

Projet d'Intelligence Artificielle

Séance #1 – Introduction à l'Intelligence Artificielle

Sylvain Bouveret - sylvain.bouveret@imag.fr

Ensimag ISI 3^{ème} année



Introduction

L'objectif est de réaliser un projet par équipes, sur un sujet librement choisi, mais **relevant du domaine de l'Intelligence Artificielle**.

L'objectif est de réaliser un projet par équipes, sur un sujet librement choisi, mais **relevant du domaine de l'Intelligence Artificielle**.

- Équipe de 4 (à constituer sur Teide)
- Sujet à choisir : relevant du domaine de l'IA au sens large (pas seulement ML)
- Projet évalué sur un livrable et une soutenance
- Toutes les séances (sauf celle-ci) sont des séances de suivi de projet. Présence **obligatoire**.

Livrable attendu sous forme d'une page web contenant *a minima* :

- la présentation du problème traité
- les données utilisées si besoin
- les méthodes utilisées et leur justification
- une bibliographie suffisamment développée
- une évaluation fine des aspects environnementaux et sociétaux
- l'ensemble des connaissances créées (par exemple sous la forme d'un Notebook résultat)
- un lien vers le code et l'infrastructure permettant de régénérer la connaissance.

NB : il y aura des livrables intermédiaires non évalués (mais utilisés pour le suivi).

- Séances :
 - Semaine 1 (28/09 — aujourd’hui) : CM
 - Semaine 2 (05/10) : TD. **Constituer les équipes**
 - Semaine 3 (19/10) : TD. **Livrable 1**
 - Semaines 4–5 (26/10 – 09/11) : TD
 - Semaine 6 (16/11) : **Pitch + livrable 2**
 - Semaines 7–8 (23/11 – 07/12) : TD
 - Semaine 9 (14/12) : TD. **Livrable 3**
 - Semaines 10–11–12 (21/12 – 11/01 – 18/01) : TD
 - **Soutenances la semaine du 22/01**
- Intervenants :
 - Sylvain Bouveret – sylvain.bouveret@imag.fr
 - Antoine Frenoy – antoine.frenoy@univ-grenoble-alpes.fr
 - Thibault Tricard – thibault.tricard@univ-grenoble-alpes.fr

Toutes les informations disponibles sur la page Chamilo du cours :
<https://chamilo.grenoble-inp.fr/courses/ENSIMAG5MMPIA/>

- Ratissez large : l'IA ne se limite pas à l'apprentissage machine
- **Formalisez** correctement votre problème avant de penser à la méthode que vous allez utiliser pour le résoudre
- Utilisez la méthode adéquate pour résoudre votre problème
- Pour cela, ne négligez pas la phase **bibliographique**
- N'oubliez pas l'évaluation des aspects **environnementaux et sociétaux**
- Sollicitez-nous : nous ne sommes pas tous experts dans tous les domaines de l'IA, mais nous pouvons probablement vous orienter.

- La séance d'aujourd'hui est une introduction générale à l'IA
- Objectifs :
 - Vous donner quelques éléments de définition générale
 - Vous donner une perspective historique globale
 - Vous donner une idée de l'étendue du domaine de l'IA
- Le cours ne vise pas à l'exhaustivité, mais à replacer le projet dans le contexte du domaine de l'IA (et à sortir un peu de l'idée reçue IA = ML)

Note : cette première partie s'inspire largement de Russel et Norvig (2010)



Russell, S. J. and Norvig, P. (2010).

Artificial intelligence : a modern approach.

Pearson Education, 3rd edition edition.

1

Première partie

Bon, c'est quoi en fait l'IA ?

2. Les différentes définitions de l'IA

Des systèmes qui agissent comme des humains

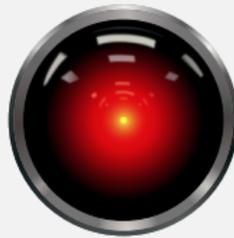
Des systèmes qui pensent comme des humains

Des systèmes qui pensent rationnellement

Des systèmes qui agissent rationnellement

3. Synthèse des approches

Les différentes définitions de l'IA



Qu'en pensez-vous ?
(2 minutes de réflexion)

Des systèmes qui pensent
comme des humains

Des systèmes qui
pensent rationnellement

Des systèmes qui agissent
comme des humains

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

Deux dimensions :

