



# Introduction Machine Learning

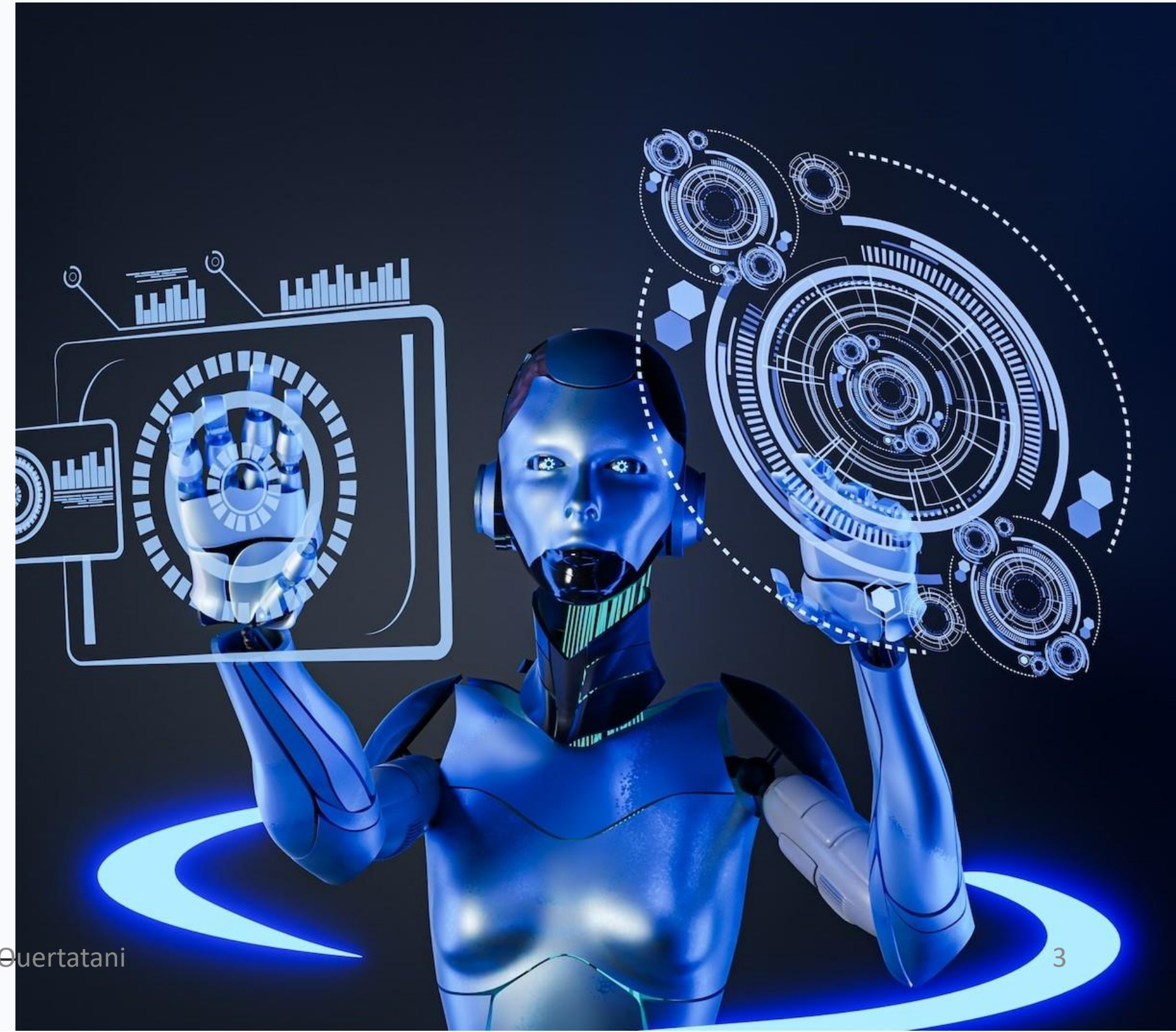
## Plan du chapitre I :

### I. Introduction

- ✓ Place du Machine Learning en IA
- ✓ Introduction au Machine Learning : Définition, ingrédients, avantages ...
- ✓ Apprentissage automatique : Applications
- ✓ Machine Learning : Problématique (Formalisation et pipeline)
- ✓ Types de Machine Learning

# INTRODUCTION

Le **Machine Learning** est une branche de l'**Intelligence Artificielle** qui permet aux systèmes de **s'améliorer** par l'expérience.

