

# L'intelligence artificielle

Cours

## Sommaire

### I Les principes de l'informatique

- A Le concept de machine universelle
- B Des données de natures diverses
- C Le code informatique

### II Les outils de l'intelligence artificielle

- A Les catégories d'apprentissage en intelligence artificielle
- B Les méthodes d'apprentissage machine
  1. La régression
  2. La classification
- C La méthode bayésienne
- D La qualité des données d'apprentissage

### III Les questions éthiques

#### RÉSUMÉ

Les principes de l'informatique qui ont permis de concevoir l'ordinateur d'aujourd'hui sont l'aboutissement d'un long processus. L'intelligence artificielle vise à reproduire le fonctionnement du cerveau humain grâce aux méthodes d'apprentissage machine. L'exploitation illimitée de données et la prise de décision par une machine posent des questions éthiques.

## I Les principes de l'informatique

Les principes de l'informatique ont évolué du concept de machine universelle développé par Turing vers des systèmes complexes. Ils peuvent traiter des données de nature diverse à l'aide d'un code informatique.

### A Le concept de machine universelle

Turing développe le premier un système de machine universelle qui permet d'effectuer quelques tâches simples.

DÉFINITION

## Machine de Turing

**La machine de Turing** est la machine la plus élémentaire possible destinée à mettre en œuvre des mécanismes de calcul comme le font notamment les ordinateurs. C'est une machine théorique qui peut effectuer des suites d'opérations très simples en se déplaçant le long d'un ruban.

Turing est considéré comme un des fondateurs de l'informatique. Il a notamment contribué à la victoire des Alliés durant la Seconde Guerre mondiale en déchiffrant le code Enigma utilisé par les Allemands pour chiffrer leurs communications sensibles.



## **B** Des données de natures diverses

Les données du monde réel sont de natures diverses : images, textes, vidéos, suites de nombres, etc. Elles sont traitées différemment selon leur type.

### DÉFINITION

#### Variable

**Une variable** est une information contenue dans une « boîte » que l'algorithme repère par son nom.

Il existe plusieurs types de variables :

<b>string</b>	Hello	Chaînes de caractères	Texte : la « boîte » contient un caractère ou une chaîne de caractères.
<b>int</b>	42	Entiers relatifs	Numérique : la « boîte » contient un nombre.
<b>bool</b>	True False	Résultats d'une comparaison < =, >=, ==, ...	Booléen : la « boîte » contient la valeur « vrai » ou la valeur « faux ».

En informatique, la variable est la brique de base que l'on peut utiliser. Elle ne peut contenir qu'un seul type de donnée selon l'utilisation que l'on souhaite en faire.

Les logiciels permettent de traiter des données différentes.

#### EXEMPLE

Gimp ou Photoshop servent à la retouche d'image.

## C Le code informatique

Pour communiquer avec un ordinateur, on écrit un code informatique ou un algorithme qui permet d'exprimer une suite d'instructions.

#### DÉFINITION

### Algorithme

Un **algorithme** est une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre une classe de problèmes.



#### REMARQUE

Un algorithme peut être écrit en langage naturel ou traduit, dans un langage de programmation, sous la forme d'un programme exécutable par un ordinateur ou une calculatrice. La division euclidienne est un exemple d'algorithme.

L'affichage d'une instruction à l'écran n'est pas un algorithme, mais une portion de code informatique.

#### PROPRIÉTÉ

Un algorithme se découpe en trois parties :

- les informations nécessaires à l'utilisation de l'algorithme ;
- la suite d'instructions à effectuer ;
- la réponse attendue à la fin.

## II Les outils de l'intelligence artificielle

À partir de grandes quantités de données issues du monde réel, les outils de l'intelligence artificielle permettent de faire des prédictions ou de prendre des décisions. Il existe deux grandes méthodes d'apprentissage : la régression et la classification. La fiabilité des prédictions est conditionnée par la qualité des données d'apprentissage.

