux, car l'objectif est la conception d'un agent **situé dans son environnement**.

Moins ambitieux, car l'agent rationnel doit agir de la **meilleure manière possible** étant donné les informations dont il dispose et ses capacités → concept de **rationalité limitée**

# L'agent rationnel

Objectifs : à la fois plus ambitieux et moins ambitieux que les approches précédentes.

Plus ambitieux, car l'objectif est la conception d'un agent **situé dans son environnement**.

Moins ambitieux, car l'agent rationnel doit agir de la **meilleure manière possible** étant donné les informations dont il dispose et ses capacités → concept de **rationalité limitée**

Cette approche s'intéresse aux principes généraux et aux composants nécessaires à la construction de tels agents rationnels.

## Synthèse des approches

# Définition

- Une définition possible de l'IA [\[Alliot et al., 2002\]](#), d'après John Mc Carthy Marvin Lee Minsky :  
*« Construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisantes par des êtres humains car elles demandent des processus mentaux de haut niveau tels que l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique. »*
- Définition par nature floue et aux contours mouvants (chaque fois qu'un problème est résolu en IA, il sort par définition du domaine de l'IA...)



Alliot, J.-M., Schiex, T., Brisset, P., and Garcia, F. (2002).

*Intelligence Artificielle et Informatique Théorique.*

Cépadùès.

# Définition

- « Automatisation des activités associées au raisonnement humain, telles que la décision, la résolution de problèmes, l'apprentissage, ... » (Bellman, 1978)

# Définition

- Concevoir des systèmes capables de reproduire le comportement de l'humain dans ses activités de raisonnement.
- Raisonner d'une manière autonome et également s'adapter aux changements de l'environnement

# Deuxième partie

---

- Philosophie
- Mathématiques
  - La logique mathématique
  - La théorie de la calculabilité
  - Le raisonnement incertain
- Économie et Recherche Opérationnelle
- Neurosciences
- Psychologie
- Informatique
- Automatique
- Linguistique

# Philosophie

# Les grandes questions philosophiques

- Lois de la pensée rationnelle : comment la partie rationnelle de l'esprit fonctionne-t-elle ? → philosophie logique.
- Esprit et matière : l'âme échappe-t-elle aux lois de la matière (dualisme, Descartes) ou non (matérialisme) ?
- Source de la connaissance : connaissance issue de l'expérience sensorielle (empirisme, Bacon, Locke...) et d'application d'un principe d'induction (Hume).
- Esprit et action : lien entre action et connaissance du résultat de cette action comme justification. Aristote, planification par régression...

# Mathématiques

# Naissance de la logique moderne

- Leibniz (1646 – 1716) : premiers langages artificiels écrits sous forme de signes
- Boole (1815–1864) : première mathématisation réelle de la logique (logique propositionnelle)
- Frege (1848–1925) : logique des prédictats, fondation de l'arithmétique comme branche de la logique
- Hilbert (1862–1943) : le programme de Hilbert → l'ensemble des mathématiques suit un ensemble d'axiomes qui peut être prouvé consistant. Système de déduction formelle.
- Bertrand Russell (1872–1970) : redéfinition de la théorie des ensembles (paradoxe de Russell), définition de la théorie des types
- Gödel (1906–1978) : théorème d'incomplétude

# Naissance de la logique moderne

- Herbrand (1908–1931) : résultat fondamental de logique, premier pas vers la déduction automatique
- Gentzen (1909–1945) : déduction naturelle, calcul des séquents
- Tarski (1902–1983) : travail sur la théorie des modèles, qui montre comment relier des objets du monde réel aux objets logiques.
- Brouwer (1898–1966) : vision constructiviste des preuves

**Économie et Recherche**

**Opérationnelle**

# Théorie de la décision

On s'intéresse aux décisions qui maximisent les gains

- Théorie de l'**utilitarisme** : Bentham (1748–1832) / Stuart Mill (1806–1873)
- Formalisation mathématique : Walras (1834–1910)
- Rationalité des agents économiques (loteries, utilités espérées...) : von Neumann et Morgenstern (1944).

# Décision collective et théorie des jeux

On s'intéresse à la prise de décision en présence d'agents coopératifs ou non

- Contexte coopératif :
  - Théorie du vote : Condorcet, Borda (XVIIIème s.)
  - Égalitarisme : Rawls (1971)
- Théorie des jeux :
  - von Neumann et Morgenstern (1944)
  - notions d'équilibres compétitifs : Nash (1951)

# Processus séquentiels

Où l'on s'intéresse à la maximisation de gains dans un contexte de décisions séquentielles

- Racines de ce problème : **Recherche Opérationnelle** (WW2)
- Étude des **processus décisionnels de Markov** : Bellman (1957)