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

Des systèmes qui pensent
comme des humains

Des systèmes qui
pensent rationnellement

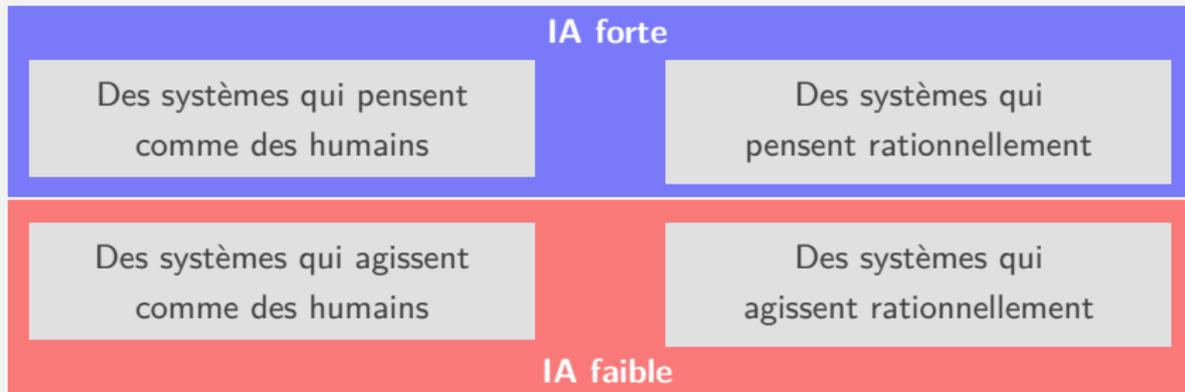
Des systèmes qui agissent
comme des humains

Des systèmes qui
agissent rationnellement

Deux dimensions :

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

Les quatre approches ont été suivies et ont mené à des voies différentes.



Deux dimensions :

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

Les quatre approches ont été suivies et ont mené à des voies différentes.

Des systèmes qui agissent comme des humains

Des systèmes qui pensent
comme des humains

Des systèmes qui agissent
comme des humains

Des systèmes qui
pensent rationnellement

Des systèmes qui
agissent rationnellement

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.

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.

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**.

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) : entre 2008 et 2020, compétition internationale récompensant les agents conversationnels satisfaisant le mieux le test de Turing (5 minutes de conversation).

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) : 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)

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.

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
- ...

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

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

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

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.

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).

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...).

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

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

Difficulté : saisir comment pensent les humains.

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

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.

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

Autre objection au test de Turing : capacité d'une machine à se comporter **comme un humain**, pas forcément **intelligemment**.

cf fautes de frappe, incohérences, mensonge...

→ est-ce réellement le but que l'IA devrait se fixer ?

Il existe toute 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

Il existe toute 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)

Il existe toute 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

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

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**

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

Une définition possible de l'IA [Alliot et al., 2002], d'après Minsky...

L'intelligence artificielle a pour but de faire exécuter par l'ordinateur des tâches pour lesquelles l'homme, dans un contexte donné, est aujourd'hui meilleur que la machine.

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.
Cépaduès.

Définition d'un « Système d'intelligence artificielle » selon l'Union Européenne :

[AI system] means software that is developed with one or more of the techniques and approaches listed in Annex I¹.

1. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206>

Définition d'un « Système d'intelligence artificielle » selon l'Union Européenne :

[AI system] means software that is developed with one or more of the techniques and approaches listed in Annex I¹.

Annex I :

1. Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning ;
2. Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems ;
3. Statistical approaches, Bayesian estimation, search and optimization methods.

1. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206>

Seuls une minorité de chercheurs se sont intéressés au test de Turing.

Seuls une minorité de chercheurs se sont intéressés au test de Turing.

Souvent, ce test est inutile, car on s'intéresse au développement de machines capables de bien résoudre un **problème précis** plutôt que de berner un humain.

Exemple : AlphaGo (ne passe certainement pas le test de Turing)

Seuls une minorité de chercheurs se sont intéressés au test de Turing.

Souvent, ce test est inutile, car on s'intéresse au développement de machines capables de bien résoudre un **problème précis** plutôt que de berner un humain.

Exemple : AlphaGo (ne passe certainement pas le test de Turing)

L'objectif de l'aéronautique est-il de créer des machines capables de se faire passer pour des pigeons ?