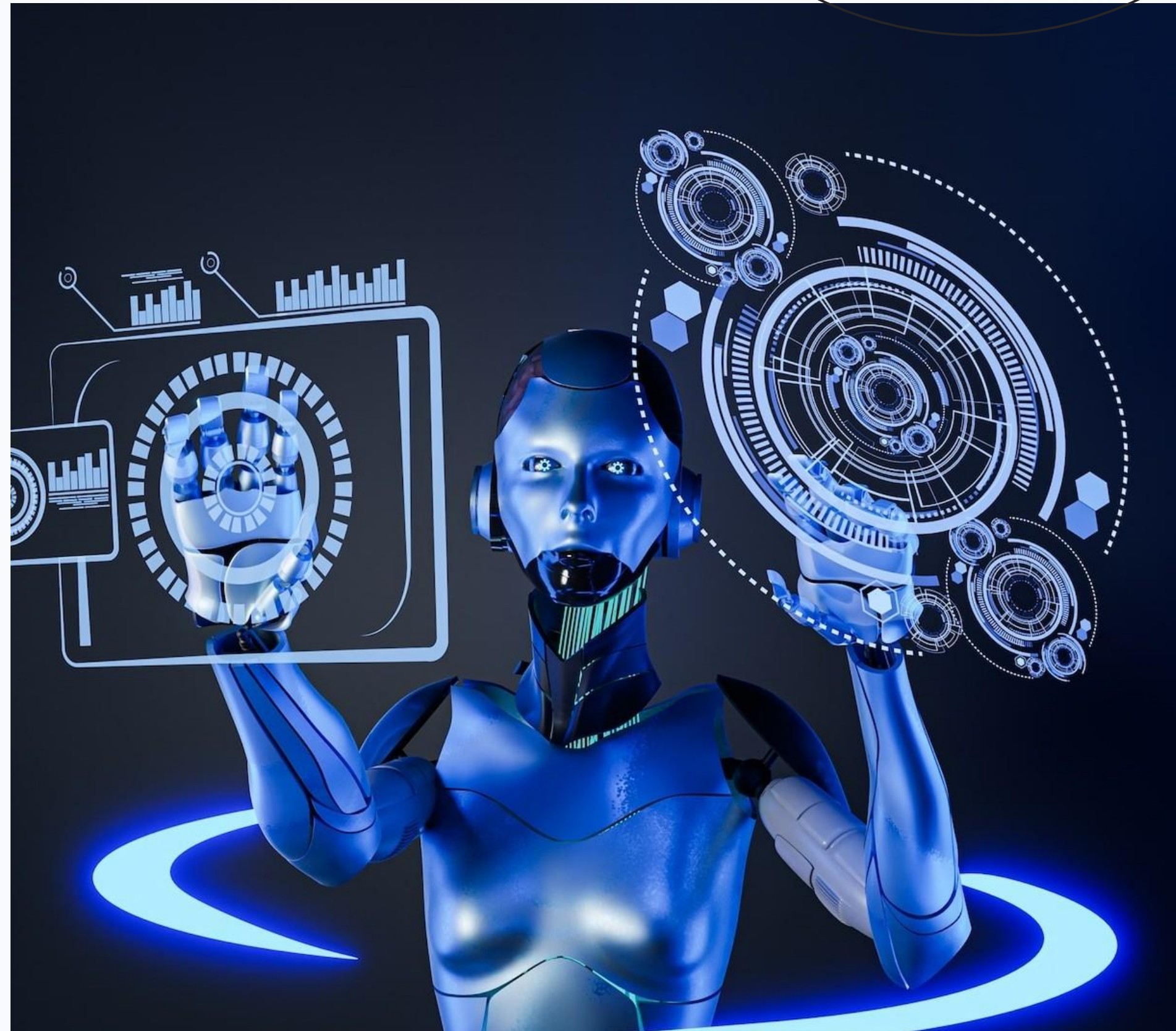




# INTRODUCTION

Le Machine Learning (ML) est une branche de **l'Intelligence Artificielle** qui permet aux systèmes de **s'améliorer** par l'expérience et qui se concentre sur :

- la création de **systèmes capables d'apprendre** et de **s'améliorer automatiquement** à partir de l'expérience sans être explicitement programmés.
- Il joue un rôle central dans l'IA en permettant aux machines de traiter des volumes de données importants et de prendre des décisions basées sur ces données.