# Neurosciences

# Les neurosciences et l'IA

On s'intéresse à comprendre comment le cerveau traite l'information

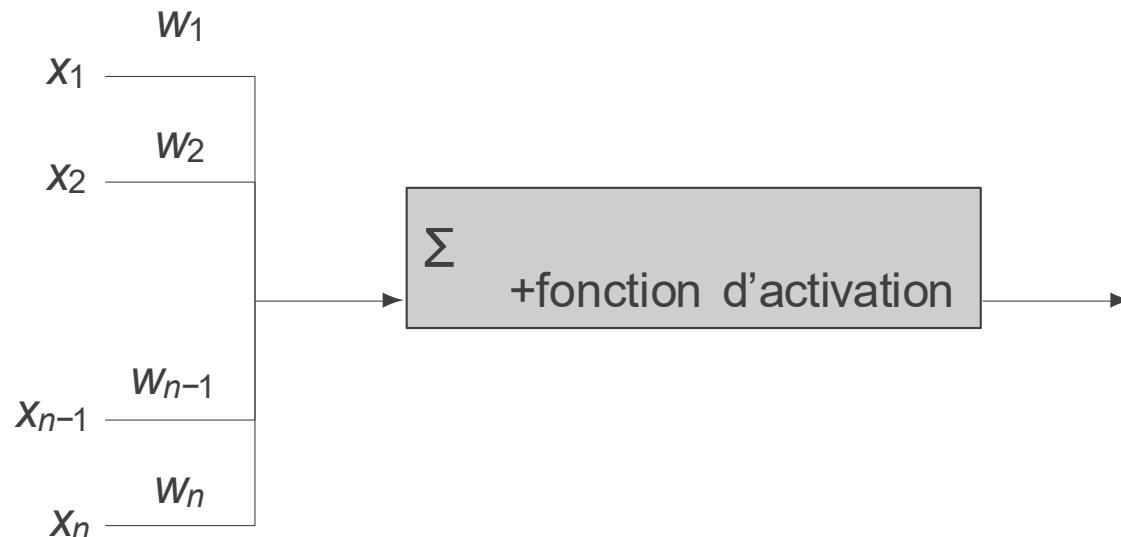
- Aires du cerveau = fonctions cognitives spécifiques : Broca (1861)
- Mise en évidence du **neurone** par coloration : Golgi (1873)
- Mesure non-invasive de l'activité cérébrale : électroencéphalogramme, Berger (1929)
- Imagerie par raisonnement magnétique (progrès récents) + stimulation électrique des neurones.

# Approches connexionnistes

- On sait encore peu de choses sur la manière dont le cerveau fonctionne réellement
- Mais l'observation des neurones a permis de définir le modèle du neurone **artificiel** (e.g. **perceptron**)...

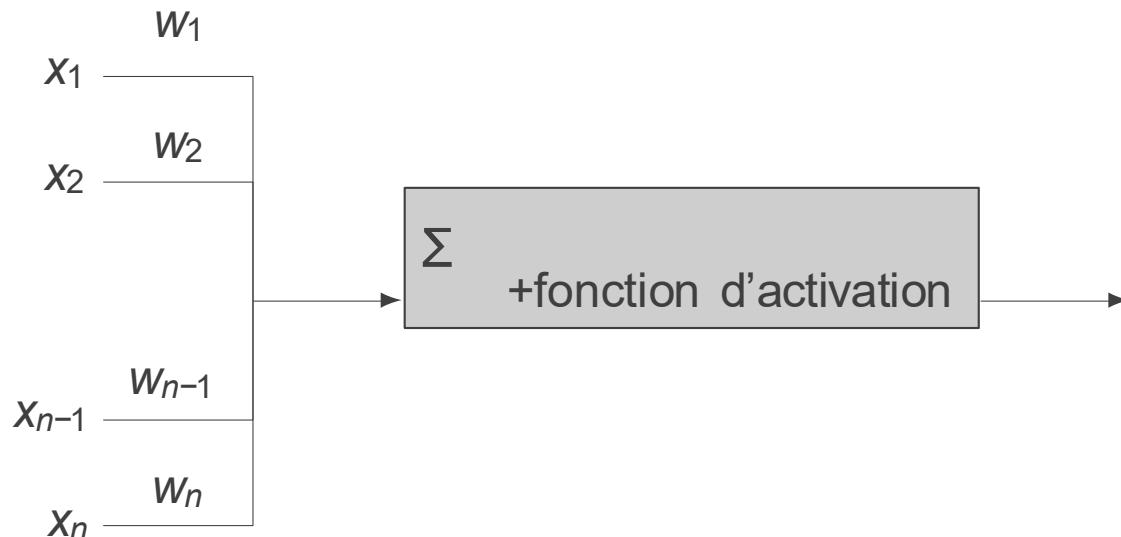
# Approches connexionnistes

- On sait encore peu de choses sur la manière dont le cerveau fonctionne réellement
- Mais l'observation des neurones a permis de définir le modèle du **neurone artificiel** (e.g. perceptron)...



# Approches connexionnistes

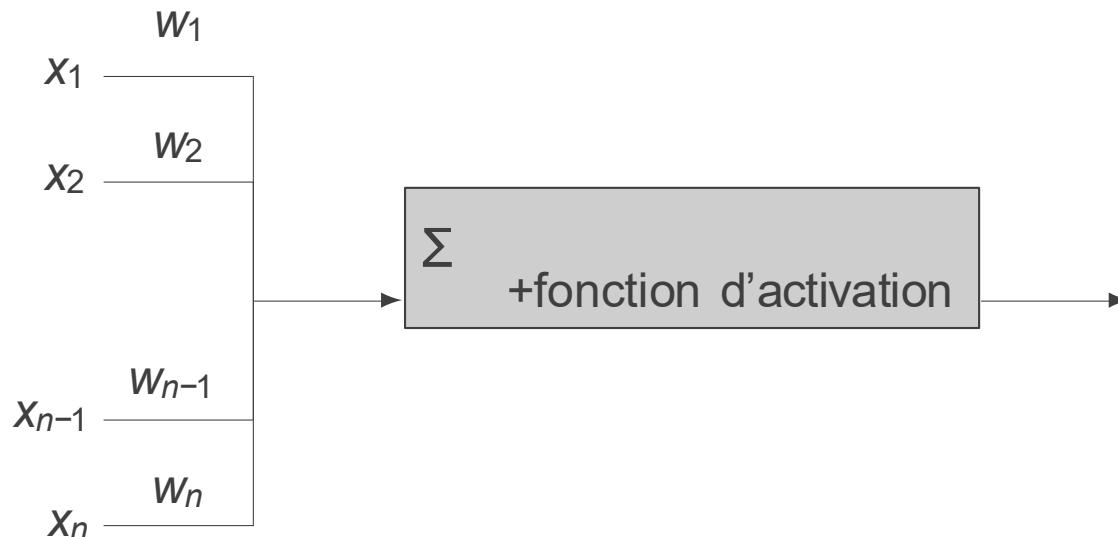
- On sait encore peu de choses sur la manière dont le cerveau fonctionne réellement
- Mais l'observation des neurones a permis de définir le modèle du **neurone artificiel** (e.g. perceptron)...