## A Les catégories d'apprentissage en intelligence artificielle

L'intelligence artificielle se divise en différentes catégories selon le type de tâche à accomplir. Si l'on dispose de données étiquetées, on parle d'apprentissage supervisé, par opposition à l'apprentissage non supervisé lorsque les algorithmes détectent des régularités sans supervision humaine. Enfin, l'apprentissage par renforcement est particulièrement performant dans les jeux vidéo ou les jeux de réflexion.

### DÉFINITION

#### Intelligence artificielle (IA)

L'**intelligence artificielle (IA)** est un ensemble de techniques permettant à des machines d'accomplir des tâches et de résoudre des problèmes normalement réservés aux humains et à certains animaux.

L'apprentissage automatique est un sous-domaine de l'IA. On assimile parfois à tort l'intelligence artificielle à cette méthode, notamment avec des techniques qui permettent de traiter d'immenses volumes de données.

### DÉFINITION

#### Apprentissage automatique

L'**apprentissage automatique** (*machine learning*) permet d'entraîner la machine sur des données (phase d'apprentissage ou d'entraînement) pour définir son comportement ultérieur (phase d'inférence ou de prédiction).

#### EXEMPLE

Pour identifier les rubriques des articles d'actualité, on peut utiliser des techniques d'apprentissage automatique.

### DÉFINITION

#### Apprentissage supervisé

L'**apprentissage supervisé** adapte les paramètres du modèle grâce à des données historiques. Cet algorithme permet ensuite de catégoriser ou de calculer une prédiction numérique sur des données inconnues.

#### EXEMPLE

Il peut s'agir de détecter un objet ou un animal dans une image.

### DÉFINITION

#### Apprentissage non supervisé

L'**apprentissage non supervisé** repère des régularités, des proximités, des corrélations pour construire lui-même son algorithme de classification. Ces méthodes demandent une plus faible intervention humaine.

#### EXEMPLE

On peut regrouper des photos d'espèces de fleurs similaires sans intervention humaine grâce à des techniques d'apprentissage non supervisé.

#### DÉFINITION

### Apprentissage par renforcement

L'**apprentissage par renforcement** consiste à induire le comportement d'un « agent » évoluant dans un « environnement » et qui apprend grâce à un système de « récompenses ».

#### EXEMPLE

L'apprentissage par renforcement est le principe que l'on trouve derrière la startup DeepMind. Elle est réputée pour avoir développé AlphaGo, la première machine à avoir triomphé du champion du monde au jeu de go.

## B Les méthodes d'apprentissage machine

Les méthodes d'apprentissage machine consistent à entraîner les programmes informatiques sur de grandes quantités de données. Elles utilisent les outils statistiques pour repérer des tendances ou faire des prédictions.

#### DÉFINITION

### Phase d'apprentissage

La **phase d'apprentissage** permet de déterminer les paramètres du modèle qui seront utilisés durant la phase de prédiction. Lors de cette phase, l'algorithme essaie d'avoir une prédiction la plus proche possible de la réalité.

La phase d'apprentissage permet d'ajuster les paramètres d'un modèle. Avant d'utiliser ces prédictions, il faut avoir une idée de l'erreur commise. Cette erreur s'estime lors de la phase de test.

#### DÉFINITION

### Phase de test

La **phase de test** permet de calculer une « erreur de prédiction » avant d'appliquer l'algorithme sur de nouvelles données. Elle permet de savoir si les résultats de l'algorithme sont cohérents avant de passer à la phase d'inférence.

Lorsque l'erreur est suffisamment faible, les prédictions peuvent être calculées pour ensuite prendre des décisions.

#### DÉFINITION

### Phase d'inférence

La **phase d'inférence** ou phase de prédiction est l'étape qui suit la phase d'entraînement. Le modèle est utilisé pour prédire un résultat sur de nouvelles données qui n'ont pas été utilisées lors de la phase d'entraînement.