# Place du Machine Learning en Intelligence Artificielle (IA)

## ➤ L'Intelligence Artificielle (IA) :

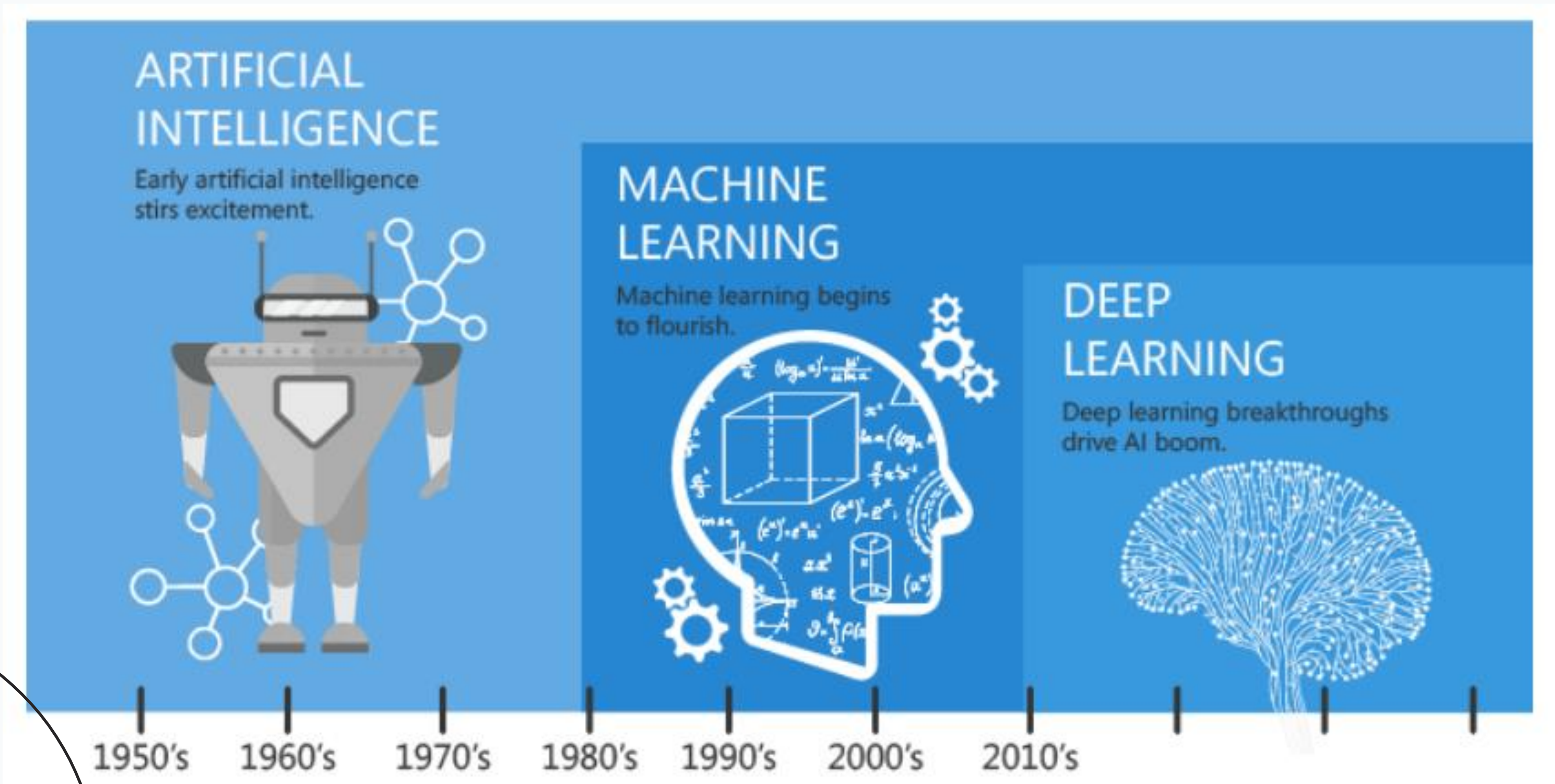
- Discipline informatique visant à créer des systèmes capables de simuler des comportements humains intelligents.
- Comprend plusieurs sous-domaines, tels que la vision par ordinateur, le traitement du langage naturel, la robotique, et **le Machine Learning**.

## ➤ Machine Learning (Apprentissage Automatique) :

- **Sous-domaine de l'IA** : Le Machine Learning permet aux systèmes de "**lire**" des données et **d'apprendre** à partir de celles-ci, sans être explicitement programmés pour une tâche spécifique.
- Le Machine Learning est essentiel pour le développement de l'IA moderne, car il permet aux machines de s'améliorer automatiquement grâce à l'expérience.



# Place du Machine Learning en Intelligence Artificielle (IA)



# Place du Machine Learning en Intelligence Artificielle (IA)



Contrairement aux systèmes traditionnels où les règles sont définies par un programmeur, le Machine Learning permet aux modèles de **découvrir des patterns** dans les données et de **prendre des décisions** basées sur ces patterns.

Le Machine Learning est une approche flexible et puissante de l'IA, utilisée dans des domaines variés comme :

- La **reconnaissance d'images**
- La **traduction automatique**
- L'**analyse prédictive**

Il joue un rôle clé dans le développement de l'**IA**, permettant aux machines d'accomplir des tâches spécifiques avec **précision et efficacité**.

# Introduction au Machine Learning

## Définition du Machine Learning

### Définition classique :

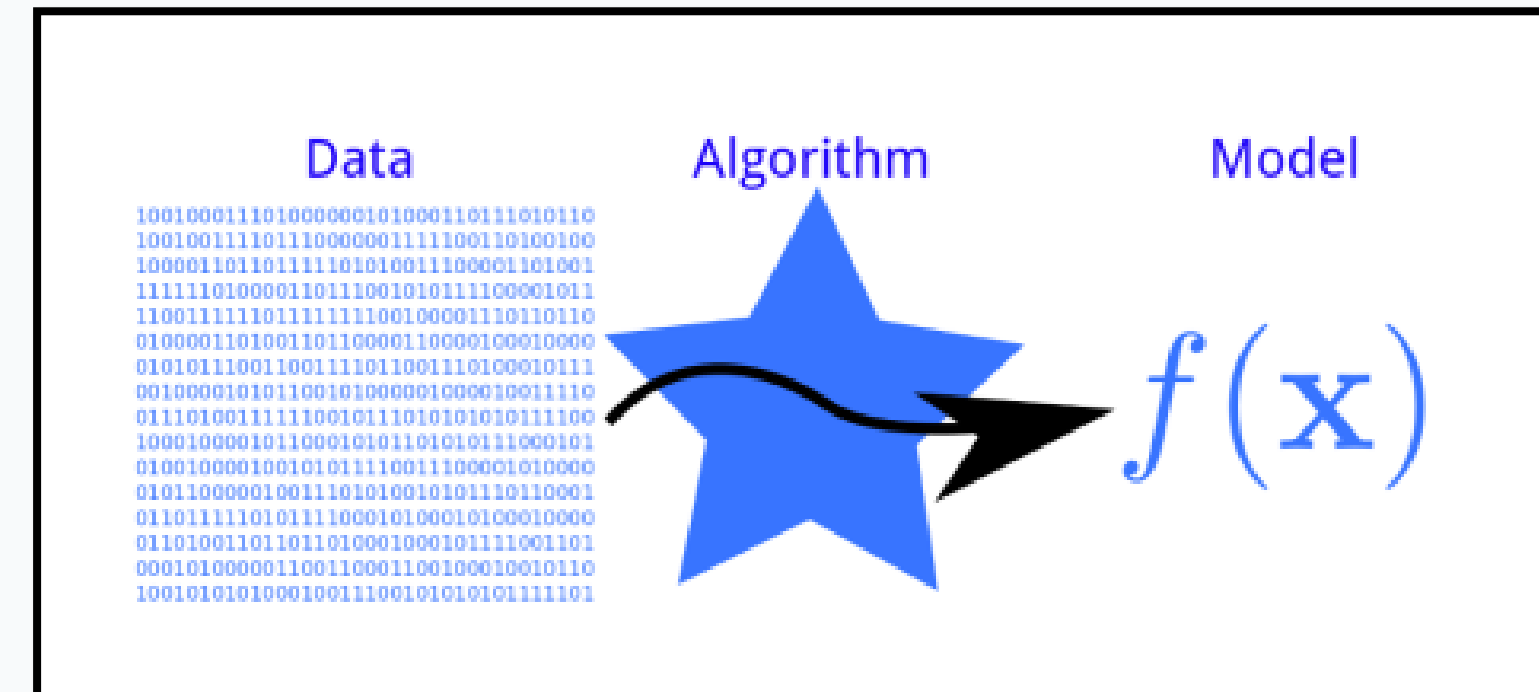
- *"Le Machine Learning est un domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés."* — Arthur Samuel (1959)

