



## R2.13 : Intelligence Artificielle, Apprentissage Supervisé (partie 1)

**Nicolas Mendez**

IUT de Béziers



# Organisation du module

- 21h TD : N. Mendez
- 21h TP : M. De Castelbajac / A. Toffano
- Une évaluation prévue en fin de Ressource (coeff 4)
- Des QCM Moodle (coeff 1)
- Des CR de TP (coeff 1)

# Sommaire

- 1 Introduction et définition d'une IA
  - Contexte
  - Qu'est-ce qu'une IA ?
- 2 Historique (bref)
  - Dates clef
  - Évolutions techniques liées à l'IA
- 3 Vocabulaire de l'IA
  - Apprentissage automatique
  - Apprentissage supervisé
  - Apprentissage non supervisé
  - Apprentissage par renforcement
  - Réseau de neurones
  - Apprentissage profond
- 4 L'apprentissage supervisé
- 5 Mise en œuvre d'une IA
- 6 L'IA par l'exemple

# Introduction

## Pourquoi parle-t-on tant d'IA ?

- Recommandation de contenu existant (Deezer, Google, ...)
- Traitement de l'image et de la vidéo (Applis smartphones, ...)
- Génération de contenus « originaux » (Dall-E, ...)
- Automatisation d'un nombre croissant de processus

## Que peut-on en retenir ?

Il n'y a pas de définition officielle de ce qu'est une intelligence artificielle. L'héritage culturel de la science-fiction (comme pour la robotique) nous conduit à mal percevoir cette technologie.

Appliquée au contexte industriel, l'IA devra :

- Résoudre des problèmes posés
- Optimiser les solutions
- Anticiper les problèmes

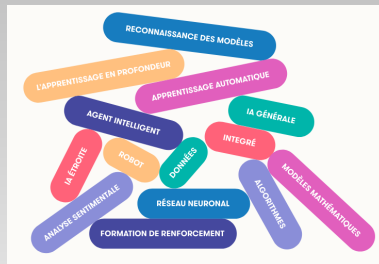
## Définition « acceptable » d'une IA

- L'intelligence artificielle (IA) est un domaine scientifique multidisciplinaire de l'informatique qui vise à créer des systèmes ou des machines capables d'imiter l'intelligence humaine.
- Elle englobe des concepts, des méthodes et des techniques pour permettre aux ordinateurs de résoudre des problèmes, d'apprendre, de raisonner, de percevoir, de planifier, de communiquer et d'interagir avec leur environnement de manière similaire à l'humain.
- Pour le Parlement européen, l'intelligence artificielle représente tout outil utilisé par une machine afin de « reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité »
- En pratique c'est un ensemble d'algorithmes et de techniques visant à imiter l'intelligence humaine.

## Rôles de l'IA

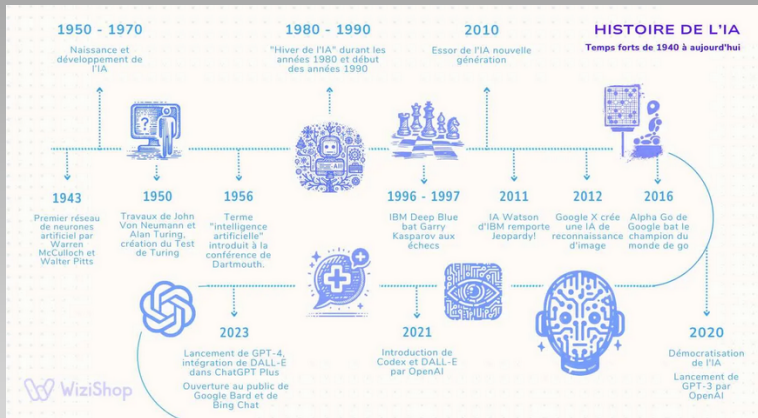
Pour mieux définir l'IA, il faudrait plutôt énumérer les caractéristiques propres à l'IA (autonomie, adaptabilité), ou ses typologies (générale, forte, étroite, super-intelligence).