2 Deuxième partie

Les racines de l'IA

4. Philosophie

5. Mathématiques

La logique mathématique

La théorie de la calculabilité

Le raisonnement incertain

6. Économie et Recherche Opérationnelle

7. Neurosciences

8. Psychologie

9. Informatique

10. Automatique

11. Linguistique

Nous n'aurons pas le temps de présenter les innombrables racines où l'IA moderne puise sa source...

Mais tout est disponible dans les planches du cours. Je vous conseille d'aller y jeter un œil.

Passer à la partie suivante...

Philosophie

- Lois de la **pensée rationnelle** : comment la partie rationnelle de l'esprit fonctionne-t-elle ? → philosophie logique. Aristote et ses syllogismes (IV^{ème} av. JC), Lull (XIV^{ème}), Hobbes (XVII^{ème}), Pascal (XVII^{ème}), Leibniz (XVIII^{ème}).
- **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

...Où l'on se demande comment formaliser le raisonnement rationnel.

- Aristote (-384 – -322)
 - syllogismes : « Tout homme est mortel. Socrate est un homme. Socrate est mortel. »
 - notion de validité, preuve par contradiction
- Les stoïciens (-500 – -200)
 - Notions de propositions atomiques et composées
 - Implication, disjonction, conjonction
- Logique médiévale scholastique (500 – 1400)

- 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 prédicats**, 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**

- 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

...Où l'on se demande si l'on peut tout calculer

- Premier algorithme formel : Euclide
- Algorithmes comme objet d'étude : Al-Khwârizmî – IXème s.
- Gödel
 - procédure effective pour démontrer tout énoncé vrai de la logique du 1^{er} ordre de Frege
 - **théorème d'incomplétude**

- Si l'arithmétique est consistante, il existe des énoncés dont on ne peut ni prouver la véracité, ni prouver le contraire
- La consistance de l'arithmétique ne peut pas être prouvée en arithmétique.

cf Démonstration à base des machines de Turing (Kleene, 1943)

- Point de départ : il existe des fonctions non calculables dans tout système formel
- Comment caractériser en pratique l'ensemble des fonctions calculables ?

- Point de départ : il existe des fonctions non calculables dans tout système formel
- Comment caractériser en pratique l'ensemble des fonctions calculables ?

Un modèle formel : la machine de Turing (rappel : des états, une mémoire linéaire infinie, une tête de lecture / écriture, des transitions...)

Thèse de Church-Turing : une fonction est calculable « intuitivement » ssi elle est calculable par une machine de Turing.

- Point de départ : il existe des fonctions non calculables dans tout système formel
- Comment caractériser en pratique l'ensemble des fonctions calculables ?

Un modèle formel : la machine de Turing (rappel : des états, une mémoire linéaire infinie, une tête de lecture / écriture, des transitions...)

Thèse de Church-Turing : une fonction est calculable « intuitivement » ssi elle est calculable par une machine de Turing.

Modèle équivalent à :

- λ -calcul de Church (racines de la programmation fonctionnelle)
- Fonctions généralement récursives de Gödel

Étapes-clefs du raisonnement :

- Toute machine de Turing \mathcal{M}_i peut être représentée comme un mot en base 17 (pour chaque couple (état, symbole lu), un triplet (état, symbole écrit, action)) \rightarrow un nombre i dans \mathbb{N}

Étapes-clefs du raisonnement :

- Toute machine de Turing \mathcal{M}_i peut être représentée comme un mot en base 17 (pour chaque couple (état, symbole lu), un triplet (état, symbole écrit, action)) \rightarrow un nombre i dans \mathbb{N}
- De même, toute entrée peut être représentée par un nombre dans \mathbb{N}

Étapes-clefs du raisonnement :

- Toute machine de Turing \mathcal{M}_i peut être représentée comme un mot en base 17 (pour chaque couple (état, symbole lu), un triplet (état, symbole écrit, action)) \rightarrow un nombre i dans \mathbb{N}
- De même, toute entrée peut être représentée par un nombre dans \mathbb{N}
- $T(i, a, n)$ est vrai ssi la machine \mathcal{M}_i appliquée à l'entrée a termine au bout de n étapes (sur un résultat $\varphi_i(a)$)

Étapes-clefs du raisonnement :