### •Principe clé :

- Le Machine Learning se base sur l'analyse des **données** pour identifier des **patterns** et construire des modèles prédictifs ou descriptifs.

### •Objectif :

- Construire des **modèles** capables de **généraliser** à partir d'exemples connus pour prendre des décisions ou faire des prédictions sur des données nouvelles et inconnues.

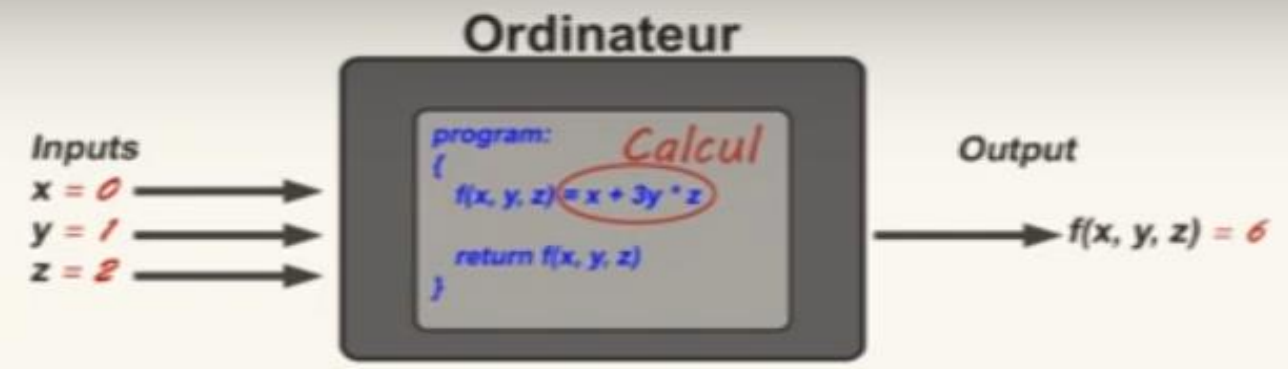




# Introduction au Machine Learning

## Comparaison avec un programme classique

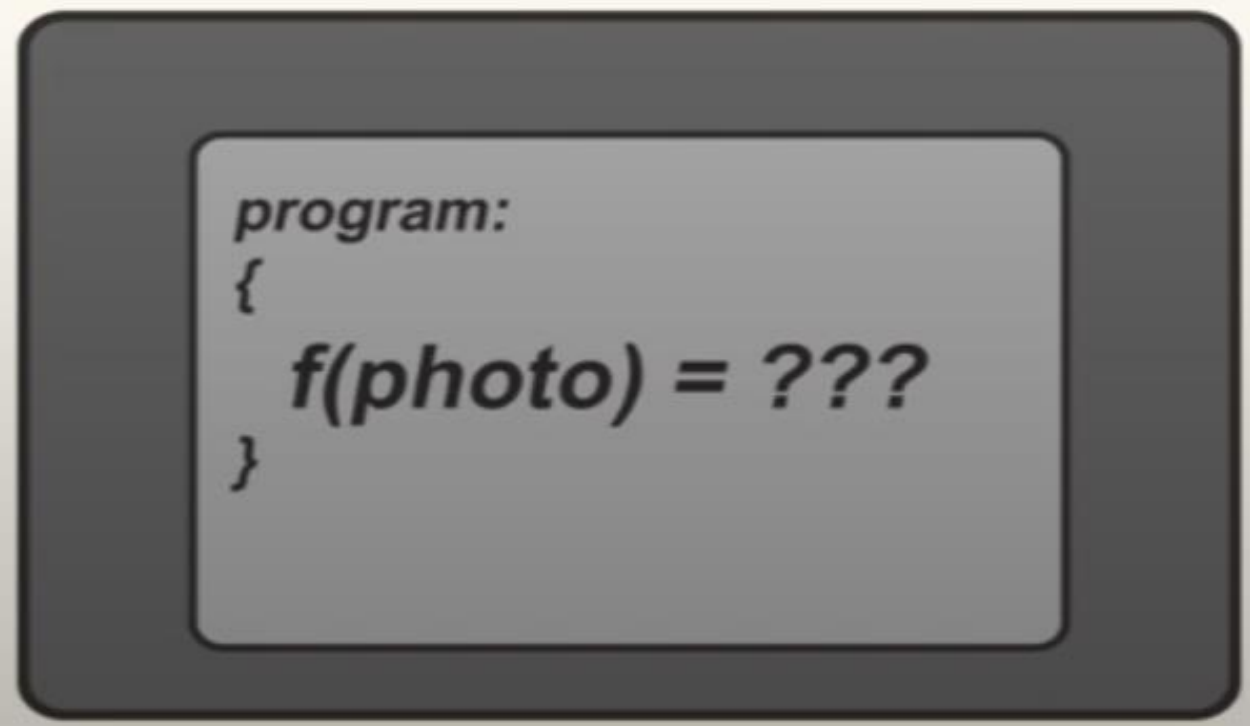