- **Autonomie** : capacité d'exécuter des tâches dans des environnements complexes, sans guidage constant de la part d'un utilisateur.
- **Adaptabilité** : capacité d'améliorer les performances grâce à l'apprentissage par l'expérience.



# Historique de l'IA





## Technologies

- 1950 : réseaux de neurones artificiels
- 1960-70 : systèmes experts, programmes informatiques qui utilisent une base de connaissances pour résoudre des problèmes spécifiques
- 1970 : algorithmes génétiques, méthodes d'optimisation inspirés de la biologie de l'évolution
- 1980 : arbres de décision, modèles de classification basés sur des critères hiérarchiques.
- 1990 : traitement automatique du langage naturel (PLN)
- 2000 : deep learning, technique d'apprentissage automatique pour les réseaux de neurones artificiels profonds
- 2022 : l'IA générative arrive sur le marché et fait grand bruit. Le grand public s'empare de l'IA.

# Vocabulaire de l'IA

## Apprentissage automatique (ou apprentissage machine)

- **Définition :**

Processus par lequel un algorithme évalue et améliore ses performances sans l'intervention d'un programmeur, en répétant son exécution sur des jeux de données jusqu'à obtenir, de manière régulière, des résultats pertinents.

- Un algorithme d'apprentissage automatique comporte un modèle dont il modifie les paramètres, de valeur initiale en général aléatoire, en fonction du résultat constaté.
- fréquemment utilisé pour le traitement du langage naturel et la vision par ordinateur, ou pour effectuer des diagnostics et des prévisions.
- Équivalent étranger : machine learning (ML).



## Apprentissage supervisé

- **Définition :**

Apprentissage automatique dans lequel l'algorithme s'entraîne à une tâche déterminée en utilisant un jeu de données assorties chacune d'une annotation indiquant le résultat attendu.

- L'apprentissage supervisé recourt le plus souvent aux réseaux de neurones artificiels.
- fréquemment utilisé par exemple, pour la reconnaissance d'images et la traduction automatique.
- Équivalent étranger : supervised learning, supervised training.
- C'est le thème de travail central du semestre : détaillé dans une partie suivante



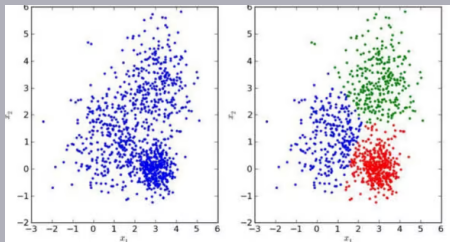
## Apprentissage non supervisé

- **Définition :**

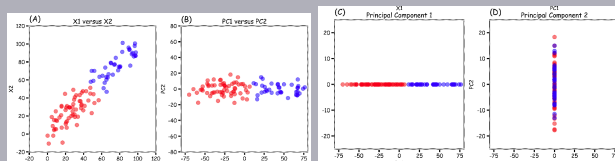
Apprentissage automatique dans lequel l'algorithme utilise un jeu de données brutes et obtient un résultat en se fondant sur la détection de similarités entre certaines de ces données.

- L'apprentissage non supervisé est utilisé, par exemple, pour l'identification de comportements et la recommandation d'achats.
- Équivalent étranger : data clustering, unsupervised learning, unsupervised training.