- Toute machine de Turing \mathcal{M}_i peut être représentée comme un mot en base 17 (pour chaque couple (état, symbole lu), un triplet (état, symbole écrit, action)) \rightarrow un nombre i dans \mathbb{N}
- De même, toute entrée peut être représentée par un nombre dans \mathbb{N}
- $T(i, a, n)$ est vrai ssi la machine \mathcal{M}_i appliquée à l'entrée a termine au bout de n étapes (sur un résultat $\varphi_i(a)$)
- Ce prédicat est décidable

Étapes-clefs du raisonnement :

- Toute machine de Turing \mathcal{M}_i peut être représentée comme un mot en base 17 (pour chaque couple (état, symbole lu), un triplet (état, symbole écrit, action)) \rightarrow un nombre i dans \mathbb{N}
- De même, toute entrée peut être représentée par un nombre dans \mathbb{N}
- $T(i, a, n)$ est vrai ssi la machine \mathcal{M}_i appliquée à l'entrée a termine au bout de n étapes (sur un résultat $\varphi_i(a)$)
- Ce prédicat est décidable
- φ_i est calculable (il suffit d'appliquer \mathcal{M}_i à l'argument et d'attendre n étapes)

Étapes-clefs du raisonnement :

- Toute machine de Turing \mathcal{M}_i peut être représentée comme un mot en base 17 (pour chaque couple (état, symbole lu), un triplet (état, symbole écrit, action)) \rightarrow un nombre i dans \mathbb{N}
- De même, toute entrée peut être représentée par un nombre dans \mathbb{N}
- $T(i, a, n)$ est vrai ssi la machine \mathcal{M}_i appliquée à l'entrée a termine au bout de n étapes (sur un résultat $\varphi_i(a)$)
- Ce prédicat est décidable
- φ_i est calculable (il suffit d'appliquer \mathcal{M}_i à l'argument et d'attendre n étapes)
- La fonction suivante n'est pas calculable :

$$\psi(a) = \begin{cases} \varphi_a(a) + 1 & \text{si } \exists n \text{ tel que } T(a, a, n) \\ 0 & \text{sinon} \end{cases}$$

Au-delà de la question de la calculabilité, une question cruciale est celle de la **calculabilité pratique**.

Au-delà de la question de la calculabilité, une question cruciale est celle de la **calculabilité pratique**.

Théorie de la **NP-complétude** : Cook (1971) et Karp (1972).

→ même si une fonction est calculable en théorie, elle peut nécessiter une telle quantité de ressources (mémoire, temps) qu'il est impossible de la calculer en pratique.

...Où l'on s'intéresse aux raisonnements que l'on peut faire dans un environnement incertain

- Aux racines, la théorie des **probabilités** : Cardan, Pascal (probabilités comme mesure de moyennes d'occurrences sur une suite de paris infinis), Bernouilli, Laplace (méthodes statistiques)
- Bayes (1702–1761) : Comment actualiser des probabilités à partir d'observations. À la base de tout le domaine du raisonnement sous incertitude : **raisonnement bayésien**.

- Les limites des probabilités :
 - Pas de distinction entre ignorance, contradiction, imprécision...
 - Probabilités \approx Certitude sur l'incertitude

- Les limites des probabilités :
 - Pas de distinction entre ignorance, contradiction, imprécision...
 - Probabilités \approx Certitude sur l'incertitude
- Ensembles **fous** : modélisation d'une appartenance imprécise (Zadeh, 1965)

- Les limites des probabilités :
 - Pas de distinction entre ignorance, contradiction, imprécision...
 - Probabilités \approx Certitude sur l'incertitude
- Ensembles **flous** : modélisation d'une appartenance imprécise (Zadeh, 1965)
- **Théorie des possibilités** (Zadeh, 1978) : représentation et raisonnement sur des informations imprécises (notion de crédibilité et plausibilité). Inclut la **théorie des probabilités**.

Économie et Recherche Opérationnelle

...Où l'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).