#### EXEMPLE

On souhaite détecter les photos sur lesquelles apparaît un chat. Lors de la phase d'entraînement, on fournit à la machine un grand nombre de photos en précisant si un chat apparaît ou non. La phase d'apprentissage permet à l'algorithme de construire une fonction de classification qui sera utilisée lors de la phase d'inférence.



## 1. La régression

La régression est une branche de l'intelligence artificielle qui permet d'effectuer des prédictions numériques sur des individus inconnus.

### DÉFINITION

#### Nuage de points

Un **nuage de points** est l'ensemble des points qui peuvent être fournis lors de la phase d'apprentissage à un algorithme d'IA. Des méthodes permettent d'ajuster ce nuage par une courbe (droite, parabole, exponentielle).

Les paramètres de cette courbe sont déterminés à l'aide des données : c'est la phase d'entraînement. La connaissance de cette courbe permet alors de prédire des valeurs pour des points de mesure absents : c'est la phase d'inférence.

Un nuage de points dans un plan en deux dimensions peut être proche d'une droite. Dans ce cas, on peut utiliser une régression linéaire.

### DÉFINITION

#### Régression linéaire

Une **régression linéaire** est un ajustement sur un nuage de points pour lequel le résultat est une droite. Il s'agit de la droite qui passe au plus près des vraies valeurs mesurées, dans un sens mathématique bien précis, celui des moindres carrés.



#### REMARQUE

La plupart des algorithmes d'apprentissage consistent à déterminer les valeurs de paramètres d'un modèle à l'aide de méthodes d'optimisation. Pour la régression linéaire, on calcule les paramètres  $a$  et  $b$  d'une droite.

### EXEMPLE

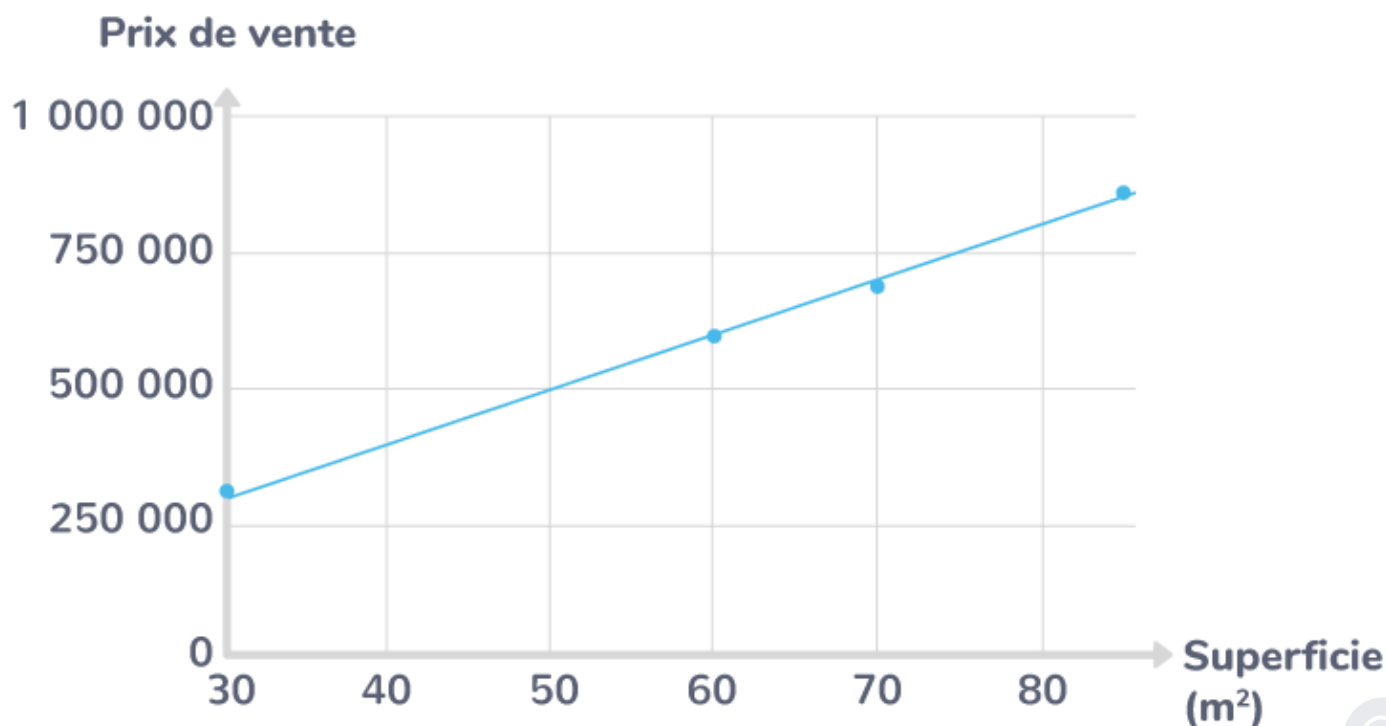
On peut utiliser des techniques d'intelligence artificielle pour calculer le prix de vente d'une maison en fonction de ses caractéristiques. Ici, on utilise une régression linéaire avec la superficie.

