



## Masterclass IA #1

50 nuances d'IA – Au cœur des modèles d'IA

---

**Sylvain Bouveret**

Laboratoire d'informatique de Grenoble, Ensimag-Grenoble INP, Univ. Grenoble-Alpes

GDR RADIA, Écoinfo

---

Institut G9+, 24 février 2025



# Caveat emptor

D'où je parle...

- Maître de conférences en Informatique (depuis 2011)
- Sujets de recherche historiques : **choix social computationnel**
- Depuis quelques années (essentiellement 2019) : impacts environnementaux et sociétaux du numérique



# Caveat emptor

D'où je parle...

- Maître de conférences en Informatique (depuis 2011)
- Sujets de recherche historiques : **choix social computationnel**
- Depuis quelques années (essentiellement 2019) : impacts environnementaux et sociétaux du numérique
- Mon rapport à l'IA :



# Caveat emptor

D'où je parle...

- Maître de conférences en Informatique (depuis 2011)
- Sujets de recherche historiques : **choix social computationnel**
- Depuis quelques années (essentiellement 2019) : impacts environnementaux et sociétaux du numérique
- Mon rapport à l'IA :
  - Choix social computationnel  $\subset$  IA (symbolique)



# Caveat emptor

D'où je parle...

- Maître de conférences en Informatique (depuis 2011)
- Sujets de recherche historiques : **choix social computationnel**
- Depuis quelques années (essentiellement 2019) : impacts environnementaux et sociétaux du numérique
- Mon rapport à l'IA :
  - Choix social computationnel  $\subset$  IA (symbolique)
  - Membre du GDR RADIA (Raisonnement, Décision et Apprentissage en IA) – <https://gdr-radia.cnrs.fr/>



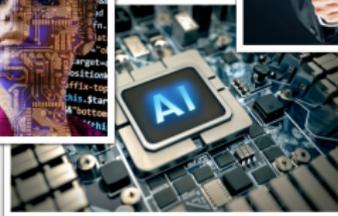
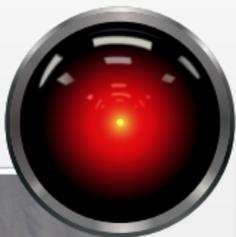
# Caveat emptor

D'où je parle...

- Maître de conférences en Informatique (depuis 2011)
- Sujets de recherche historiques : **choix social computationnel**
- Depuis quelques années (essentiellement 2019) : impacts environnementaux et sociétaux du numérique
- Mon rapport à l'IA :
  - Choix social computationnel  $\subset$  IA (symbolique)
  - Membre du GDR RADIA (Raisonnement, Décision et Apprentissage en IA) – <https://gdr-radia.cnrs.fr/>
  - Intéressé par l'apprentissage profond / de l'IA générative (surtout sous le prisme des impacts environnementaux et sociétaux), mais **non spécialiste** de ces domaines



# Pour débiter...





# Pour débiter...

## Quiz

Qu'est-ce que l'IA ?



Exprimez-vous ici :

<http://quiz.noiraudes.net/>



# Menu du jour...

1. **Qu'est-ce que l'IA ?**
2. **Les racines scientifiques de l'IA**
3. **Perspectives historiques**
4. **IA statistique vs IA symbolique**
5. **Les grands domaines de l'IA**
6. **Conclusion**

## Partie 1

---

Qu'est-ce que l'IA ?



## Les deux dimensions de l'IA

Des systèmes qui pensent  
comme des humains

Des systèmes qui agissent  
comme des humains

Des systèmes qui  
pensent rationnellement

Des systèmes qui agis-  
sent rationnellement



# Les deux dimensions de l'IA

Des systèmes qui pensent  
comme des humains

Des systèmes qui  
pensent rationnellement

Des systèmes qui agissent  
comme des humains

Des systèmes qui agis-  
sent rationnellement

Deux dimensions [Russell and Norvig, 2010] :

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



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

*Artificial intelligence: a modern approach.*

Pearson Education, 3rd edition edition.



# Les deux dimensions de l'IA

## IA forte

Des systèmes qui pensent  
comme des humains

Des systèmes qui  
pensent rationnellement

Des systèmes qui agissent  
comme des humains

Des systèmes qui agis-  
sent rationnellement

## IA faible

Deux dimensions [Russell and Norvig, 2010] :

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



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

*Artificial intelligence: a modern approach.*  
Pearson Education, 3rd edition edition.



## À propos de ChatGPT...

Des systèmes qui pensent  
comme des humains

Des systèmes qui  
pensent rationnellement

Des systèmes qui agissent  
comme des humains

Des systèmes qui agis-  
sent rationnellement

### Quiz

À votre avis, où se situe un système comme ChatGPT dans cette nomenclature ?



Exprimez-vous ici :  
<http://quiz.noiraudes.net/>

## 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



## Le 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.



# Le 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.



## Le 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**.



## Le 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**.

cf 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).



# 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
- ...

## Des systèmes qui pensent 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 agis-  
sent rationnellement



## IA forte et conscience

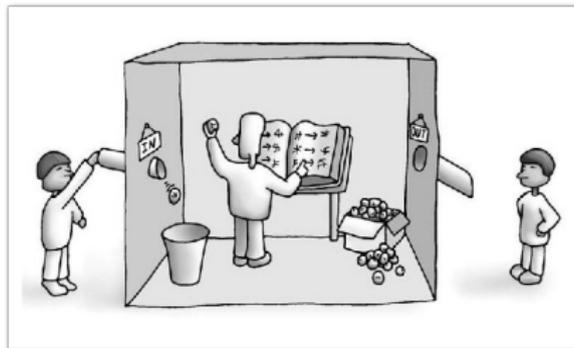
Argument de la **conscience** : « Une machine ne peut pas être dite intelligente si elle n'est pas consciente de ses actes »



## 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 agis-  
sent rationnellement



## De la stupidité humaine

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 ?



## Le modèle de la logique

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



## Le modèle de la logique

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)



# Le modèle de la logique

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 agissent  
comme des humains

Des systèmes qui  
pensent rationnellement

Des systèmes qui agissent  
rationnellement



# 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.



# Une définition de l'IA

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.



# L'IA selon l'UE

Définition d'un « Système d'intelligence artificielle » selon l'Union Européenne (AI Act, 2024) :

## Quiz

À votre avis, qu'est-ce qui caractérise un système d'IA pour l'union européenne ?



Exprimez-vous ici :

<http://quiz.noiraudes.net/>



## L'IA selon l'UE

Définition d'un « Système d'intelligence artificielle » selon l'Union Européenne (AI Act, 2024) :

'AI system' means a machine-based system that is designed to operate with *varying levels of autonomy* and that may exhibit *adaptiveness* after deployment, and that, for explicit or implicit objectives, *infers*, from the input it receives, how to generate *outputs* such as predictions, content, recommendations, or decisions *that can influence physical or virtual environments*<sup>a</sup>.

<sup>a</sup><https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689#d1e2090-1-1> (emphase ajoutée)



## L'IA selon l'UE

Définition d'un « Système d'intelligence artificielle » selon l'Union Européenne (AI Act, 2024) :

'AI system' means a machine-based system that is designed to operate with *varying levels of autonomy* and that may exhibit *adaptiveness* after deployment, and that, for explicit or implicit objectives, *infers*, from the input it receives, how to generate *outputs* such as predictions, content, recommendations, or decisions *that can influence physical or virtual environments*<sup>a</sup>.

⇒ pas de frontière claire entre programme informatique et système d'IA...