- ...Et les approches connexionnistes (réseaux de neurones)

# Approches connexionnistes

- On sait encore peu de choses sur la manière dont le cerveau fonctionne réellement
- Mais l'observation des neurones a permis de définir le modèle du **neurone artificiel** (e.g. **perceptron**)...



- ...Et les approches connexionnistes (réseaux de neurones)
- Interprétation de certains résultats (vision, traitement de la parole...) à l'aune des connaissances sur les processus de raisonnement du cerveau.

# Psychologie

# La psychologie et l'IA

On s'intéresse à la manière dont les humains (et les animaux) pensent et agissent

Deux courants antagonistes :

- Le **behaviorisme** (Watson, 1878–1958) : impossibilité de raisonner sur des états mentaux, car aucun moyen fiable d'accéder à l'**introspection**.
- La **psychologie cognitive** (James, 1842–1910) : modélise le processus de pensée comme un « programme informatique » qui convertit des stimuli en représentations internes, puis en actions.

Informatique

# Informatique et IA

...Où l'on s'intéresse à construire un ordinateur capable d'exécuter des calculs performants

- Premières machines **programmables** : métiers à tisser Jacquard (1752–1834), Babbage et (Ada) Lovelace (1792–1871)
- Enigma, premier ordinateur opérationnel (Turing, 1940) Z3 (Zuse, 1941) et l'ENIAC (Atanasoff et Berry, 1942)
- Augmentation constante des performances des CPU jusqu'en 2005
- À partir de 2005 : multicœurs et architectures parallèles
- Changement de paradigme, à partir de 2010 :
  - Données massives
  - Calculs sur GPU

Automatique

# Automatique et IA

...Où l'on s'intéresse aux systèmes dont le but est de maximiser une fonction objectif au cours du temps.

- À l'origine, des systèmes autorégulés : régulateur du moteur à vapeur (Watt, 1736–1819), thermostat (Drebbel, 1572–1633)
- Invention de la **théorie du contrôle** (automatique, cybernétique) : Wiener (1894–1964)
- Contrôle stochastique optimal : applique dans le domaine continu ce que l'on cherche à faire dans le domaine discret en informatique

Linguistique

# Linguistique et IA

...Où l'on s'intéresse aux rapports entre le langage et la pensée.

- Naissance de la linguistique moderne : Chomsky (1957), s'opposant au behaviorisme. Il existe des modèles syntaxiques permettant à une personne de construire une infinité de phrases à partir d'un ensemble fini d'exemples (structure innée)
- Naissance de la linguistique computationnelle en même temps que l'IA (compréhension de la langue naturelle, traduction automatique)
- Après les grandes espérances du début, désillusions (années 1960)
- Depuis 2010, succès des approches connexionnistes

# Troisième partie

- Débuts de l'IA (1943–1969)
- L'IA se spécialise (1969–1979)
- L'âge de raison de l'IA (1980–2000)
- L'explosion des données
- L'IA aujourd'hui
- Les grands domaines de l'IA

# Historique

## Précurseurs de l'IA

### (1943-1956)

- Pendant cette période furent menés les premiers travaux qui peuvent être considérés comme les débuts de l'intelligence artificielle.
- On peut citer les travaux de McCulloch et Pitts qui ont introduit en 1943 un modèle de neurones artificiels.
- Premiers modèles de neurones et de circuits à bases de neurones : McCulloch & Pitts (1943). Modèle d'apprentissage des neurones : Hebb (1949)
- Le SNARC, 1er ordinateur simulant un réseau de (40) neurones : Minsky et Edmonds (1950)
- Computing Machinery and Intelligence (Turing, 1950) : test de Turing, Turing a publié son fameux article dans lequel introduit le test de Turing.

# Historique

## Naissance d'IA

### Conférence de Dartmouth

#### (1956)

C'est durant cette année qu'un petit groupe d'informaticiens (Minsky, McCarthy, Shannon, Rochester) intéressés par l'étude de l'intelligence se réunirent pour une conférence sur ce thème.

- Séminaire de deux mois à l'été 1956, sous l'initiative de **John McCarthy** (Princeton)
- Une dizaine de spécialistes éminents dans la théorie des automates, les réseaux neuronaux, l'étude de l'intelligence.
- Première apparition du terme **Intelligence Artificielle**
- Première présentation du Logic Theorist de Newell et Simon (Carnegie Tech)

# Historique

## Espoirs grandissants

### (1952-1969)

Ce fut une période très active pour le jeune domaine de l'IA.

- Les programmes Logic Theorist (par Newell et Simon) et Geometry Theorem Prover furent en mesure de prouver certains théorèmes mathématiques.
- Le General Problem Solver de Newell et Simon réussissait quant à lui à résoudre des puzzles simples avec un raisonnement semblable au raisonnement humain.
- Hypothèse du **système symbolique matériel** : « un système symbolique matériel contient les moyens nécessaires et suffisants pour un comportement généralement intelligent »
- La recherche sur les réseaux de neurones fut également poursuivie Premiers travaux sur les **perceptrons** (Rosenblatt, 1962).
- Shakey, le premier robot à être capable de raisonner sur ses propres actions.

Quelques prédictions de Simon : dans 10 ans, un ordinateur serait champion d'échec et un théorème mathématique important serait démontré par une machine. + prédictions sur la traduction automatique.