...Où l'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)

...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)

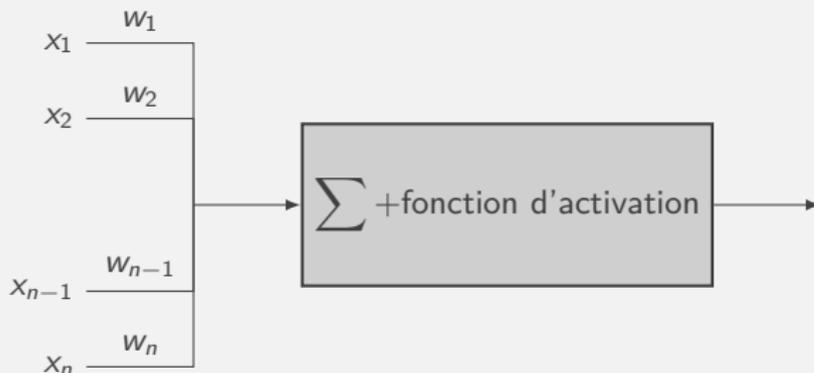
Neurosciences

...Où l'on s'intéresse à comprendre comment le cerveau traite l'information

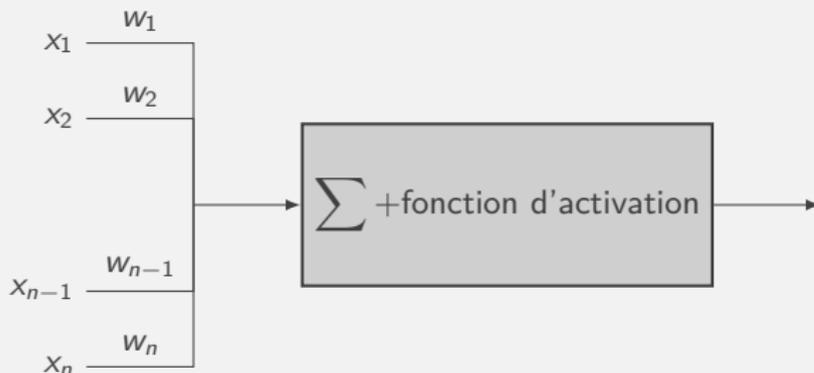
- Aires du cerveau = fonctions cognitives spécifiques : Broca (1861)
- Mise en évidence du **neurone** par coloration : Golgi (1873)
- Mesure non-invasive de l'activité cérébrale : électroencéphalogramme, Berger (1929)
- Imagerie par raisonnement magnétique (progrès récents) + stimulation électrique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

- 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**)...

- 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**)...

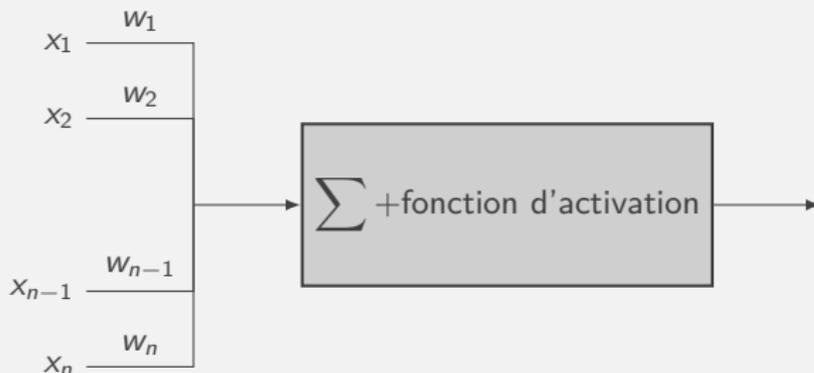


- On sait encore peu de choses sur la manière dont le cerveau fonctionne réellement
- Mais l'observation des neurones a permis de définir le modèle du **neurone artificiel** (e.g. **perceptron**)...



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

- 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

...Où l'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

...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

...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

...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

3 Troisième partie

Perspectives historiques

- 12. Débuts de l'IA (1943–1969)
- 13. L'IA se spécialise (1969–1979)
- 14. L'âge de raison de l'IA (1980–2000)
- 15. L'explosion des données
- 16. L'IA aujourd'hui
- 17. Les grands domaines de l'IA

Débuts de l'IA (1943–1969)

- 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, apprentissage artificiel et par renforcement, algorithmes génétiques.

- Séminaire de deux mois à l'été 1956, sous l'initiative de **John McCarty** (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)

- *Logic Theorist* de Newell et Simon suivi du *General Problem Solver* : raisonnement par décomposition en sous-buts, séparation claire entre résolution et connaissance du problème
- Hypothèse du **système symbolique matériel** : « un système symbolique matériel contient les moyens nécessaires et suffisants pour un comportement généralement intelligent »
- Plusieurs programmes d'IA chez IBM (e.g. *Geometry Theorem Prover*)
- Les **micromondes** de Minsky
- Premiers travaux sur les **perceptrons** (Rosenblatt, 1962).