<sup>a</sup><https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689#d1e2090-1-1> (emphase ajoutée)



## Les chercheurs et le test de Turing

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



## Les chercheurs et 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)



## Les chercheurs et 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 ?

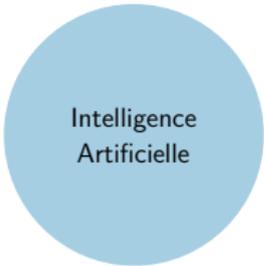
## Partie 2

---

### Les racines scientifiques de l'IA



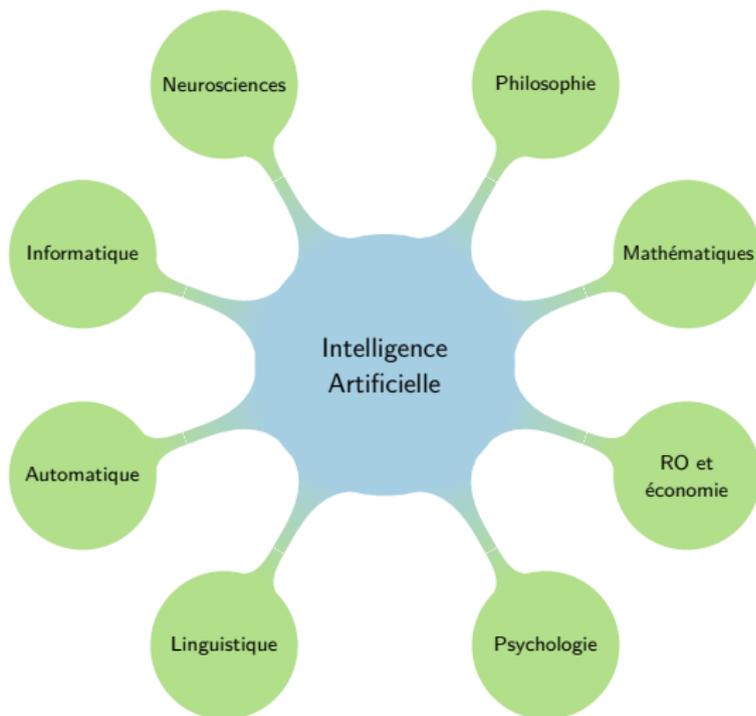
# Les racines de l'IA



Intelligence  
Artificielle

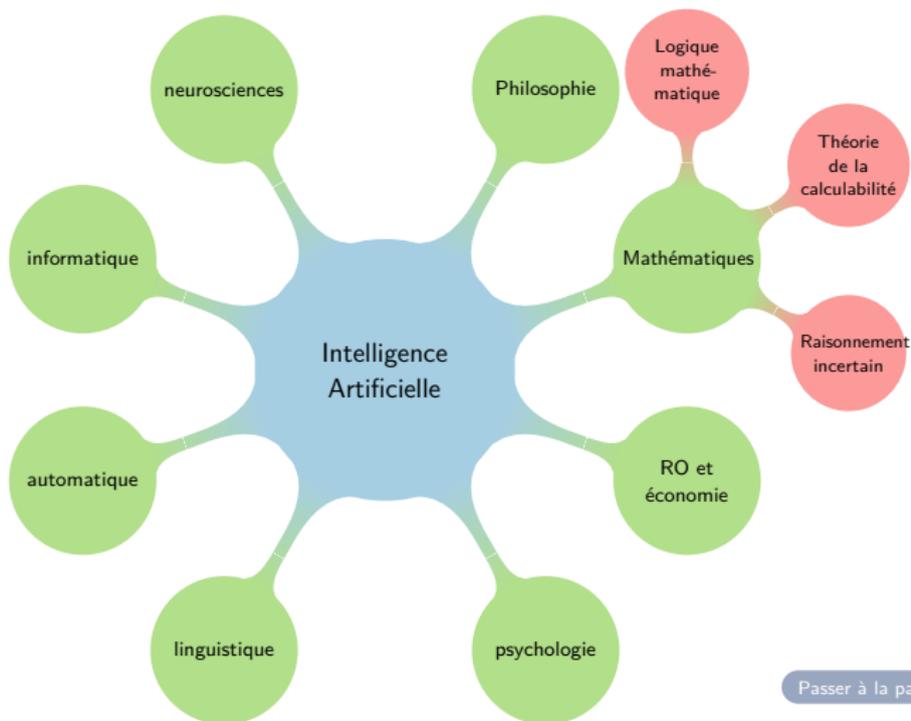


# Les racines de l'IA





# Les racines de l'IA



Passer à la partie suivante...

# **Philosophie**



## Les grandes questions philosophiques

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

# Mathématiques

–

La logique mathématique



# La logique mathématique, les racines

...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)



## 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 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**



# 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

# Mathématiques

–

La théorie de la calculabilité



# Théorie de la calculabilité

...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<sup>er</sup> ordre de Frege
  - **théorème d'incomplétude**



## Théorème de Gödel (1931)

- 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)



# Calculabilité selon Church-Turing

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



# Calculabilité selon Church-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.



# Calculabilité selon Church-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 :

- $\lambda$ -calcul de Church (racines de la programmation fonctionnelle)
- Fonctions généralement récursives de Gödel



## Une fonction non calculable...

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  $\mathbb{N}$



## Une fonction non calculable...

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  $\mathbb{N}$



## Une fonction non calculable...

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  $\varphi_i(a)$ )



## Une fonction non calculable...

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



# Une fonction non calculable...

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
- $\varphi_i$  est calculable (il suffit d'appliquer  $\mathcal{M}_i$  à l'argument et d'attendre  $n$  étapes)



# Une fonction non calculable...

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
- $\varphi_i$  est calculable (il suffit d'appliquer  $\mathcal{M}_i$  à l'argument et d'attendre  $n$  étapes)
- La fonction suivante n'est pas calculable :

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



# Complexité et tractabilité

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



## Complexité et tractabilité

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.

# Mathématiques

–

Le raisonnement incertain



## Raisonnement incertain

...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**.



# Au-delà des probabilités : les possibilités

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



## Au-delà des probabilités : les possibilités

- 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)



# Au-delà des probabilités : les possibilités

- 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



# Théorie de la décision

...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).



# Décision collective et théorie des jeux

...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)



# 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

...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.



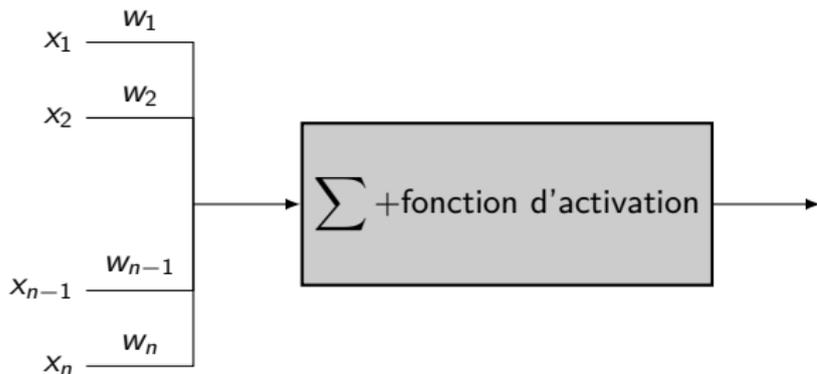
# Approches connexionnistes

- 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

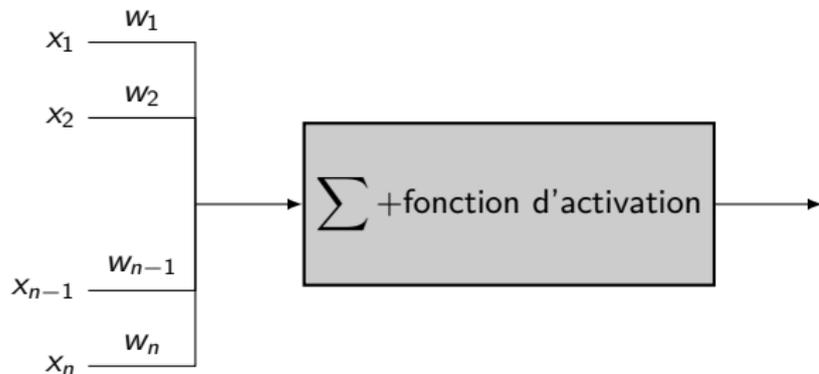
- On sait encore peu de choses sur la manière dont le cerveau fonctionne réellement
- Mais l'observation des neurones a permis de définir le modèle du **neurone artificiel** (e.g. **perceptron**)...





# 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**)...

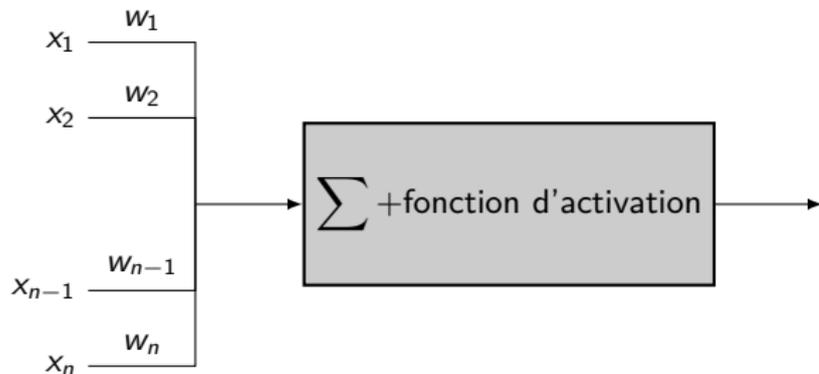


- ...Et les approches connexionnistes (réseaux de neurones)



# 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

...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**



# 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

## Partie 3

---

### Perspectives historiques



## Avant de débiter...

### Quiz

Qui peut être considéré comme le père du terme "Intelligence Artificielle" ?