# Shakey, first robot

- Créé en 1967
- 1h de calcul avant d'effectuer un mouvement
- Trop de bruits



# Historique

## Premières Déceptions

(1966-1973)

- Limite de puissance de calcul
- Ne respectent pas le passage à l'échelle
- Limite du savoir
- Fin des investissements
- Beaucoup de critiques

# Historique

## Premières Déceptions

(1966-1973)

- Les chercheurs n'avaient compté que 5 ans pour réaliser un traducteur automatique, mais se sont vite rendu compte que leur approche purement syntaxique n'étaient pas suffisante (pour bien traduire un texte, il faut d'abord le comprendre).
- Cet échec a provoqué l'annulation en 1966 de tout le financement du gouvernement américain pour les projets de traduction automatique.

# Historique

## Premières Déceptions

(1966-1973)

Minsky et Papert prouvent dans leur livre “Perceptrons” de 1969 que les réseaux de neurones de l'époque ne pouvaient pas calculer certaines fonctions pourtant très simples, ce qui mit en cause toute la recherche en apprentissage automatique, entraînant une crise dans cette branche de l'IA.

# Historique

## L'IA de spécialise

(1969-1979)

Très rapidement, l'IA se heurte à trois écueils :

- la limite des manipulations syntaxiques dans des domaines tels que la traduction automatique
- l'explosion combinatoire
- la capacité de représentation limitée des structures fondamentales permettant de générer un comportement intelligent (e.g. perceptrons)

# Historique

## Systèmes Experts

(1969-1979)

- Le premier système expert, appelé **DENDRAL**, créé en 1969 pour la tâche spécialisée consistant à déterminer la structure moléculaire d'une molécule étant données sa formule et les résultats de sa spectrométrie de masse.
- Le système **MYCIN**, créé en 1972, réussissait à diagnostiquer à un niveau proches des experts humains et considérablement meilleur que celui les jeunes médecins.

# Historique Industrialisation

- Premier système expert à succès commercial : R1 (DEC, 1982)
- Projet **Cinquième Génération** lancé par le Japon (1981) : construction d'ordinateurs intelligents programmés en Prolog
- USA : création du consortium **MCC**
- Hausse constante des budgets consacrés à l'IA dans les entreprises...
- ...Suivie d'une phase de désillusion : **l'hiver de l'IA**
- Mais dans le même temps, l'IA devient de plus en plus une science, organisée en sous-domaines

# Historique

## Retour des réseaux de neurones

### (1986-présent)

- Au milieu des années 80, quatre groupes de chercheurs ont découvert indépendamment la règle d'apprentissage "back-propagation" qui permit le développement de réseaux de neurones capables d'apprendre des fonctions très complexes (curieusement, cette règle avait déjà été proposée en 1969, mais n'avait eu aucun écho dans la communauté scientifique).
- Depuis, l'apprentissage automatique est devenu l'un des domaines les plus actifs de l'IA, et a été appliqué avec succès à de nombreux problèmes pratiques.
- Premiers succès sur des problèmes très spécifiques (e.g. reconnaissance de caractères)
- Mais les succès ne sont pas au rendez-vous sur d'autres types de problèmes (e.g. reconnaissance d'images)

# Historique

## Les agents

(1986-présent)

- Retour de l'étude du problème de l'agent total dans les années 1980
- Agent considéré dans son environnement, avec des entrées sensorielles continues et des interactions avec d'autres entités
- Un regain d'intérêt encore plus marqué avec l'avènement du web (bots, systèmes de recommandation, moteurs de recherche...)

# Historique

## Les jeux

- **Deep Blue** (IBM, 1997) : premier ordinateur ayant battu un champion mondial d'échecs
- Succès de l'IA faible (algorithme de parcours d'arbre couplé à une gigantesque base de données)

# Les données et le renouveau de l'IA

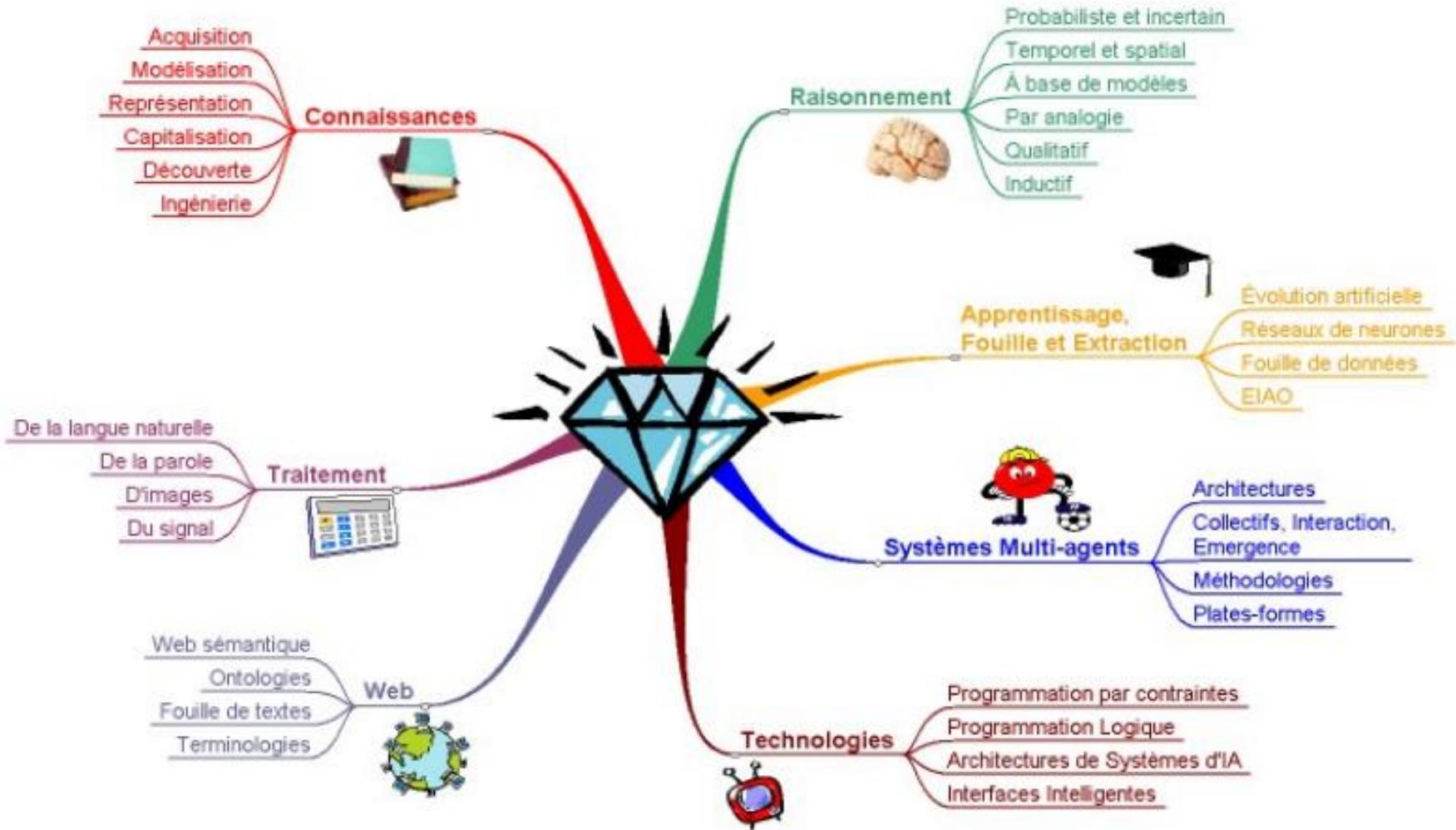
- Avènement du web et des systèmes connectés → gigantesques corpus de données
- Premier exemple : désambiguïsation (Yarowsky, 1995)
- Watson (IBM, 2011)
- Le renouveau des **réseaux de neurones** (cf AlexNet, 2012)