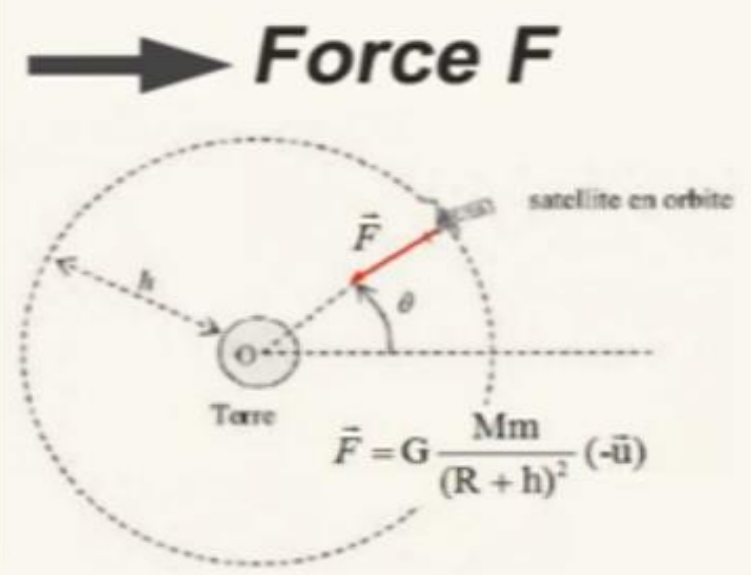
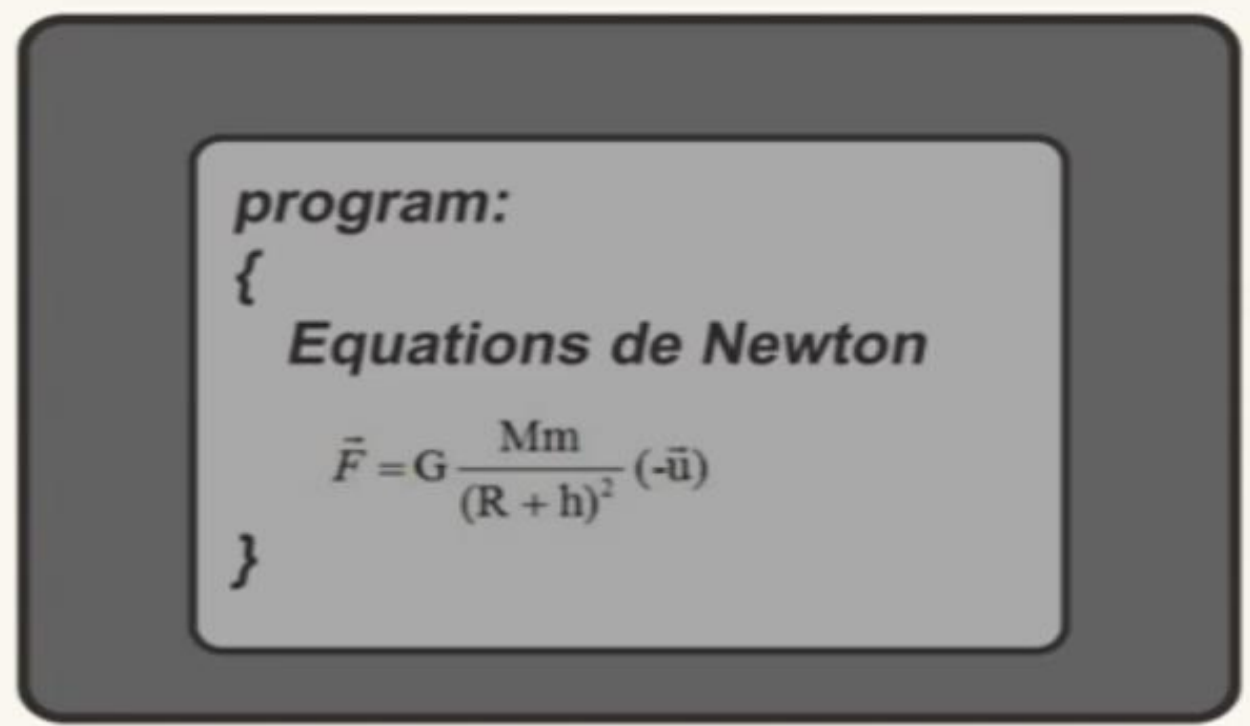
- **Programme classique** : Utilise une procédure et les données reçues en entrée pour produire des réponses en sortie.
  - **Programme d'apprentissage automatique** : Utilise les données et les réponses pour produire la procédure permettant d'obtenir les réponses à partir des données.
- **Machine Learning** : Une application pratique de l'intelligence artificielle permettant aux machines d'apprendre et de s'adapter à partir de l'expérience, sans intervention humaine directe.