Exprimez-vous ici :  
<http://quiz.noiraudes.net/>



## Avant de débiter...

### Quiz

Quand a été inventé le premier algorithme de classification par perceptron ?



Exprimez-vous ici :  
<http://quiz.noiraudes.net/>



## Avant de débiter...

### Quiz

Quand l'algorithme de rétropropagation pour des réseaux de neurones multicouches a-t-il été redécouvert par la communauté IA ?



Exprimez-vous ici :  
<http://quiz.noiraudes.net/>

## **Début de l'IA (1943–1969)**



## Les travaux fondateurs

- 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.



# La conférence de Dartmouth

- 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)



## L'enthousiasme des débuts (1952–1969)

- *Logic Theorist* de Newell et Simon suivi du *General Problem Solver* : raisonnement par décomposition en sous-but, 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).



## L'enthousiasme des débuts (1952–1969)

- *Logic Theorist* de Newell et Simon suivi du *General Problem Solver* : raisonnement par décomposition en sous-but, 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)**



# La désillusion

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



# La désillusion

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

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



# La désillusion

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 désillusion

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)



# Les systèmes experts

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



# Les systèmes experts

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

Mais... Rapidement, explosion combinatoire.



# Les systèmes experts

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.



# Les systèmes experts

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.

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

+ développement de Prolog (Colmerauer, 1970)

## **L'âge de raison de l'IA (1980–2000)**



# 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...



# 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



# Les réseaux de neurones

- 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)



# Les agents

- 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...)



# 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)

# **L'explosion des données et des capacités de calcul (2000–)**



# Les données et le renouveau de l'IA

- Avènement du web et des systèmes connectés → gigantesques corpus 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)



# 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)



# 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)
- ...précurseurs du renouveau des **réseaux de neurones**, et en particulier des **réseaux de neurones profonds**, permis par :



# Les données et le renouveau de l'IA

- Avènement du web et des systèmes connectés → gigantesques corpus de données
- Premier exemple : désambiguïstation (Yarowsky, 1995)
- Watson (IBM, 2011)
- ...précurseurs du renouveau des réseaux de neurones, et en particulier des réseaux de neurones profonds, permis par :
  - la disponibilité de grands corpus de données (→ entraînement)



# 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)
- ...précurseurs du renouveau des **réseaux de neurones**, et en particulier des **réseaux de neurones profonds**, permis par :
  - la disponibilité de grands corpus de données (→ entraînement)
  - l'augmentation considérable des capacités de calcul (en particulier GPU)



# L'ère du deep learning

- **AlexNet, 2012**, réseau 8 couches (5 convolutives, 3 entièrement connectées), vainqueur d'ImageNet 2012 [Krizhevsky et al., 2012]
- **Transformers, 2017**, pour le TALN [Vaswani et al., 2017], qui a permis l'entraînement des grands modèles de langue (LLM)
- Avènement de l'IA générative et démocratisation des outils (ChatGPT : 1 million d'utilisateurs en 5 jours<sup>1</sup>)



**Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).**

Imagenet classification with deep convolutional neural networks.  
*Advances in neural information processing systems*, 25.



**Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017).**

Attention is all you need.  
*Advances in neural information processing systems*, 30.

<sup>1</sup><https://fr.statista.com/infographie/29187/temps-qu-il-a-fallu-a-certains-services-plateformes-en-ligne-pour-atteindre-1->

# **L'IA aujourd'hui**



# Les applications les plus visibles

L'IA aujourd'hui...



# Les applications les plus visibles

## 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...)

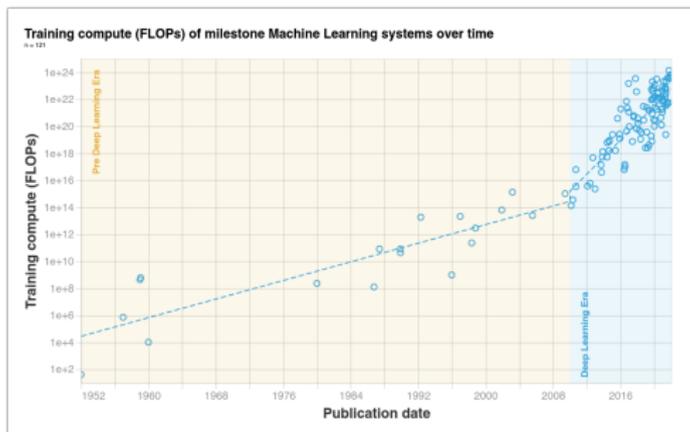


## Les applications les plus visibles (2)

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



# Tendances observées dans le ML



(sur la période bleue, doublement tous les 5,6 mois environ)



Sevilla, J., Heim, L., Ho, A., Besiroglu, T., Hobbahn, M., and Villalobos, P. (2022).

Compute Trends Across Three Eras of Machine Learning.

*arXiv:2202.05924 [cs]*.



# Efficacité vs exponentielle

Un petit exercice de pensée :

- chaque FLOP consomme  $E$  joules
- début 2025, le nombre moyen de FLOPS effectué par l'entraînement d'une machine est de  $N$  (donc consommation totale  $N \times E$  joules)
- courant 2025, une innovation de rupture permettant de diviser par 10 la consommation unitaire d'une FLOP arrive



# Efficacité vs exponentielle

Un petit exercice de pensée :

- chaque FLOP consomme  $E$  joules
- début 2025, le nombre moyen de FLOPS effectué par l'entraînement d'une machine est de  $N$  (donc consommation totale  $N \times E$  joules)
- courant 2025, une innovation de rupture permettant de diviser par 10 la consommation unitaire d'une FLOP arrive

## Quiz

En quelle année aura-t-on à nouveau dépassé la consommation totale  $NE$  joules si les tendances se poursuivent ?



Exprimez-vous ici :

<http://quiz.noiraudes.net/>

## Partie 4

---

IA statistique vs IA symbolique



## Avant de débiter cette partie...

### Quiz

Connaissez-vous la différence entre IA statistique et IA symbolique ?



Exprimez-vous ici :

<http://quiz.noiraudes.net/>



## Avant de débiter cette partie...

### Quiz

Connaissez-vous la différence entre IA statistique et IA symbolique ?



Exprimez-vous ici :

<http://quiz.noiraudes.net/>

Pour illustrer cette différence, nous allons nous pencher sur la question :  
*ChatGPT* est-il capable de raisonner logiquement ?

# **IA statistique et logique**



## Le paradoxe du jardinier...

Bonjour ChatGPT. J'ai un problème de logique. Dans mon village, le jardinier taille les arbres de toutes les personnes qui ne peuvent pas le faire elles-mêmes. Est-ce que le jardinier taille ses propres arbres ?



# Le paradoxe du jardinier...

Bonjour ! Ah, cette question me rappelle le fameux paradoxe du barbier, formulé par Bertrand Russell. Le problème que vous décrivez est un cas classique de paradoxe logique. Regardons cela de plus près :

## Énoncé

- Le jardinier taille les arbres de **tous ceux qui ne taillent pas leurs propres arbres**.
- La question est : **taille-t-il ses propres arbres ?**

## Raisonnement

1. **Supposons qu'il taille ses propres arbres.**
  - S'il taille ses propres arbres, alors il ne fait plus partie des personnes qui ne taillent pas leurs arbres.
  - Donc, il ne devrait pas tailler ses propres arbres. Contradiction.
2. **Supposons qu'il ne taille pas ses propres arbres.**
  - S'il ne taille pas ses propres arbres, alors il appartient au groupe des personnes dont les arbres doivent être taillés par le jardinier.
  - Donc, il devrait tailler ses propres arbres. Encore une contradiction.

## Conclusion

Ce paradoxe illustre une situation où les règles définies sont contradictoires et ne permettent pas de résoudre la question de manière cohérente. Ce n'est pas un problème de logique "classique" à résoudre, mais plutôt une manière de montrer les limites de certains systèmes de raisonnement.

C'est un casse-tête intéressant, n'est-ce pas ? 😊



# Le paradoxe du jardinier...

Bonjour ! Ah, cette question me rappelle le fameux paradoxe du barbier, formulé par Bertrand Russell. Le problème que vous décrivez est un cas classique de paradoxe logique. Regardons cela de plus près :

## Énoncé

- Le jardinier taille les arbres de **tous ceux qui ne taillent pas leurs propres arbres**.
- La question est : **taille-t-il ses propres arbres ?**

## Raisonnement

1. **Supposons qu'il taille ses propres arbres.**
  - S'il taille ses propres arbres, alors il ne fait plus partie des personnes qui ne taillent pas leurs arbres.
  - Donc, il ne devrait pas tailler ses propres arbres. Contradiction.
2. **Supposons qu'il ne taille pas ses propres arbres.**
  - S'il ne taille pas ses propres arbres, alors il appartient au groupe des personnes dont les arbres doivent être taillés par le jardinier.
  - Donc, il devrait tailler ses propres arbres. Encore une contradiction.