- *Logic Theorist* de Newell et Simon suivi du *General Problem Solver* : raisonnement par décomposition en sous-buts, séparation claire entre résolution et connaissance du problème
- Hypothèse du **système symbolique matériel** : « un système symbolique matériel contient les moyens nécessaires et suffisants pour un comportement généralement intelligent »
- Plusieurs programmes d'IA chez IBM (e.g. *Geometry Theorem Prover*)
- Les **micromondes** de Minsky
- Premiers travaux sur les **perceptrons** (Rosenblatt, 1962).

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.

L'IA se spécialise (1969–1979)

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

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

- la limite des manipulations syntaxiques dans des domaines tels que la traduction automatique

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

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)

Approches de résolutions des premiers systèmes : application systématique de règles élémentaires pour résoudre un problème complet.

Approches de résolutions des premiers systèmes : application systématique de règles élémentaires pour résoudre un problème complet.

Mais... Rapidement, explosion combinatoire.

Approches de résolutions des premiers systèmes : application systématique de règles élémentaires pour résoudre un problème complet.

Mais... Rapidement, explosion combinatoire.

Idée : restreindre le domaine d'application et injecter beaucoup de connaissance de ce domaine, pour restreindre le champ des inférences et orienter la recherche.

Approches de résolutions des premiers systèmes : application systématique de règles élémentaires pour résoudre un problème complet.

Mais... Rapidement, explosion combinatoire.

Idee : restreindre le domaine d'application et injecter beaucoup de connaissance de ce domaine, pour restreindre le champ des inférences et orienter la recherche.

→ Avènement des **systèmes experts** (e.g. DENDRAL, MYCIN...)

+ développement de Prolog (Colmerauer, 1970)

L'âge de raison de l'IA (1980–2000)

- 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...

- 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

- Retour des réseaux de neurones dans les années 1980
- Algorithme de **rétropropagation**
- La communauté se scinde en deux :
 - Étude des propriétés mathématiques des neurones et des réseaux
 - Mise au point d'architectures efficaces
- 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)

- 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...)

- **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)

L'explosion des données

- 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)

L'IA aujourd'hui

- **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...)

- **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 diagnostique médical**
- **IA prédictive** (embauches, justice...)
- ...

Les grands domaines de l'IA

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

- 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

- Inférence en logique. Dédution, 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

- 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.

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

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

Formalisation de l'intelligence collective et des comportements humains.

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

- 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.

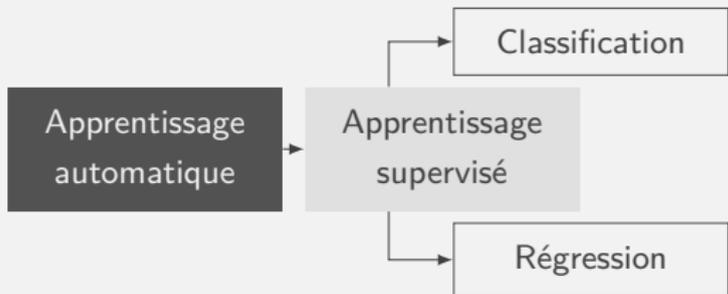
L'apprentissage désigne la faculté à créer de la connaissance, permettant de résoudre des tâches pour lesquels l'algorithme n'a pas été explicitement programmé.

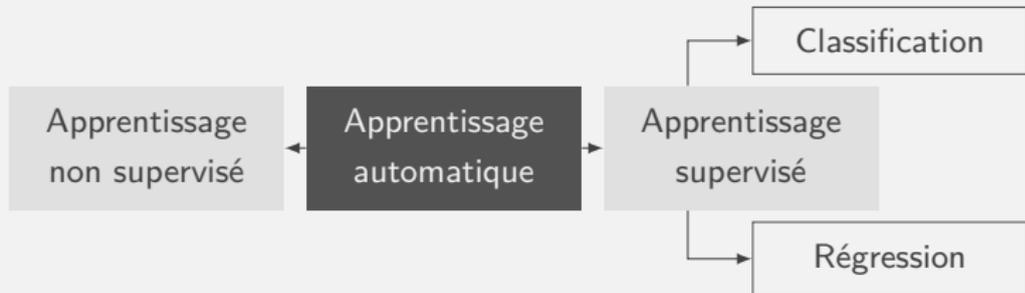
L'apprentissage désigne la faculté à créer de la connaissance, permettant de résoudre des tâches pour lesquels l'algorithme n'a pas été explicitement programmé.