## Clustering



## Réduction de dimension





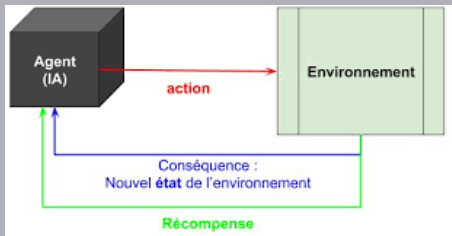
## Apprentissage par renforcement

- **Définition :**

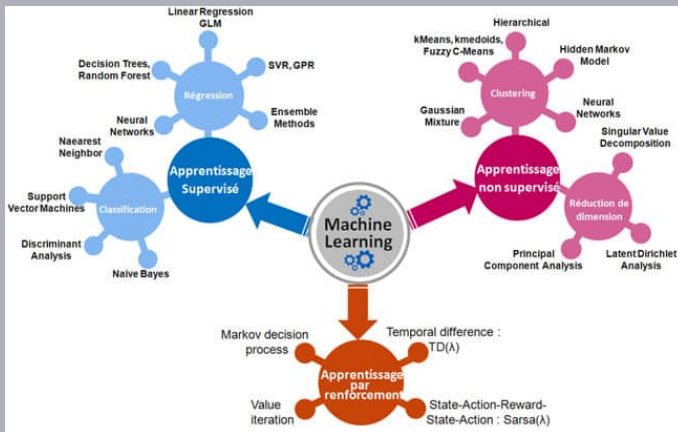
Apprentissage automatique dans lequel un programme extérieur évalue positivement ou négativement les résultats successifs de l'algorithme, l'accumulation des résultats permettant à l'algorithme d'améliorer ses performances jusqu'à ce qu'il atteigne un objectif préalablement fixé.

- L'apprentissage par renforcement est fréquemment utilisé dans la robotique.
- L'efficacité de l'apprentissage par renforcement a été attestée dans certains jeux stratégiques comme le jeu de go.
- Équivalent étranger : reinforcement learning.





## Schéma récapitulatif



## Tableau de Synthèse

	APPRENTISSAGE SUPERVISÉ	APPRENTISSAGE NON-SUPERVISÉ	APPRENTISSAGE PAR RENFORCEMENT
DÉFINITION	L'algorithme apprend à partir de données labellisées	L'algorithme est entraîné à partir de données non labellisées sans indications particulières	L'algorithme interagit avec son environnement en réalisant des actions et en apprenant de ses erreurs et succès
TYPE DE PROBLÈMES	Régression et classification	Association et Clustering	Basés sur un système de récompense
TYPE DE DONNÉES	Données labellisées	Données non labellisées	Pas de données fournies au préalable
APPROCHE	Étudie les relations sous-jacentes qui lient les données en entrée aux labels	Découvre les motifs communs au sein des données d'entrée	Apprend une stratégie de comportement en fonction d'expériences passées et des récompenses perçues

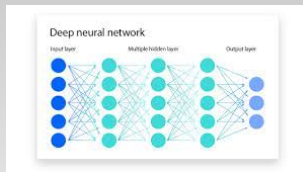
DataScientest • com

## Neurone artificiel / Réseau de neurones

- **Définition :**

Dispositif à plusieurs entrées et une sortie, qui simule certaines propriétés du neurone biologique.

- La valeur de sortie du neurone artificiel est une fonction non linéaire, généralement à seuil, d'une combinaison de valeurs d'entrée dont les paramètres sont ajustables.
- Un ensemble de neurones artificiels interconnectés qui constitue une architecture de calcul est appelé un **réseau de neurones**
- Équivalent étranger : artificial neuron, artificial neurone, formal neuron, formal neurone, artificial neural network, neural network.

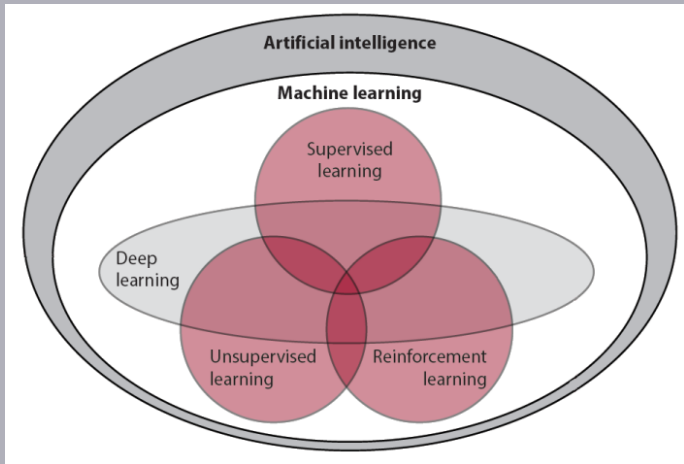


## Apprentissage profond

- **Définition :**

Apprentissage automatique qui utilise un réseau de neurones artificiels composé d'un grand nombre de couches dont chacune correspond à un niveau croissant de complexité dans le traitement et l'interprétation des données.