## Conclusion

Ce paradoxe illustre une situation où les règles définies sont contradictoires et ne permettent pas

Peut-on pour autant en déduire que ChatGPT possède une forme de raisonnement logique formalisé ?



# À propos de génération de textes

- La plupart des algorithmes de génération de textes s'appuient sur des **modèles de langue**



## À propos de génération de textes

- La plupart des algorithmes de génération de textes s'appuient sur des **modèles de langue**
- **Modèle de langue** = distribution de probabilité sur les phrases d'une langue : par ex. « Le chat mange ses croquettes » est (beaucoup) plus probable que « Chien ses chat polymorphisme pédaler avec »



## À propos de génération de textes

- La plupart des algorithmes de génération de textes s'appuient sur des **modèles de langue**
- **Modèle de langue** = distribution de probabilité sur les phrases d'une langue : par ex. « Le chat mange ses croquettes » est (beaucoup) plus probable que « Chien ses chat polymorphisme pédaler avec »
- Comment utiliser ces modèles de langue ?



## À propos de génération de textes

- La plupart des algorithmes de génération de textes s'appuient sur des **modèles de langue**
- **Modèle de langue** = distribution de probabilité sur les phrases d'une langue : par ex. « Le chat mange ses croquettes » est (beaucoup) plus probable que « Chien ses chat polymorphisme pédaler avec »
- Comment utiliser ces modèles de langue ?
  - Prédiction





# À propos de génération de textes

- La plupart des algorithmes de génération de textes s'appuient sur des **modèles de langue**
- **Modèle de langue** = distribution de probabilité sur les phrases d'une langue : par ex. « Le chat mange ses croquettes » est (beaucoup) plus probable que « Chien ses chat polymorphisme pédaler avec »
- Comment utiliser ces modèles de langue ?

- Prédiction



- Masquage





## À propos de génération de textes

- La plupart des algorithmes de génération de textes s'appuient sur des **modèles de langue**
- **Modèle de langue** = distribution de probabilité sur les phrases d'une langue : par ex. « Le chat mange ses croquettes » est (beaucoup) plus probable que « Chien ses chat polymorphisme pédaler avec »
- Comment utiliser ces modèles de langue ?

- Prédiction



- Masquage



- Types de modèles de langue : Markoviens, réseaux de neurones entraînés par **apprentissage supervisé**



# 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

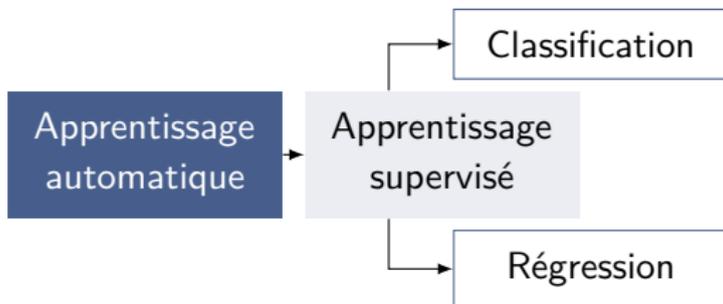
Apprentissage  
automatique



Apprentissage  
supervisé

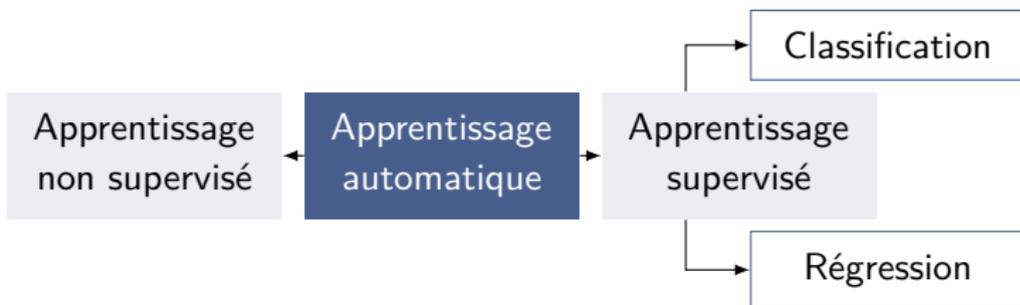


# Différents domaines de l'apprentissage



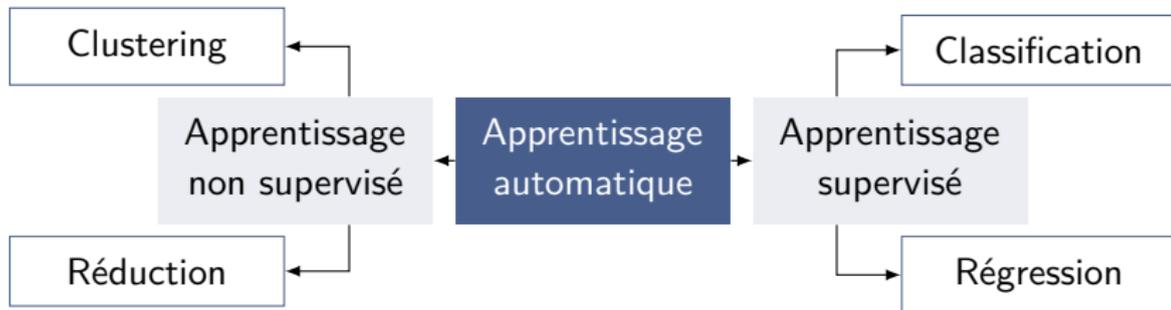


# Différents domaines de l'apprentissage



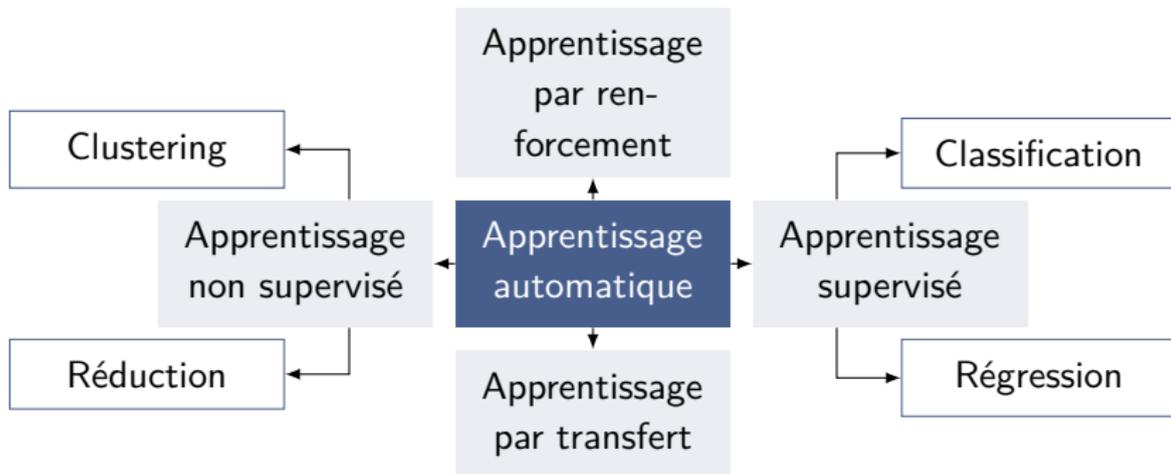


# Différents domaines de l'apprentissage





# Différents domaines de l'apprentissage





# À propos d'apprentissage supervisé

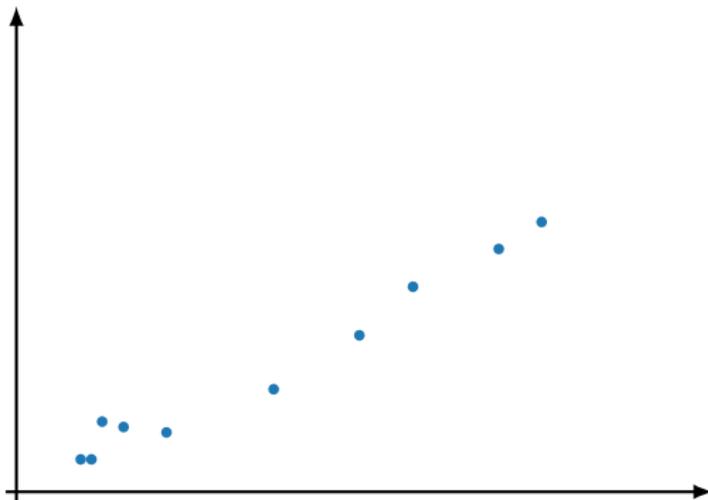
La régression :