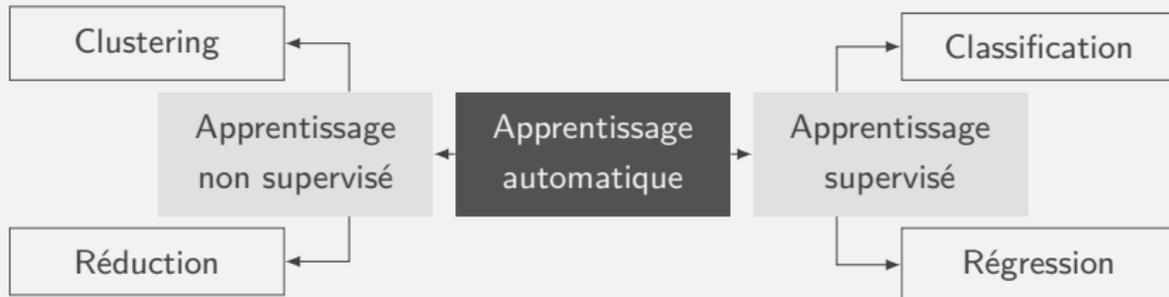
Pour certains scientifiques, il s'agit de l'**essence-même** de l'intelligence.

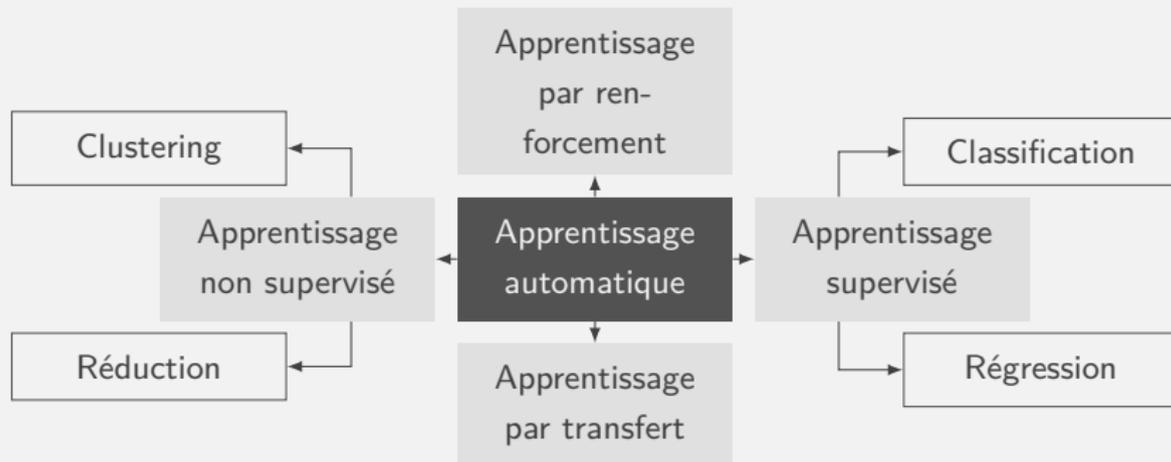
Apprentissage
automatique











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)

4 Quatrième partie

Conclusion

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



Russell, S. J. and Norvig, P. (2010).

Artificial intelligence : a modern approach.

Pearson Education, 3rd edition edition.

...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 ?

...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 ?

...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 ?*

- Travaux de Kate Crawford (cf <https://katecrawford.net/>) + Anatomy of an AI System (<https://anatomyof.ai/>)
- Cathy O'Neil, *Weapons of Math Destruction* (2016)
- Sur ChatGPT à l'Ensimag (<https://www-verimag.imag.fr/~maraninx/blog/chatgpt-pro/>)
- Travaux d'Ecoinfo sur l'impact des systèmes d'IA : <https://ecoinfo.cnrs.fr/category/ia/>
- Valérie d'Acremont, sur les systèmes d'IA dans la santé (https://www.youtube.com/watch?v=oKcy_cYOQ0w)