# Historique IA Moderne

(1987-présent)

- L'intelligence artificielle est devenue au fil du temps une matière scientifique de plus en plus rigoureuse et formelle.
- La plupart des approches étudiées aujourd'hui sont basées sur des théories mathématiques ou des études expérimentales plutôt que sur l'intuition, et sont appliquées plus souvent aux problèmes issus du monde réel.

# Domaines d'application



# Les applications les plus visibles

- **Véhicules autonomes :**
  - Challenge DARPA en 2005 (désert du Mohave)
  - Google commence à développer une voiture autonome en 2009, suivi par plusieurs constructeurs.
  - 2012 : premier permis de circulation pour un véhicule autonome (Nevada)
  - 2018 : premier accident mortel impliquant un piéton (Arizona)
- **Robots :**
  - robots aspirateurs, tondeuses, domotique
  - robots guides (musées...)

# Les applications les plus visibles (2)

- **Jeux :**
  - Échecs : Deep Blue (IBM, 1997)
  - Jeopardy : Watson (IBM, 2011)
  - Go : AlphaGo (Google, 2017)
- **(Chat)bots :**
  - Jeux
  - SAV / FAQ d'entreprises...
  - ChatGPT (2022)
- **Assistants Intelligents** (Siri, Google Assistant, Alexa...)
- **Reconnaissance faciale** (photo, vidéo)
- **Aide au diagnostic médical**
- **IA prédictive** (embauches, justice...)
- ...

# Les grands domaines de l'IA

# Représentation de la connaissance

- Logique propositionnelle, logique du premier ordre
- Ontologies, ingénierie des connaissances
- Informations incertaines, probabilités
- Préférences, utilités, élicitation

# Représentation de la connaissance

- Logique propositionnelle, logique du premier ordre
- Ontologies, ingénierie des connaissances
- Informations incertaines, probabilités
- Préférences, utilités, élicitation

IA essentiellement **symbolique** visant à la **formalisation** de connaissances humaines

# Raisonnement

- Inférence en logique. Déduction, induction, abduction
- Logiques non monotones
- Raisonnement sur le temps et l'espace
- Raisonnement sur l'incertain
- Raisonnement en situation de coopération ou d'adversité
- Planification
- Problèmes de satisfaction de contraintes

# Raisonnement

- Inférence en logique. Déduction, induction, abduction
- Logiques non monotones
- Raisonnement sur le temps et l'espace
- Raisonnement sur l'incertain
- Raisonnement en situation de coopération ou d'adversité
- Planification
- Problèmes de satisfaction de contraintes

IA essentiellement **symbolique**, visant à la **formalisation** des règles du raisonnement rationnel, et à la **mise en œuvre** de ces règles sur des problèmes concrets.

# Systèmes multiagents

- Architectures
- Interaction, protocoles
- Émergence
- Simulation
- Plates-formes

# Systèmes multiagents

- Architectures
  - Interaction, protocoles
  - Émergence
  - Simulation
  - Plates-formes
- 
- Formalisation de l'intelligence collective et des comportements humains.

# Communication, perception, action

- Traitement du langage naturel / reconnaissance de la parole
- Perception / reconnaissance d'images
- Robotique

# Communication, perception, action

- Traitement du langage naturel / reconnaissance de la parole
- Perception / reconnaissance d'images
- Robotique

Une partie de ces domaines fait appel aux bases théoriques présentés dans les planches précédentes (+ apprentissage). Par exemple : planification pour la robotique.

# Apprentissage

L'apprentissage désigne la faculté à créer de la connaissance, permettant de résoudre des tâches pour lesquels l'algorithme n'a pas été explicitement programmé.

# Apprentissage

L'apprentissage désigne la faculté à créer de la connaissance, permettant de résoudre des tâches pour lesquels l'algorithme n'a pas été explicitement programmé.

Pour certains scientifiques, il s'agit de l'essence-même de l'intelligence.