Calculer la position du satellite dans l'espace :

m : La masse du Satellite  
M : La masse de la Terre

masse m  
masse M  
position h



“Chat”



# Introduction au Machine Learning

## Exemple d'application du machine learning

### •Exemple 1 : Calcul de dépenses

- Objectif : Calculer le montant total dépensé par un client à partir de ses factures.
- **Solution** : Algorithme classique, tel qu'une simple addition.
- **Conclusion** : Le Machine Learning n'est pas nécessaire ici.

### •Exemple 2 : Prédiction des achats futurs

- Objectif : Prédire les produits que le client est le plus susceptible d'acheter dans un mois.
- **Défi** : Les informations directement disponibles ne suffisent pas pour une prédiction précise.
- **Solution** : Utiliser l'historique d'achat d'un grand nombre de clients et un algorithme de Machine Learning pour créer un modèle prédictif.



# Introduction au Machine Learning

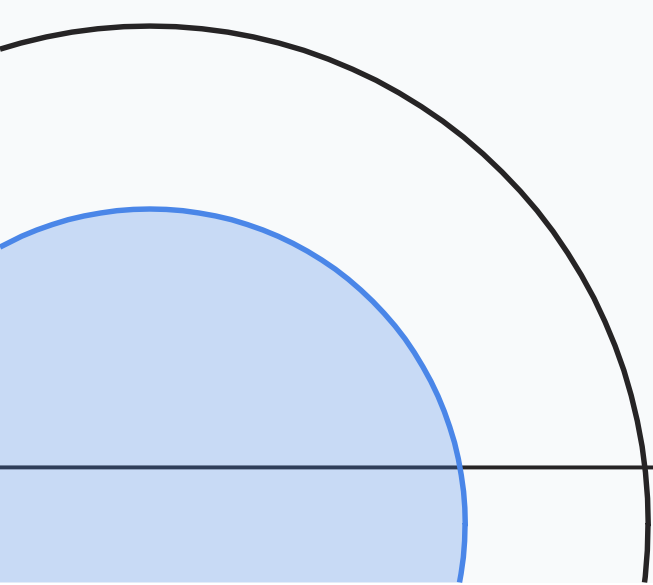
## Pourquoi utiliser le machine learning?

- **Le machine learning peut être utilisé pour résoudre trois types de problèmes :**
  - ✓ **Problèmes inconnus :**
    - Ceux que l'on ne sait pas résoudre (ex. : prédiction d'achats).
  - ✓ **Problèmes connus mais non formalisables :**
    - Ceux que l'on sait résoudre, mais pour lesquels on ne sait pas formaliser les étapes en termes algorithmiques (ex. : reconnaissance d'images, compréhension du langage naturel).
  - ✓ **Problèmes connus mais coûteux en ressources :**
    - Ceux que l'on sait résoudre, mais avec des procédures trop coûteuses en ressources informatiques (ex. : prédiction d'interactions entre molécules de grande taille pour lesquelles les simulations sont très lourdes).

# Introduction au Machine Learning

## Quand utiliser le machine learning ?

- Quand les données sont **abondantes**.
- Quand les connaissances sont **peu accessibles** ou **peu développées**.



# Introduction au Machine Learning

## Ingrédients du machine learning

Deux piliers fondamentaux du Machine Learning :

1. Les données : Ce sont les exemples à partir desquels l'algorithme apprend.
  - Les données doivent être pertinentes :
  - Le concept de "garbage in, garbage out" signifie qu'un algorithme ne peut produire que des prédictions de mauvaise qualité si les données fournies sont de mauvaise qualité.

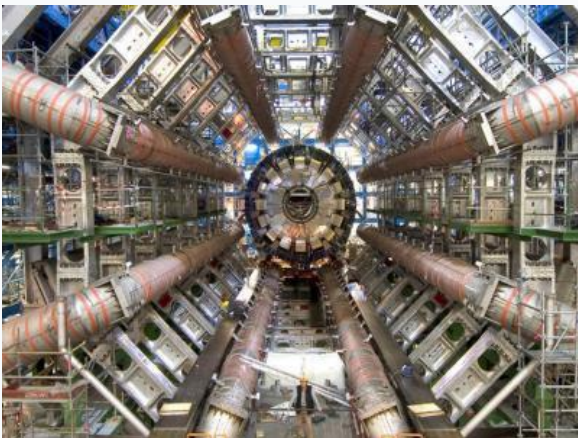


# Introduction au Machine Learning

Ingrédients du machine Learning :