	Superficie (m <sup>2</sup> )	Prix de vente
Maison A	30	306 000
Maison B	60	595 000

Maison C	70	688 700
Maison D	85	862 100

On peut représenter graphiquement le prix de vente de ces maisons en fonction de leur superficie.

### Nuage de points et régression linéaire sur le prix de vente des maisons en fonction de la superficie



Un tableur donne l'équation de la droite  $y = 100\,004x + 176$ . Le nombre 100 004 est le coefficient directeur, 176 est l'ordonnée à l'origine.



REMARQUE

Il existe d'autres courbes d'ajustement (ajustement polynomial, exponentiel, logarithmique, etc.).

## 2. La classification

La classification est une branche de l'intelligence artificielle qui permet de classer des individus dans des groupes.

### DÉFINITION

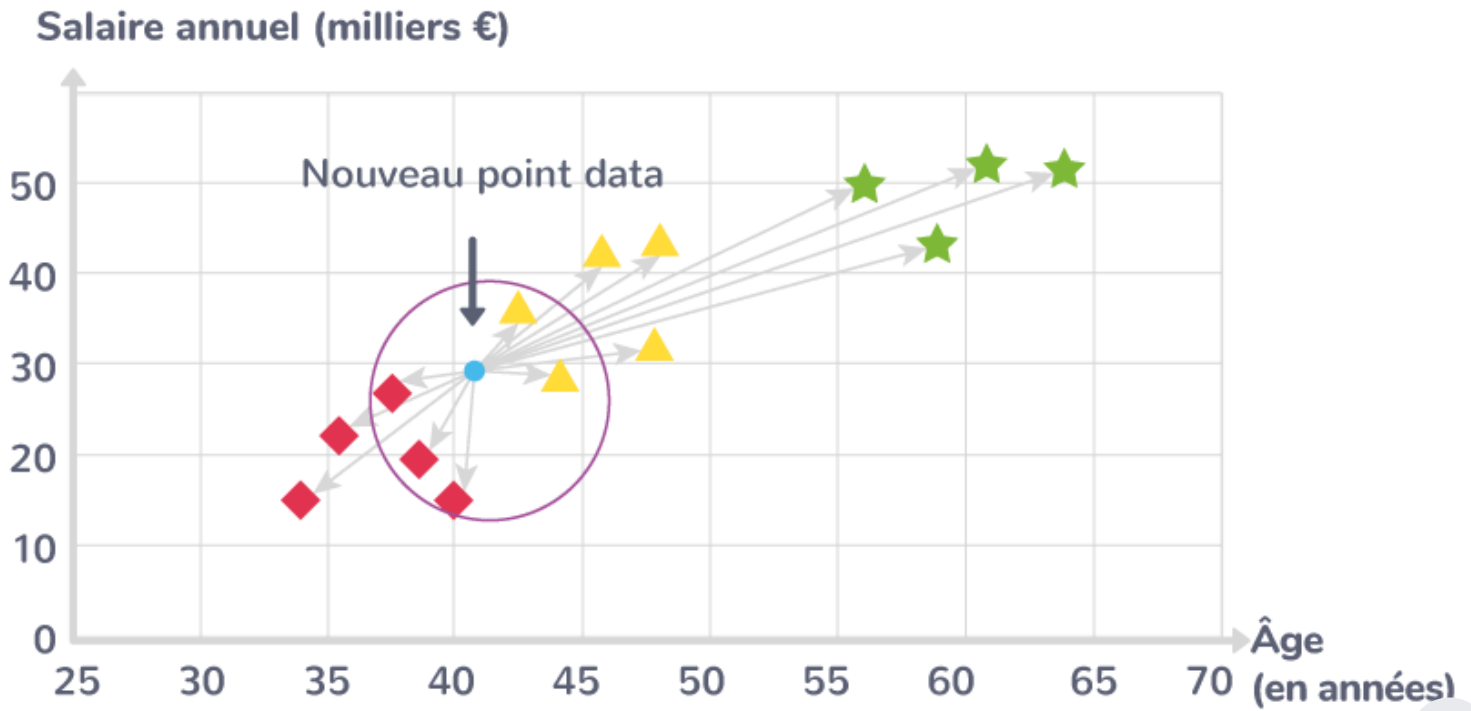
#### Classification

**La classification** est un processus pour prédire la classe à laquelle appartiennent des points. On appelle parfois ces classes la « cible » ou la « catégorie » des points.

#### EXEMPLE

Sur le graphique suivant, on cherche à classer le salaire d'une personne en fonction de son âge en trois catégories : moyen (rouge), haut (jaune), élevé (vert). Pour attribuer une nouvelle catégorie au point bleu, on peut utiliser l'algorithme des plus proches voisins pour lui assigner la catégorie de ses voisins les plus proches.

On décide de sélectionner ses cinq plus proches voisins. Parmi eux, trois sont rouges et deux sont jaunes. La catégorie du nouveau point sera donc « rouge », car ses voisins rouges sont les plus nombreux.