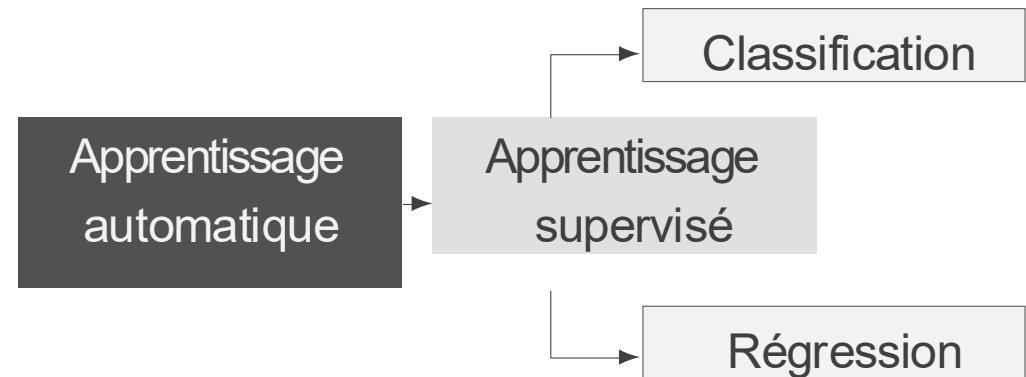
# Différents domaines de l'apprentissage

Apprentissage automatique

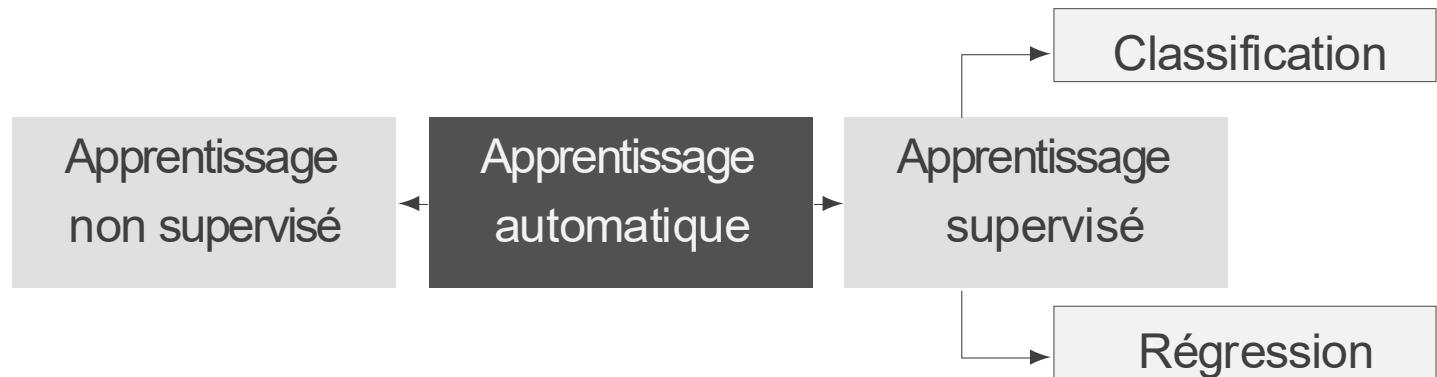
# Différents domaines de l'apprentissage



# Différents domaines de l'apprentissage



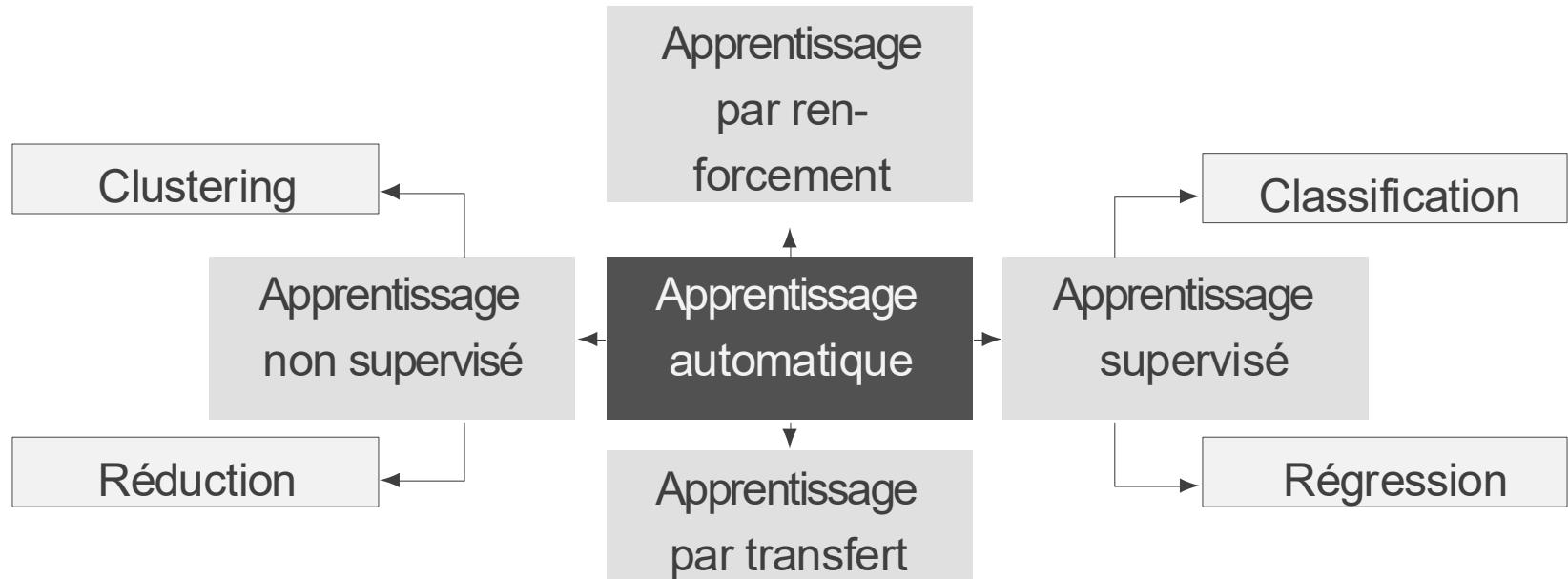
# Différents domaines de l'apprentissage



# Différents domaines de l'apprentissage



# Différents domaines de l'apprentissage



# Algorithmes d'apprentissage

Voici quelques exemples de méthodes d'apprentissage classiques que vous pourriez être amenés à rencontrer (non exhaustif) :

## Apprentissage non supervisé :

Clustering

Partitionnement (k-moyennes,...)