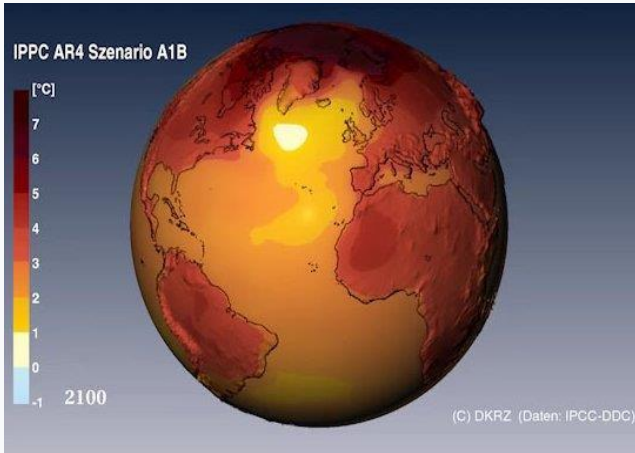
Données = carburant du ML

CERN /  
Large Hadron Collider  
~1 PetaOctets/jour



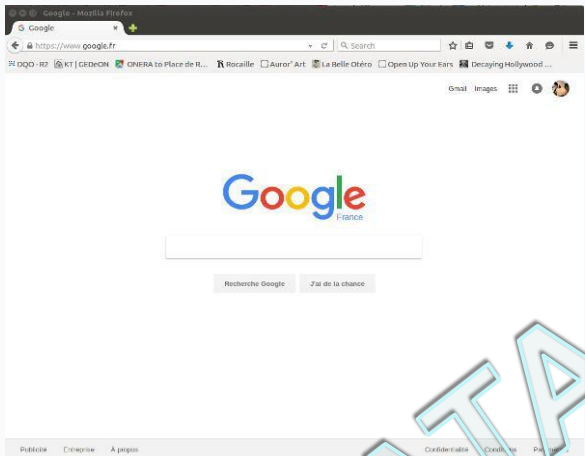
CERN: l'Organisation  
européenne pour la  
recherche nucléaire

DKRZ (Climat)  
500 Po



Le **Centre de calcul  
climatique allemand  
(DKRZ)** est une  
infrastructure de  
recherche dédiée à la  
simulation climatique.

Google :  
24 PetaOctets/jour



des données  
d'observation  
de la Terre



Square Kilometer Array  
1376 Po/an (en 2024)



Radiotélescope :  
Astronomie

BIG DATA

# Introduction au Machine Learning

## Ingrédients du machine learning

2. L'algorithme d'apprentissage : La procédure qui transforme les données en modèle.

- **L'algorithme doit être adapté** même avec des **données pertinentes**, un algorithme inadapté entraînera un modèle de **mauvaise qualité**.
- Entraînement : Le processus de faire tourner l'algorithme sur un jeu de données pour créer un modèle.

# Introduction au Machine Learning

## Exemple de machine learning

### Exemple 1 - Commerce Électronique

- Problème : Identifier des types représentatifs de clientèle pour un site marchand.
  - Données : Transactions passées des clients (achats, montants dépensés, comportements d'achat).
  - Objectif d'Optimisation : Maximiser la proximité entre les clients assignés à un même type (segmentation des clients).
- Finalité : Créer des segments de clientèle homogènes pour des campagnes marketing ciblées ou des recommandations personnalisées.



# Apprentissage automatique : applications

- Anti-Spam (*Classifieur Bayésien*)



- *Des compétitions contre des champions humains:*

1997 : **DeepBlue** (développé par IBM) bat un champion du monde d'échecs en activité, Garry Kasparov.

2017: **AlphaGo**, développé par DeepMind, a battu Ke Jie, alors numéro un mondial du jeu de Go.

2019: **AlphaStar**, également développé par DeepMind, a atteint un niveau professionnel dans le jeu de stratégie en temps réel StarCraft II, en battant des joueurs humains de haut niveau.