## C La méthode bayésienne

**La méthode bayésienne est une méthode de calcul de probabilités de causes à partir des probabilités de leurs effets. Elle permet de détecter des anomalies.**

La formule de Bayes permet de calculer la probabilité a posteriori d'une cause (par exemple, être malade) à partir des probabilités de l'effet observé (les chances de test positif lorsqu'on est malade et lorsqu'on ne l'est pas), en prenant en compte la probabilité a priori de la cause. Elle permet de réviser le pourcentage de chances a priori d'être malade en fonction des observations.

#### FORMULE

#### Formule de Bayes

Étant donné deux événements  $A$  et  $B$ , le théorème de Bayes permet de déterminer la probabilité de  $A$  sachant  $B$ , si l'on connaît les probabilités :

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}$$

#### EXEMPLE



Le principe du filtrage des spams repose sur le fait que les mots du dictionnaire n'apparaissent pas avec la même fréquence dans un e-mail légitime ou dans un spam. L'algorithme doit apprendre ces fréquences lors de la phase d'apprentissage, à partir de données étiquetées par des humains.

Chaque message reçu est catégorisé en « spam » ou « non spam », et l'algorithme ajuste les probabilités de rencontrer ce mot dans un spam ou dans un courrier légitime.

On note  $P_S(M)$  la probabilité que le mot  $M$  apparaisse dans un spam et  $P_{\bar{S}}(M)$  la probabilité que le mot  $M$  apparaisse dans un e-mail légitime. On estime ces probabilités sur les données d'entraînement, ainsi que la probabilité  $P(S)$  qu'un message soit un spam.