Regroupement hiérarchique (dendrogramme)

Réduction de dimension

Analyse en composantes principales

## Apprentissage supervisé :

Régression

Régression linéaire / polynomiale

Régression logistique

Classification

Machines à vecteurs supports (SVM)

Arbres de décisions

k plus proches voisins

Forêts aléatoires

Réseaux de neurones (profonds ou non)

# L'IA est-elle souhaitable ?

Quelques questions éthiques à emporter...

...En 2010 (cf Russell et Norvig) :

- l'IA pourrait faire perdre des emplois
- les gens pourraient avoir trop (ou pas assez) de temps libre
- l'humanité pourrait perdre le sens de son unicité
- l'IA pourrait être utilisée à des fins indésirables (cf slaughterbots)
- l'IA pourrait entraîner une perte de responsabilité
- l'avènement de l'IA pourrait entraîner la fin de l'humanité

# L'IA est-elle souhaitable ?

...En 2019 :

- comment la machine doit-elle résoudre des dilemmes moraux ?
- est-il acceptable de remplacer la décision humaine par une décision automatique ?
- qui est responsable en cas d'accident ?
- l'IA renforce-t-elle des biais existants ?
- une IA peut-elle / doit-elle être raciste / misogyne / discriminatoire... ?
- l'IA renforce-t-elle des structures de pouvoir existantes ?

# L'IA est-elle souhaitable ?

...En 2023 :

- Doit-on accepter d'utiliser ChatGPT à l'école / dans les universités ?
- Comment les images générées par l'IA (photos de presse, *deepfakes*, etc. modifie-t-elle notre rapport à la vérité ?

103 /  
73

...Et des questions plus indirectes, mais fondamentales :

- Sur quels matériaux bruts l'IA actuelle (au sens apprentissage profond) est-elle construite ?
  - Matières premières
  - Données brutes (→ biais potentiels, droits d'utilisation...)
  - Microtravail, donc exploitation humaine
- Corollaire : *À qui bénéficie l'IA, et au détriment de qui ?*

# Les trois théories morales

Déontologie

Respect des règles et  
devoirs

Conséquentialisme  
Conséquences des actions

Éthique de la vertu  
Caractère moral de la  
personne

# Les trois théories morales :

## Déontologie

- Basée sur le respect des règles et des devoirs
- Les actions sont jugées moralement correctes si elles respectent des principes universels
- Exemple : Dire la vérité, même si cela entraîne des conséquences négatives

# Les trois théories morales : Conséquentialisme

- Juge la moralité d'une action selon ses conséquences
- L'utilitarisme est la forme la plus connue : maximiser le bonheur collectif
- Exemple : Choisir l'action qui apporte le plus grand bien au plus grand nombre

# Les trois théories morales : Éthique de la vertu

- Se concentre sur le caractère moral de la personne
- Vise à développer des vertus comme la justice, le courage, la prudence
- Exemple : Agir honnêtement car c'est une qualité d'une bonne personne

# Définitions à partir de la philosophie morale

## Morale

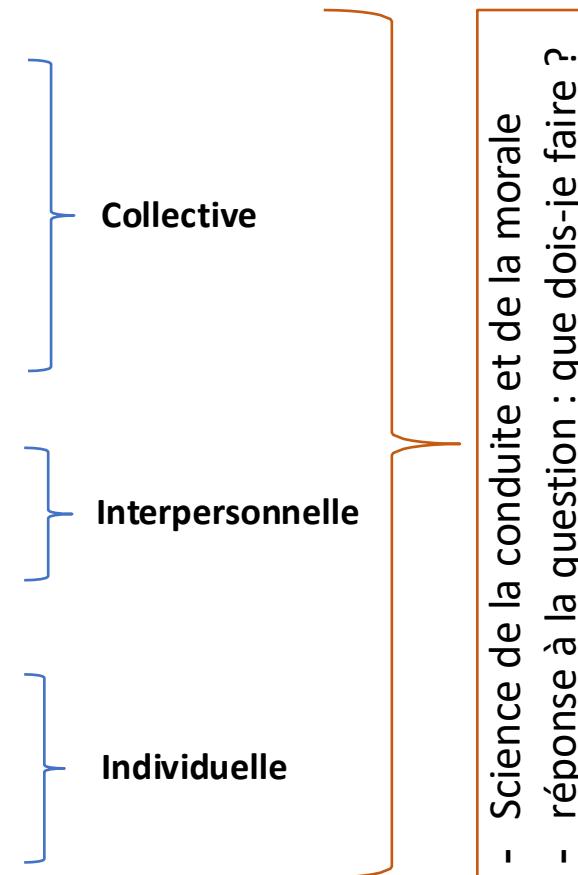
- Ensemble de valeurs et de principes qui permettent de différencier le bien du mal, le juste de l'injuste, l'acceptable de l'inacceptable, et auxquels il faudrait se conformer
- Ensemble des normes propres à un individu, un groupe social ou un peuple, à un moment précis de son histoire
- Notion de droits, de devoirs et d'interdits
- Notion de bonne/mauvaise action