# À propos d'apprentissage supervisé

La régression :

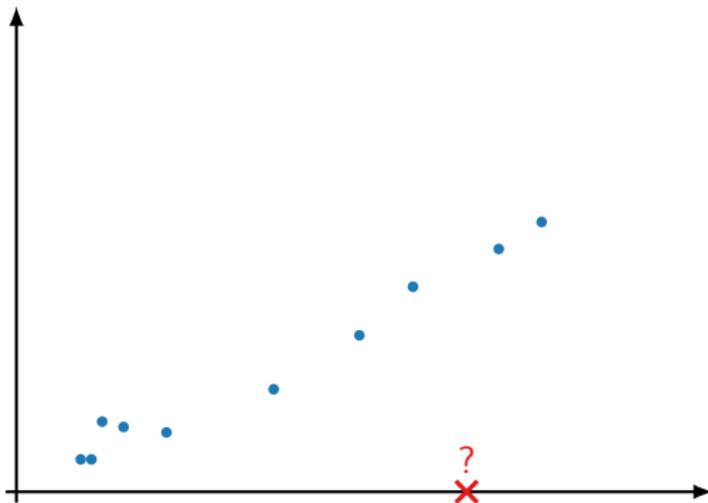


- Exemples



# À propos d'apprentissage supervisé

La régression :

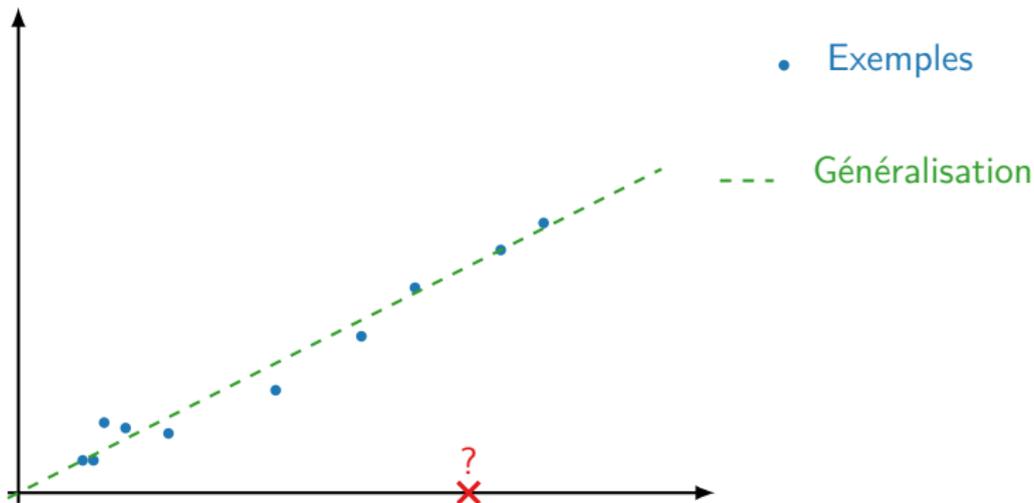


- Exemples



# À propos d'apprentissage supervisé

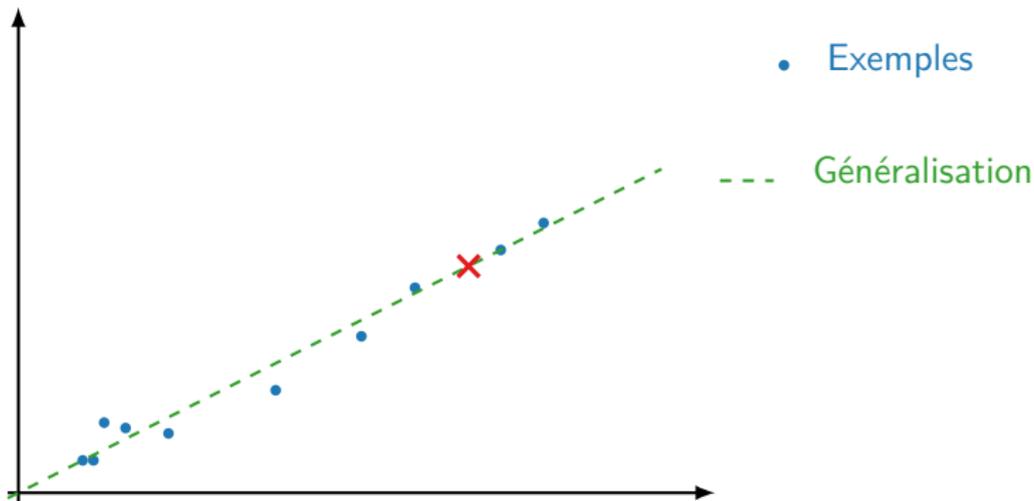
La régression :





# À propos d'apprentissage supervisé

La régression :





# À propos d'apprentissage supervisé

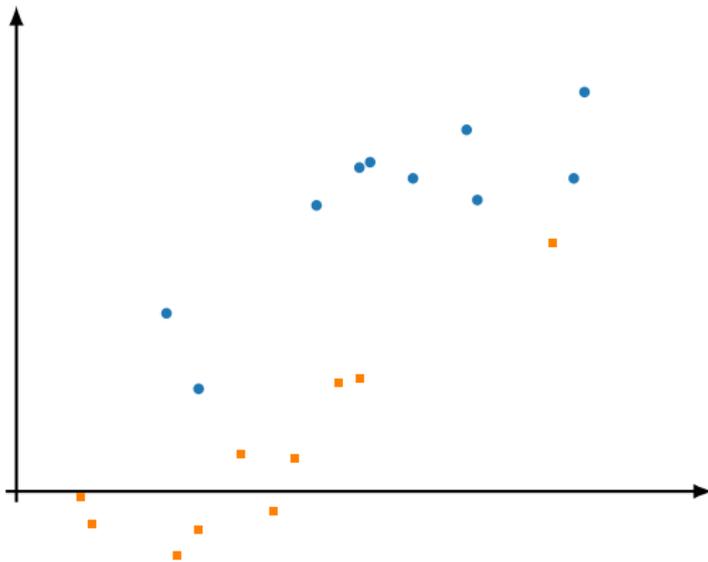
La classification :





# À propos d'apprentissage supervisé

La classification :

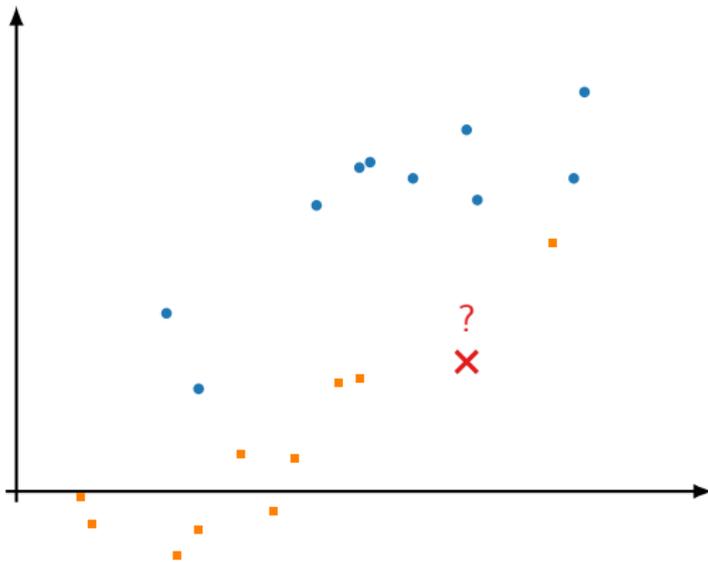


● ■ Exemples (2 classes)



# À propos d'apprentissage supervisé

La classification :

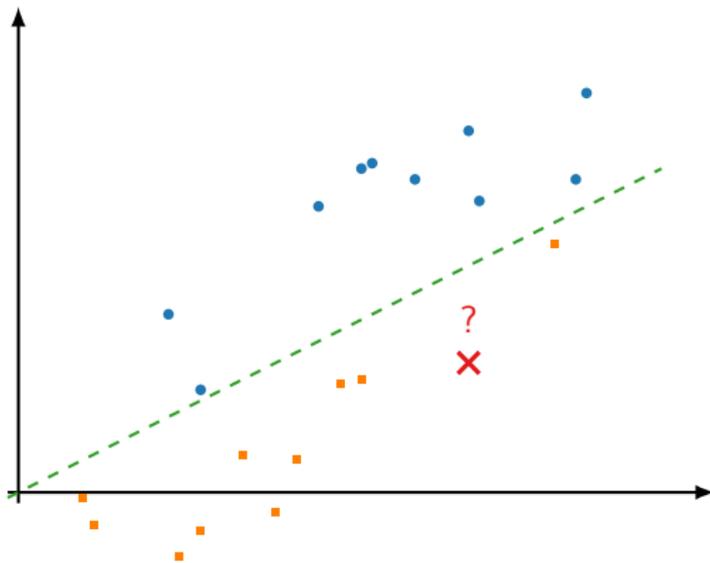


● ■ Exemples (2 classes)