- L'apprentissage profond est notamment utilisé dans la détection automatique d'objets au sein d'images et dans la traduction automatique.
- Équivalent étranger : deep learning, deep structured learning, hierarchical learning.



# L'apprentissage supervisé

## Principe

Si on reprend le TP1 de la R113 sur la reconnaissance d'un chien et d'un chat par une IA :

- La machine reçoit des données comportant des **caractéristiques** (couleur, forme des oreilles, position des yeux, etc...)
- Chaque image est **labellisée** « chien » ou « chat »
- L'algorithme supervisé va prédire la catégorie d'une nouvelle image en fonction de ses caractéristiques.

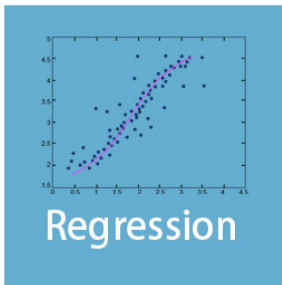




## Problématique de l'IA supervisée

A partir d'une entrée donnée et par un traitement algorithmique un système IA supervisé produit une sortie permettant de résoudre deux types de problèmes :

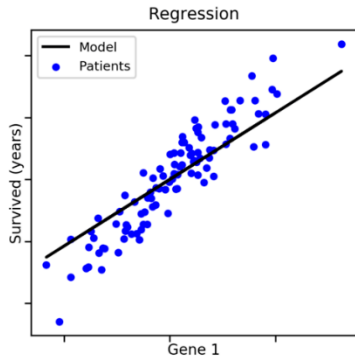
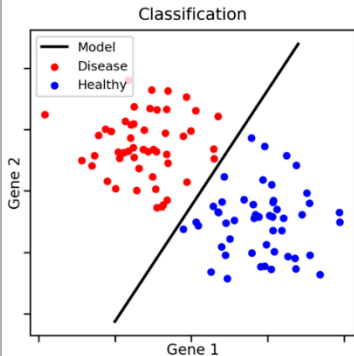
- La **classification** (l'exemple précédent)
- La **régression** (à venir)



VS



## Par exemple : en médecine



## Classification

- **objectif** : catégoriser les points de données dans des classes ou catégories.
- **variable de sortie** : catégorie, une étiquette ou une classe.
- **par exemple** : déterminer si un e-mail est du spam ou non, classer des images d'animaux en différentes espèces ou prédire si un objet est une clé USB ou un stylo.
- **sortie d'un algorithme de classification discrète (ou qualitative)** : Il attribue des points de données à l'une des classes en fonction des caractéristiques des données.
- **algorithmes couramment utilisés pour la classification** : les arbres de décision, les forêts aléatoires, les machines à vecteurs de support (SVM), les k-plus proches voisins (KNN) et les réseaux neuronaux.

## Régression

- **objectif** : prédire des valeurs numériques continues en fonction des caractéristiques d'entrée.
- **variable de sortie** : valeur réelle ou une variable continue.
- **par exemple** : prédire les prix des maisons en fonction de caractéristiques telles que la taille, l'emplacement et le nombre de chambres, prévoir les prix des actions ou estimer la température en fonction de données historiques.
- **sortie d'un algorithme de régression est continue (ou quantitative)** : Il prédit une quantité plutôt que d'attribuer des points de données à des catégories.
- **algorithmes couramment utilisés pour la régression** : la régression linéaire, la régression polynomiale, la régression par vecteurs de support (SVR), les arbres de décision, les forêts aléatoires et les réseaux neuronaux avec des fonctions d'activation linéaires ou sigmoïdes dans la couche de sortie.