## Déontologie

- Règles et devoirs qui encadrent une profession ou un groupe de personnes

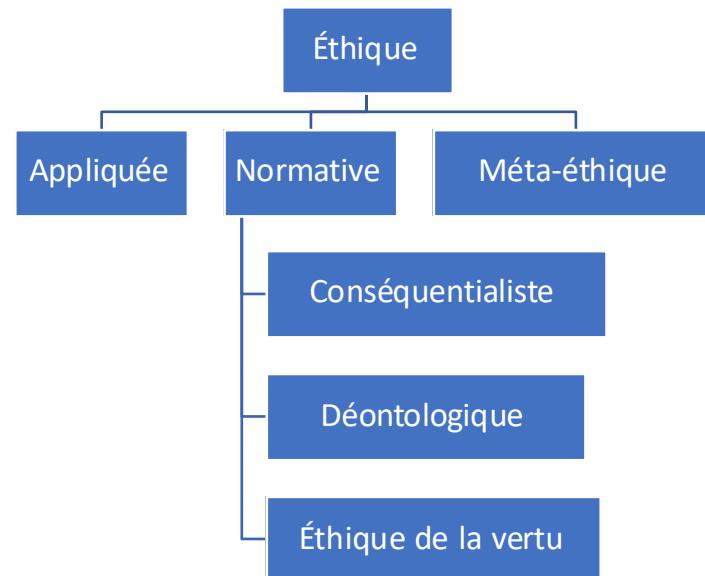
## Éthique

- Ce n'est pas un ensemble de valeurs ni de principes en particulier
- Il s'agit d'une réflexion argumentée en vue du bien-agir
- Relativise la notion de la bonne/mauvaise action



# Éthique

De façon générale, l'éthique propose de s'interroger sur les valeurs morales et les principes moraux qui devraient orienter nos actions, dans différentes situations, dans le but d'agir conformément à ceux-ci.



# Éthique appliquée

Analyse de situations concrètes qui soulèvent des questions éthiques :

- Accent mis sur le soutien à la prise de décision face à des enjeux concrets.
- Ne concerne le point de vue de la forme et du processus décisionnel que du point de vue substantiel, c'est-à-dire des valeurs et principes en jeu et de leurs rapports entre eux.

Exemples :

- Bioéthique : procréation artificielle, manipulations génétiques...
- Éthique de l'environnement : développement durable, responsabilité face aux générations futures, biodiversité...

# Éthique normative (1/2)

Proposition de règles pour évaluer une action d'un point de vue moral.

Ne permet pas de déterminer, entre deux actions, laquelle est moralement meilleure.

## 1. Éthique conséquentialiste (ou conséquentialisme)

- S'intéresse à l'ensemble des conséquences
- Une action est bonne si ses conséquences sont bonnes

## 2. Éthique déontologique (ou déontologisme)

- Notion de devoir, d'obligation et d'impératif moral
- Un acte est moralement bon ou mauvais indépendamment de ses conséquences
- Exemple : devoir de parent

## 3. Éthique de la vertu

- Traits de caractère dont témoignent les actions
- Exemple : l'aveu est toujours associé à la vertu d'honnêteté

# Éthique normative (2/2)

Une conception du bien

- Quelles sont les meilleures conséquences ?
- Quelles sont mes obligations morales ?
- Comment savoir quelles vertus adopter ?
- Une conception des bonnes conséquences, des devoirs moraux fondamentaux, vertus à privilégier
- Théories morales incluses dans ces trois grandes approches

# Problèmes et questions posés par l'IA

- Atteinte à la vie privée et aux libertés
- Contrôle et risque d'asservissement des personnes
- Risques liés à l'utilisation des réseaux sociaux
- Espionnage, cyberattaques et criminalité
- Systèmes autonomes (Robots, armes létales...)
- Prolétarisation
- Atomisation de la société
- Perte de la solidarité
- ...

# Intelligence Artificielle

## Symbolique Vs Numérique

### IA Numérique : Machine Learning (ML)

- ✓ Construit des modèles basés sur les données (données d'apprentissage) pour réaliser des tâches spécifiques
- ✓ Capacités de prédiction s'améliorent au fur et à mesure de leur apprentissage
- ✓ Prédiction de nouvelles situations
- Entièrement guidée par les données
- Les modèles les plus performants sont opaques

### IA Symbolique

- ✓ Repose sur le raisonnement logique et formel
- ✓ À l'origine des premiers systèmes experts
- ✓ Permet de tracer le processus de raisonnement
- N'explique pas les raisons des décisions (solution/échec à trouver une solution)
- Basée sur une expertise limitée

# Initiatives pour limiter les risques liés à Intelligence Artificielle

- Lettre ouverte en 2015 initiée par Stephen Hawking et Elon Musk pour interdire les robots tueurs (signée par plus de 1000 chercheurs)
- RGPD, 2016
- Yoshua Bengio en 2018 appelle les gouvernements pour réglementer avec des lois en vue du développement de l'IA pour le bien social
- Déclaration de Montréal en 2018, pour une IA responsable (signée par : Yoshua Bengio et Yann LeCun)
- ONU : Appel du 15 septembre 2021 pour l'instauration d'un moratoire sur certaines applications de l'IA
- Conférence mondiale à l'UNESCO en 2019 : « Principes pour l'Intelligence Artificielle : vers une approche humaniste ? »
- International Telecommunication Union (ITU) & AI for good Foundation : "*What if AI were developed to serve humanity rather than commerce ?*"
- Global AI Ethics Institute : think tank mondial sur l'éthique appliquée à l'IA
- Lettre ouverte : *Research Priorities for Robust and Beneficial Artificial Intelligence* ([Stuart Russell & al., 2015])
- *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*
- The Ethics and Governance of Artificial Intelligence Initiative : projet commun au MIT Media Lab et the Harvard Berkman-Klein Center for Internet and Society (fairness, human autonomy, and justice)
- Global Catastrophic Risk Institute
- ...