# À propos d'apprentissage supervisé

La classification :



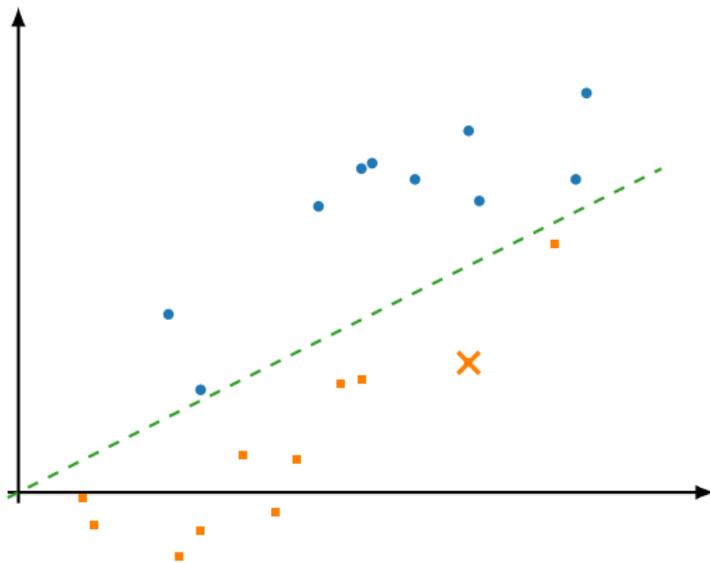
● ■ Exemples (2 classes)

--- Généralisation



# À propos d'apprentissage supervisé

La classification :



● ■ Exemples (2 classes)

--- Généralisation



# Apprentissage et génération de texte

- La plupart des modèles de langue actuels ont été calculés par des algorithmes de **classification** dont l'objectif est de prédire le **mot suivant** à partir des **mots précédents**. Vision simplifiée :



# Apprentissage et génération de texte

- La plupart des modèles de langue actuels ont été calculés par des algorithmes de **classification** dont l'objectif est de prédire le **mot suivant** à partir des **mots précédents**. Vision simplifiée :
  - **Entrée** : codage des mots précédents de la phrase



# Apprentissage et génération de texte

- La plupart des modèles de langue actuels ont été calculés par des algorithmes de **classification** dont l'objectif est de prédire le **mot suivant** à partir des **mots précédents**. Vision simplifiée :
  - **Entrée** : codage des mots précédents de la phrase
  - **Sortie** : une probabilité associée à chaque mot suivant possible (un mot = une classe)



# Apprentissage et génération de texte

- La plupart des modèles de langue actuels ont été calculés par des algorithmes de **classification** dont l'objectif est de prédire le **mot suivant** à partir des **mots précédents**. Vision simplifiée :
  - **Entrée** : codage des mots précédents de la phrase
  - **Sortie** : une probabilité associée à chaque mot suivant possible (un mot = une classe)
- Ces probabilités ont été apprises en analysant un corpus de textes (sites web, extraits de conversations, livres, etc.)



# Apprentissage et génération de texte

- La plupart des modèles de langue actuels ont été calculés par des algorithmes de **classification** dont l'objectif est de prédire le **mot suivant** à partir des **mots précédents**. Vision simplifiée :
  - **Entrée** : codage des mots précédents de la phrase
  - **Sortie** : une probabilité associée à chaque mot suivant possible (un mot = une classe)
- Ces probabilités ont été apprises en analysant un corpus de textes (sites web, extraits de conversations, livres, etc.)
- Les capacités de **généralisation** de l'algorithme d'apprentissage font qu'il est capable de prédire un mot même à partir d'une phrase qu'il n'a jamais rencontrée auparavant



# Apprentissage et génération de texte

- La plupart des modèles de langue actuels ont été calculés par des algorithmes de **classification** dont l'objectif est de prédire le **mot suivant** à partir des **mots précédents**. Vision simplifiée :
  - **Entrée** : codage des mots précédents de la phrase
  - **Sortie** : une probabilité associée à chaque mot suivant possible (un mot = une classe)
- Ces probabilités ont été apprises en analysant un corpus de textes (sites web, extraits de conversations, livres, etc.)
- Les capacités de **généralisation** de l'algorithme d'apprentissage font qu'il est capable de prédire un mot même à partir d'une phrase qu'il n'a jamais rencontrée auparavant
- Lorsque l'on pose une question à un chatbot, il génère donc les mots les uns après les autres en se servant du prompt + du début de la phrase générée comme entrée de l'algorithme



---

# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :



---

# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :
  1. **Apprentissage :**



# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :
  1. **Apprentissage :**
    - On fournit à l'algorithme un certain nombre d'exemples étiquetés. Pour le chatbot, ce sont les phrases extraites des corpus de textes.



# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :
  1. **Apprentissage :**
    - On fournit à l'algorithme un certain nombre d'exemples étiquetés. Pour le chatbot, ce sont les phrases extraites des corpus de textes.
    - Cette phase peut être exécutée hors-ligne.



# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :
  1. **Apprentissage :**
    - On fournit à l'algorithme un certain nombre d'exemples étiquetés. Pour le chatbot, ce sont les phrases extraites des corpus de textes.
    - Cette phase peut être exécutée hors-ligne.
  2. **Inférence :**



# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :
  1. **Apprentissage :**
    - On fournit à l'algorithme un certain nombre d'exemples étiquetés. Pour le chatbot, ce sont les phrases extraites des corpus de textes.
    - Cette phase peut être exécutée hors-ligne.
  2. **Inférence :**
    - On fournit à l'algorithme un nouvel exemple (éventuellement jamais vu). Pour le chatbot, c'est le prompt + le début de la phrase déjà générée.



# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :
  1. **Apprentissage :**
    - On fournit à l'algorithme un certain nombre d'exemples étiquetés. Pour le chatbot, ce sont les phrases extraites des corpus de textes.
    - Cette phase peut être exécutée hors-ligne.
  2. **Inférence :**
    - On fournit à l'algorithme un nouvel exemple (éventuellement jamais vu). Pour le chatbot, c'est le prompt + le début de la phrase déjà générée.
    - L'algorithme calcule la classe de sortie en fonction de ce qu'il a appris. Pour le chatbot, c'est le mot suivant le plus probable.



# Apprentissage vs inférence

- Les deux phases d'exécution d'un algorithme d'apprentissage :
  1. **Apprentissage :**
    - On fournit à l'algorithme un certain nombre d'exemples étiquetés. Pour le chatbot, ce sont les phrases extraites des corpus de textes.
    - Cette phase peut être exécutée hors-ligne.
  2. **Inférence :**
    - On fournit à l'algorithme un nouvel exemple (éventuellement jamais vu). Pour le chatbot, c'est le prompt + le début de la phrase déjà générée.
    - L'algorithme calcule la classe de sortie en fonction de ce qu'il a appris. Pour le chatbot, c'est le mot suivant le plus probable.
    - Cette phase a lieu en ligne, par interaction avec l'utilisateur.



# Réseaux de neurones

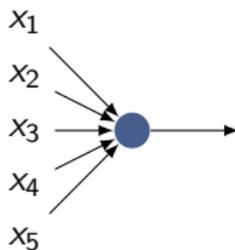
- Les algorithmes de classification les plus efficaces à l'heure actuelle (sur du texte, des images, etc.) s'appuient sur des **réseaux de neurones profonds**



# Réseaux de neurones

- Les algorithmes de classification les plus efficaces à l'heure actuelle (sur du texte, des images, etc.) s'appuient sur des **réseaux de neurones profonds**

À la base, un modèle de neurone, le **perceptron**...