La formule de Bayes nous permet de calculer  $P_M(S)$ , c'est-à-dire la probabilité qu'un message  $S$  soit un spam s'il contient le mot  $M$ .

$$P(S|M) = \frac{P(M \cap S)}{P(M)} = \frac{P(M|S) \times P(S)}{P_S(M) \times P(S) + P_{\bar{S}}(M) \times (1 - P(S))}$$

On se fixe un seuil, et si cette probabilité est supérieure au seuil, alors le message est catégorisé comme un spam.

## D La qualité des données d'apprentissage

La qualité des données d'apprentissage est essentielle pour avoir un modèle d'intelligence artificielle sans biais. Elle repose sur la représentativité des données utilisées pour l'entraînement.

Les données sont la matière première d'un algorithme d'intelligence artificielle. Pour s'assurer qu'elles n'auront pas des conséquences indésirables sur les prédictions, il faut vérifier les critères suivants :

- la qualité des données : elles doivent être exactes. Il faut anticiper les tentatives de malveillance et mettre en place des systèmes de contrôle ;
- la quantité de données : les algorithmes d'apprentissage ont besoin d'avoir un échantillon représentatif de la population sur laquelle ils effectuent ensuite leurs prédictions ;
- la représentativité des données : elles doivent être les plus exhaustives possibles.

### EXEMPLE

Amazon utilise depuis 2014 une intelligence artificielle pour classer des CV et améliorer le processus de recrutement. Toutefois, il a été remarqué que l'IA discriminait systématiquement les profils féminins pour des postes techniques, habituellement réservés à des hommes. En réalité, l'IA d'Amazon ne faisait que reproduire le biais sexiste d'embauche de ses ingénieurs.

## III Les questions éthiques

La substitution du cerveau humain par l'intelligence artificielle pose des questions éthiques. Les algorithmes d'apprentissage machine reproduisent les biais du monde réel tels que les discriminations. L'exploitation des données personnelles doit faire l'objet d'un contrôle attentif.

L'intelligence artificielle se retrouve dans de nombreux domaines de la vie quotidienne ou professionnelle. Elle remplace les caissiers des supermarchés par des capteurs automatiques ou les ressources humaines en triant automatiquement les CV des candidats. Un des enjeux est la déshumanisation de certains métiers.

Le respect de la quantité et de la représentativité des données est par ailleurs essentiel pour une bonne utilisation de l'IA dans la prise de décision. Si les prédictions sont biaisées et si elles alimentent des décisions réelles, elles peuvent amplifier le biais au lieu de les résoudre. On parle alors de boucle de rétroaction.

#### DÉFINITION

### Rétroaction

**Une rétroaction** désigne l'action en retour d'un système à la modification d'un paramètre. Si la réponse du système amplifie le phénomène, on parle alors de rétroaction positive. Si elle l'atténue, on parle à l'inverse de rétroaction négative.

#### DÉFINITION

### Boucle de rétroaction

**Une boucle de rétroaction** est un cycle où l'action d'un système se reproduit perpétuellement. Une boucle de rétroaction positive amplifie la perturbation, une boucle de rétroaction est négative si elle tend à atténuer la perturbation.

#### EXEMPLE

PredPol, le modèle de police prédictive de Los Angeles, indique les lieux dangereux où des crimes sont susceptibles de survenir pour y déployer des forces policières. Les contrôles sont plus réguliers dans ces zones, souvent des quartiers pauvres où vivent majoritairement des personnes de couleur. Ce sont donc des personnes de couleur qui sont le plus ciblées par des contrôles : la base de données d'entraînement de Los Angeles est donc biaisée. On voit ici comment un modèle d'intelligence artificielle est modifié par un comportement discriminant fondé sur la couleur de peau. En utilisant une base de données biaisée, comme c'est le cas pour la police de Los Angeles, le prédictif se transforme en boucle de rétroaction, discriminatoire dans ce cas, qui s'auto-alimente. L'algorithme prédit plus de crimes dans les quartiers qui sont déjà les plus contrôlés, alors plus de crimes sont détectés donc plus de contrôles ont lieu, et la boucle est sans fin.