## Avantages

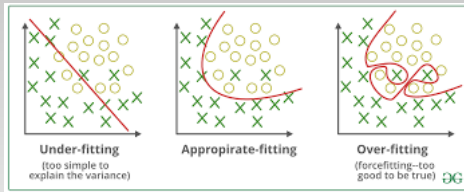
L'apprentissage supervisé est pertinent par rapport à d'autres types d'apprentissage automatique pour plusieurs raisons :

- **Clarté des objectifs** : modèle entraîné sur des données étiquetées où les entrées sont associées à des sorties connues et permet de mesurer directement la performance du modèle par rapport à ces étiquettes.
- **Performance élevée dans des tâches spécifiques** : Dans des cas où des données étiquetées sont disponibles.
- **Interprétabilité** : il est plus facile de comprendre comment les caractéristiques d'entrée contribuent aux prédictions du modèle.
- **Capacité à généraliser** : vise souvent à produire des modèles capables de généraliser à de nouvelles données.
- **Large gamme d'applications** : reconnaissance de formes, la vision par ordinateur, le traitement du langage naturel, la recommandation de produits, la détection de pannes, la prédiction de séries temporelles, et bien d'autres.

## ATTENTION TOUTEFOIS

Cependant, il est également important de reconnaître que l'apprentissage supervisé a ses limites.

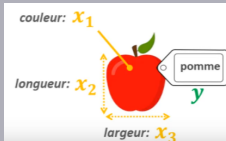
- **Taille training-set** : nécessite des données étiquetées pour l'entraînement, ce qui peut être coûteux et laborieux à obtenir dans certains domaines.
- **Sensible au surapprentissage** : lorsque le modèle s'adapte trop précisément aux données d'entraînement et ne généralise pas bien à de nouvelles données. Dans de telles situations, d'autres approches comme l'apprentissage non supervisé ou l'apprentissage par renforcement peuvent être plus appropriées.



# Mise en œuvre

## Les étapes incontournables

### 1 Préparation des données d'entraînement (data set → training set)



avec des données :

- caractérisées par des variables (features) notées  $x$
- annotées par une variable (label ou target) notée  $y$

- 2 Choix du modèle utilisé et des ses « hyperparamètres »
- 3 Entraînement du modèle
- 4 Évaluation du modèle
- 5 Utilisation du modèle pour prédire un résultat



## Et en Python ?

- L'utilisation de la bibliothèque `scikit-learn` va nous permettre de programmer facilement ces étapes
- NumPy permettra la mise en forme des données :
  - Une feature matrix  $X$  de dimension  $n$  samples par  $p$  features
  - Un target vector  $y$  de dimension  $n$  samples par 1
- La structure du script « type »

```
model = choix_modele(hyperparametres)
model.fit(X, y)
model.score(X, y)
model.predict(X)
```
- [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)

## Comment évaluer la pertinence d'un modèle ?

- Après l'entraînement des modèles, il est important de les évaluer pour mesurer leur performance et s'assurer qu'ils répondent aux critères de réussite définis au préalable
- Nécessité de créer des outils de contrôle : métrique, fonction de coût par exemple
- La notion d'optimisation est centrale ici.
- A chaque fois, l'algorithme doit pouvoir estimer sa capacité à prédire correctement le résultat (souvent exprimé dans scikit-learn par un nombre entre 0 et 1!).

## Les algorithmes que nous allons aborder

Je vous propose une organisation de type « cours inversé »  
Chaque groupe choisira un algorithme à présenter le plus clairement possible

- 1 KNN ( $k$  plus proches voisins)
- 2 Classifieur Bayésien Naïf
- 3 Algorithmes de régression linéaire, polynomiale, logistique
- 4 Les arbres de décisions (Random Forest par exemple)
- 5 L'Analyse de Composantes Principales
- 6 Le Perceptron et la méthode de descente du gradient
- 7 Machine à Supports de Vecteurs (SVM)

# l'IA par l'exemple

A vous de jouer !