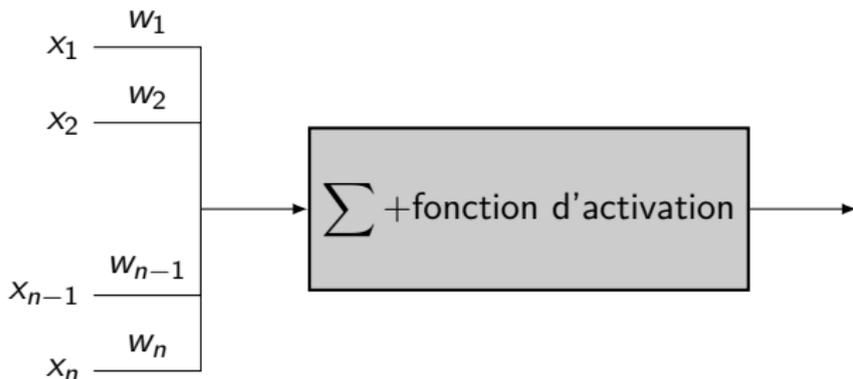




# Réseaux de neurones

- Les algorithmes de classification les plus efficaces à l'heure actuelle (sur du texte, des images, etc.) s'appuient sur des **réseaux de neurones profonds**

À la base, un modèle de neurone, le **perceptron**...

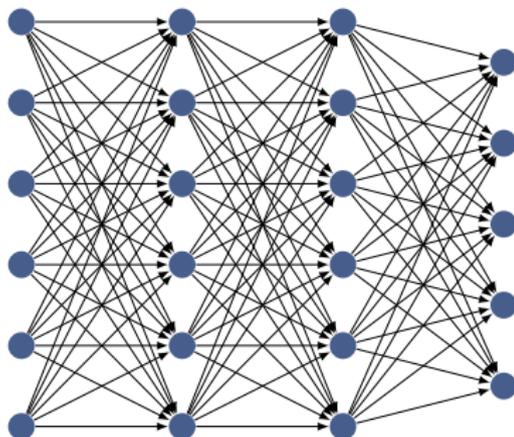




# Réseaux de neurones

- Les algorithmes de classification les plus efficaces à l'heure actuelle (sur du texte, des images, etc.) s'appuient sur des **réseaux de neurones profonds**

On organise tout ça en **réseau multicouche** pour pouvoir apprendre des fonctions non linéaires :

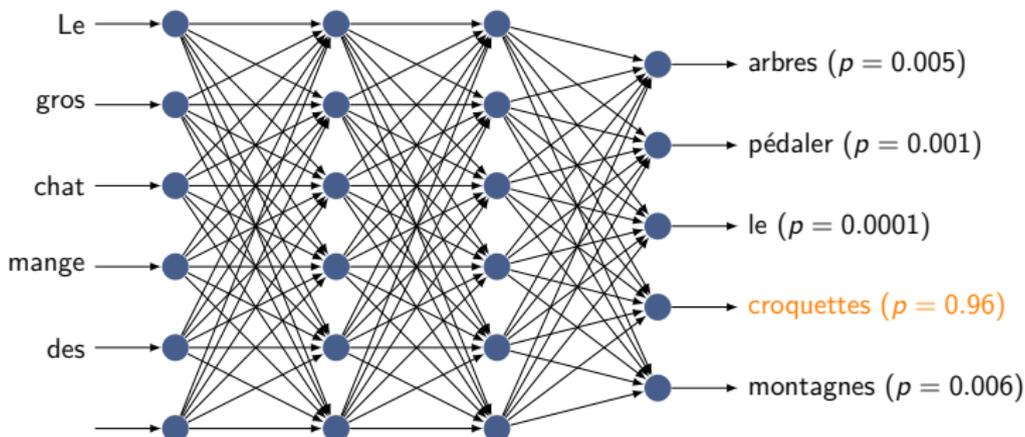




# Réseaux de neurones

- Les algorithmes de classification les plus efficaces à l'heure actuelle (sur du texte, des images, etc.) s'appuient sur des **réseaux de neurones profonds**

On organise tout ça en **réseau multicouche** pour pouvoir apprendre des fonctions non linéaires :





# Réseaux de neurones

- Les algorithmes de classification les plus efficaces à l'heure actuelle (sur du texte, des images, etc.) s'appuient sur des **réseaux de neurones profonds**
- Ce qui est appris dans les réseaux de neurone est l'ensemble des **poids** portés par les connexions entre neurones.



---

# Modèle, architecture

- Selon la tâche à réaliser, plusieurs architectures de réseaux sont possibles



# Modèle, architecture

- Selon la tâche à réaliser, plusieurs architectures de réseaux sont possibles
- Définitions:



# Modèle, architecture

- Selon la tâche à réaliser, plusieurs **architectures** de réseaux sont possibles
- Définitions:
  - **Architecture** : organisation des neurones entre eux (connexions entre neurones) et opérations réalisées dans chaque neurone



# Modèle, architecture

- Selon la tâche à réaliser, plusieurs **architectures** de réseaux sont possibles
- Définitions:
  - **Architecture** : organisation des neurones entre eux (connexions entre neurones) et opérations réalisées dans chaque neurone
  - **Checkpoint** : poids chargés dans une architecture donnée



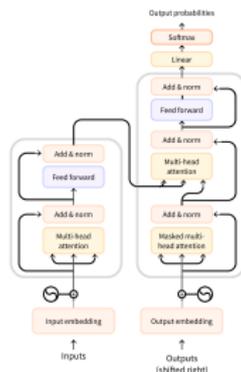
# Modèle, architecture

- Selon la tâche à réaliser, plusieurs **architectures** de réseaux sont possibles
- Définitions:
  - **Architecture** : organisation des neurones entre eux (connexions entre neurones) et opérations réalisées dans chaque neurone
  - **Checkpoint** : poids chargés dans une architecture donnée
  - **Modèle** : terme ambigu qui peut désigner, selon le contexte, une architecture (sans les poids), ou un ensemble architecture + checkpoint



# Architecture transformer

- Pour la génération de texte, l'architecture la plus performante à l'heure actuelle est l'architecture **transformer** [Vaswani et al., 2017]



Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017).  
Attention is all you need.

*Advances in neural information processing systems*, 30.



## Quelques ordres de grandeurs

- **GPT-3 (OpenAI) :**
  - $1,75 \times 10^{11}$  paramètres (= 175 milliards), 96 couches
  - 32000 mots (tokens) en entrée
  - 12000 dimensions dans l'espace des mots
- **Llama 3.1 (Meta) :**
  - 405 milliards de paramètres, 126 couches
  - 128000 mots (tokens) en entrée
  - coût d'entraînement : 440 000 petaFLOP-jour (=  $440 \times 10^{18}$  FLOP/sec  $\times$  1 jour  $\rightarrow 3,8016 \times 10^{40}$  FLOP)



## Affinage de modèle

- La plupart des LLM sont **génériques** (dédiés à des tâches textuelles générales)



## Affinage de modèle

- La plupart des LLM sont **génériques** (dédiés à des tâches textuelles générales)
- Pour des tâches spécifiques (typiquement sur un corpus particulier), besoin d'un entraînement spécifique



## Affinage de modèle

- La plupart des LLM sont **génériques** (dédiés à des tâches textuelles générales)
- Pour des tâches spécifiques (typiquement sur un corpus particulier), besoin d'un entraînement spécifique
- Méthode générale :



## Affinage de modèle

- La plupart des LLM sont **génériques** (dédiés à des tâches textuelles générales)
- Pour des tâches spécifiques (typiquement sur un corpus particulier), besoin d'un entraînement spécifique
- Méthode générale :
  - on part d'un LLM générique (architecture + checkpoint)



# Affinage de modèle

- La plupart des LLM sont **génériques** (dédiés à des tâches textuelles générales)
- Pour des tâches spécifiques (typiquement sur un corpus particulier), besoin d'un entraînement spécifique
- Méthode générale :
  - on part d'un LLM générique (architecture + checkpoint)
  - on relance un entraînement sur le corpus particulier



## Affinage de modèle

- La plupart des LLM sont **génériques** (dédiés à des tâches textuelles générales)
- Pour des tâches spécifiques (typiquement sur un corpus particulier), besoin d'un entraînement spécifique
- Méthode générale :
  - on part d'un LLM générique (architecture + checkpoint)
  - on relance un entraînement sur le corpus particulier
  - mais en général sans entraîner tout le réseau : on n'entraîne que les couches de sortie (et on fige le reste des poids)



# Affinage de modèle

- La plupart des LLM sont **génériques** (dédiés à des tâches textuelles générales)
- Pour des tâches spécifiques (typiquement sur un corpus particulier), besoin d'un entraînement spécifique
- Méthode générale :
  - on part d'un LLM générique (architecture + checkpoint)
  - on relance un entraînement sur le corpus particulier
  - mais en général sans entraîner tout le réseau : on n'entraîne que les couches de sortie (et on fige le reste des poids)
  - cette technique est appelée **affinage de modèle** (*fine-tuning*)



# Chatbot et modèles de langue

- Un modèle de langue (dont LLM) ne fait que de la prédiction de mot suivant (et donc peut faire de la génération de texte)



# Chatbot et modèles de langue

- Un modèle de langue (dont LLM) ne fait que de la prédiction de mot suivant (et donc peut faire de la génération de texte)
- Un chatbot, c'est un peu plus que ça. Il faut en plus, *a minima* :



# Chatbot et modèles de langue

- Un modèle de langue (dont LLM) ne fait que de la prédiction de mot suivant (et donc peut faire de la génération de texte)
- Un chatbot, c'est un peu plus que ça. Il faut en plus, *a minima* :
  - être capable de générer des dialogues **réalistes**. Pour cela, on ajoute en général une couche algorithmique supplémentaire : le RLHF (*Reinforcement Learning on Human Feedback*)



# Chatbot et modèles de langue

- Un modèle de langue (dont LLM) ne fait que de la prédiction de mot suivant (et donc peut faire de la génération de texte)
- Un chatbot, c'est un peu plus que ça. Il faut en plus, *a minima* :
  - être capable de générer des dialogues **réalistes**. Pour cela, on ajoute en général une couche algorithmique supplémentaire : le RLHF (*Reinforcement Learning on Human Feedback*)
  - enrober ça dans une interface (web par exemple)



## Conclusion sur ChatGPT...

- Cœur de l'algorithme : génération **mot à mot**
- Modèle de langue déterminé par classification supervisée, par **réseau de neurones profond**, affiné par apprentissage par renforcement
- Il n'y a *a priori* rien de plus, donc en particulier pas de formalisation du raisonnement logique ni règle syntaxique de formation des phrases
- Apparente capacité de déduction logique ← capacité de **généralisation statistique** à partir d'**exemples**

**IA symbolique**



# IA symbolique

Une approche complètement différente...

1. On traduit notre problème dans un langage formel (*i.e* mathématique), typiquement la **logique** (propositionnelle, premier ordre...). En général, un problème =



# IA symbolique

Une approche complètement différente...

1. On traduit notre problème dans un langage formel (*i.e* mathématique), typiquement la **logique** (propositionnelle, premier ordre...). En général, un problème =
  - une base de connaissances (faits + données)



# IA symbolique

Une approche complètement différente...

1. On traduit notre problème dans un langage formel (*i.e* mathématique), typiquement la **logique** (propositionnelle, premier ordre...). En général, un problème =
  - une base de connaissances (faits + données)
  - une requête



# IA symbolique

Une approche complètement différente...

1. On traduit notre problème dans un langage formel (*i.e* mathématique), typiquement la **logique** (propositionnelle, premier ordre...). En général, un problème =
  - une base de connaissances (faits + données)
  - une requête
2. L'algorithme de raisonnement calcule de nouveaux faits à partir de la base de connaissances et de règles de déduction, et donne la réponse à la requête



# IA symbolique

Une approche complètement différente...

1. On traduit notre problème dans un langage formel (*i.e* mathématique), typiquement la **logique** (propositionnelle, premier ordre...). En général, un problème =
  - une base de connaissances (faits + données)
  - une requête
2. L'algorithme de raisonnement calcule de nouveaux faits à partir de la base de connaissances et de règles de déduction, et donne la réponse à la requête
3. On retraduit la réponse dans notre langue naturelle



# IA symbolique

Une approche complètement différente...

1. On traduit notre problème dans un langage formel (*i.e* mathématique), typiquement la **logique** (propositionnelle, premier ordre...). En général, un problème =
  - une base de connaissances (faits + données)
  - une requête
2. L'algorithme de raisonnement calcule de nouveaux faits à partir de la base de connaissances et de règles de déduction, et donne la réponse à la requête
3. On retraduit la réponse dans notre langue naturelle
  - Toutes les étapes sont vérifiables et explicables



## Exemple

- Les règles...

```
pyDatalog.create_terms('X, Y, Z, ancestor')  
  
parent(X, Y) <= mother(X, Y)  
parent(X, Y) <= father(X, Y)  
child(X, Y) <= parent(Y, X)  
ancestor(X, Y) <= parent(X, Y)  
ancestor(X, Y) <= parent(X, Z) & ancestor(Z, Y)
```



## Exemple

- Les règles...

```
pyDatalog.create_terms('X, Y, Z, ancestor')

parent(X, Y) <= mother(X, Y)
parent(X, Y) <= father(X, Y)
child(X, Y) <= parent(Y, X)
ancestor(X, Y) <= parent(X, Y)
ancestor(X, Y) <= parent(X, Z) & ancestor(Z, Y)
```

- Les faits...

```
+mother('Jon Snow', 'Lyanna Stark')
+father('Jon Snow', 'Eddard Stark')
+mother('Eddard Stark', 'Lycarra Stark')
+father('Eddard Stark', 'Rickard Stark')
+mother('Arya Starck', 'Catelyn Stark')
...
```



## Exemple

- La requête...

```
print(ancestor('Arya Starck', X))
```



## Exemple

- La requête...

```
print(ancestor('Arya Starck', X))
```

- La réponse...

```
X
-----
Minisa Whent (House Whent)
Hoster Tully (House Tully)
Arya Flint
Rodrik Stark (House Stark)
Catelyn Stark (House Tully)
Eddard Stark (House Stark)
Lyarra Stark
Rickard Stark (House Stark)
...
```



# IA statistique vs symbolique

- **IA statistique / numérique**  $\approx$  inférence de relations et de faits (cachés) derrière des données (numériques)
- **IA symbolique**  $\approx$  **formalisation** de la connaissance et utilisation de ce modèle pour inférer de nouveaux faits en utilisant un **mécanisme de preuve**
- Avantages de l'IA symbolique : **explicabilité et transparence**
  - toutes les étapes de raisonnement pour atteindre une conclusion s'appuient sur des représentations symboliques lisibles par les humains
- Inconvénients de l'IA symbolique : **robustesse et passage à l'échelle**
  - les connaissances et les règles doivent être codées à la main (mais on trouve de plus en plus de travaux sur l'apprentissage de règles à partir de données)
  - les algorithmes de raisonnement génériques ont une grande complexité computationnelle (dans le pire cas)



# IA statistique vs symbolique

- Les deux formes d'IA sont utiles car adaptées à la résolution de problèmes différents
- cf les deux formes d'intelligence (intelligence rapide, intuitive, inconsciente vs planification, raisonnement logique)
- l'avenir est probablement dans le couplage neuro-symbolique :
  - cf par exemple AlphaGo (2017) ou Nukkai pour le bridge

## Partie 5

---

### Les grands domaines de l'IA



# IA $\neq$ ML

Équation souvent lue (ou entendue) :

$$\text{IA} = \text{ML}$$

Et son équation jumelle plus restrictive :

$$\text{IA} = \text{DL}$$

En fait, le domaine de l'IA est beaucoup plus large que cela  
[Marquis et al., 2014]



**Marquis, P., Papini, O., and Prade, H. (2014).**

*Panorama de l'intelligence artificielle.*

Editions Cépaduès, Toulouse.



# 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. 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



# Raisonnement

- 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

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.



# Raisonnement

- 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

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.

**Applications** : planification logistique, problèmes de configuration, affectation (cf Parcoursup), recherche d'information (Web sémantique), systèmes de recommandation...



# Systemes multiagents

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

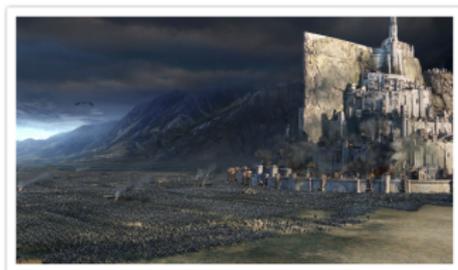


# Systèmes multiagents

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

Formalisation de l'intelligence collective et des comportements humains.

**Applications** : optimisation, génération de scènes réalistes, simulation de comportements de foules (risques naturels)...





# 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

Le domaine de l'apprentissage a déjà été largement évoqué dans cette présentation. Quelques méthodes d'apprentissage classiques :

- **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)

## Partie 6

---

### Conclusion



# 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



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

*Artificial intelligence: a modern approach.*

Pearson Education, 3rd edition edition.



# 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 2025 :

- 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 ?
  - Quel système politico-économique l'IA sert-elle ?



## Quelques pistes de réflexion

- Travaux de Kate Crawford (cf <https://katecrawford.net/>) + Anatomy of an AI System (<https://anatomyof.ai/>)
- Cathy O'Neil, *Weapons of Math Destruction* (2016)
- À propos de perroquets stochastiques [Bender et al., 2021]
- Sur ChatGPT dans le monde académique (<https://academia.hypotheses.org/58766>)
- 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\\_cY0Q0w](https://www.youtube.com/watch?v=oKcy_cY0Q0w))



**Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. (2021).**

On the dangers of stochastic parrots: Can language models be too big?

In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623.