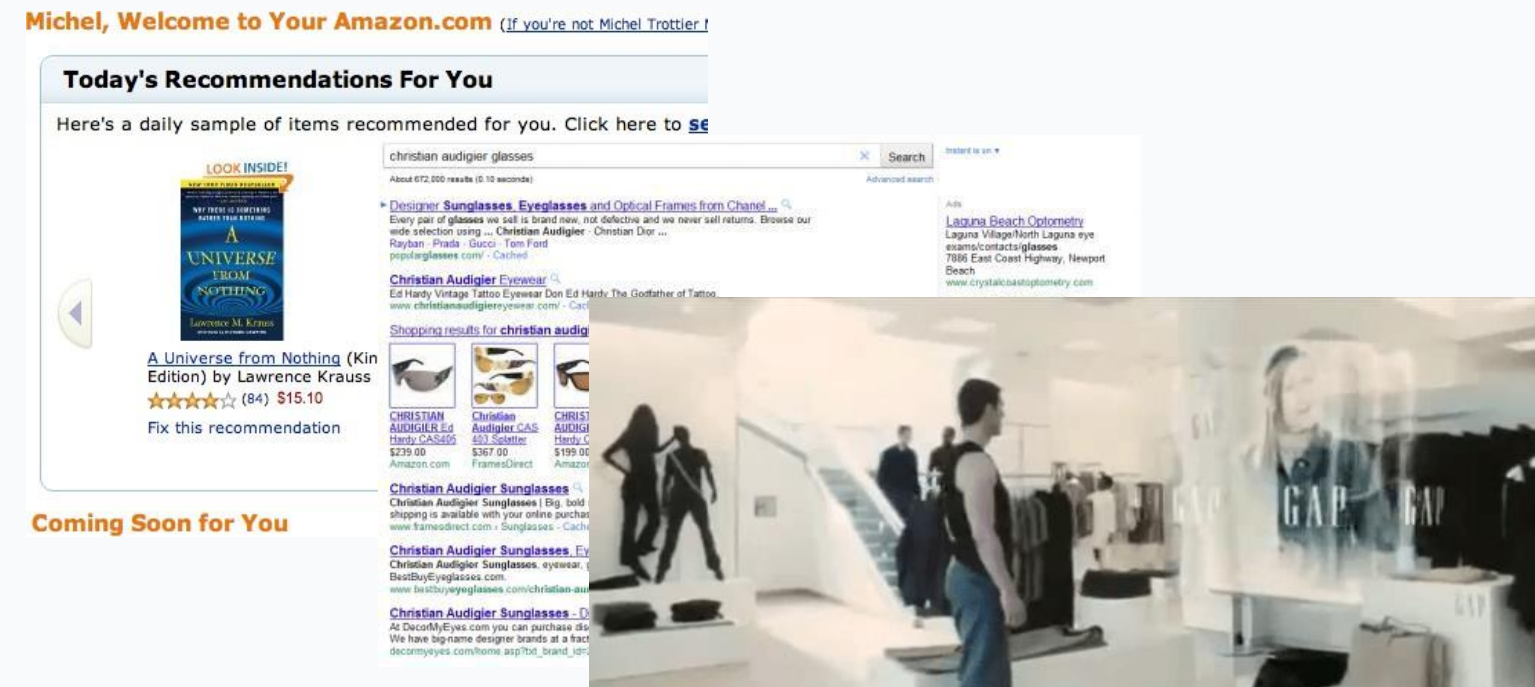


- Tri postal automatique (*détection de chiffres manuscrits par réseaux de neurones*)



# Apprentissage automatique : applications

- Recommandation ciblée (*régression logistique*)



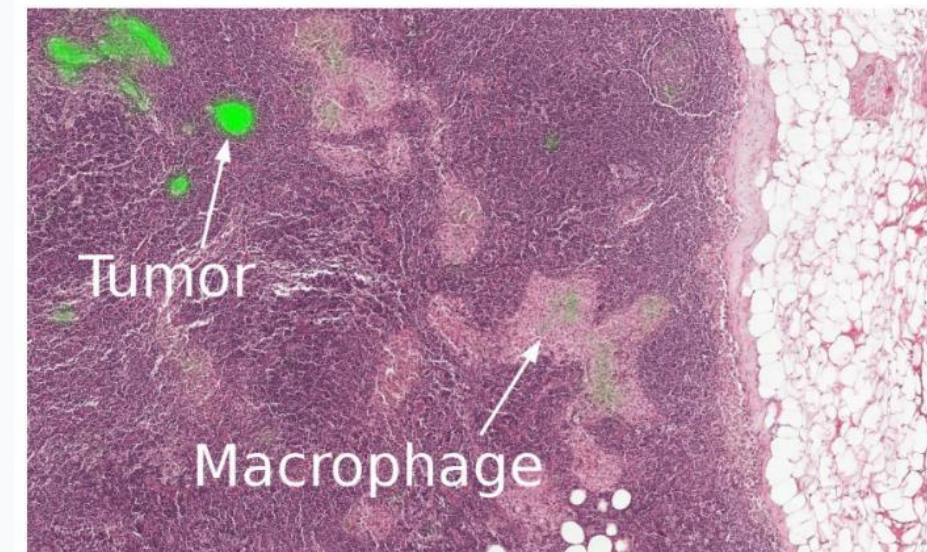
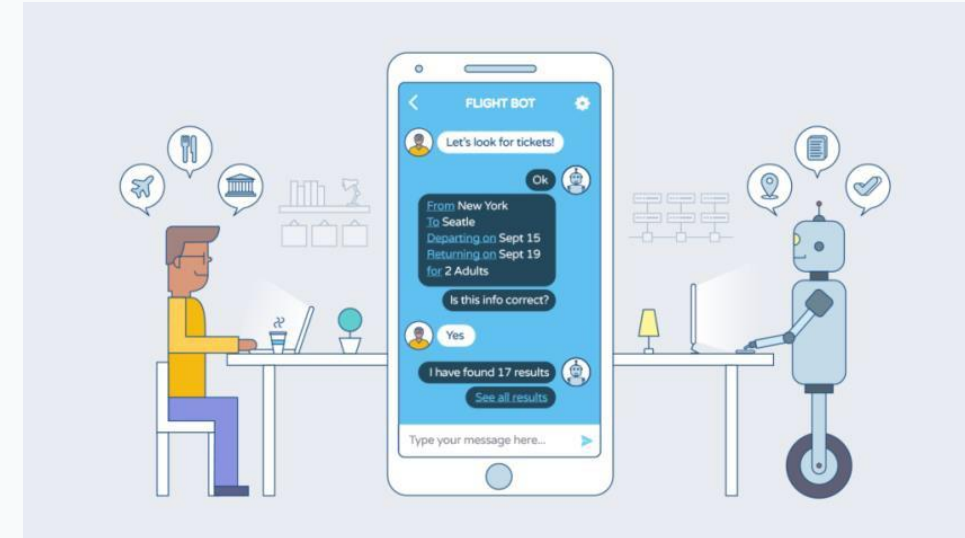
- Appareil photo avec détection de visages (*boosting*)





# Apprentissage automatique : applications

- Chat Bots  
(*Réseaux de neurones*)
- Diagnostic médical  
(*Réseaux de neurones*)
- Traduction multi-lingue  
(*Réseaux de neurones*)





# Facteurs d'émergence du Machine Learning

## 1. Augmentation des données (Big Data)

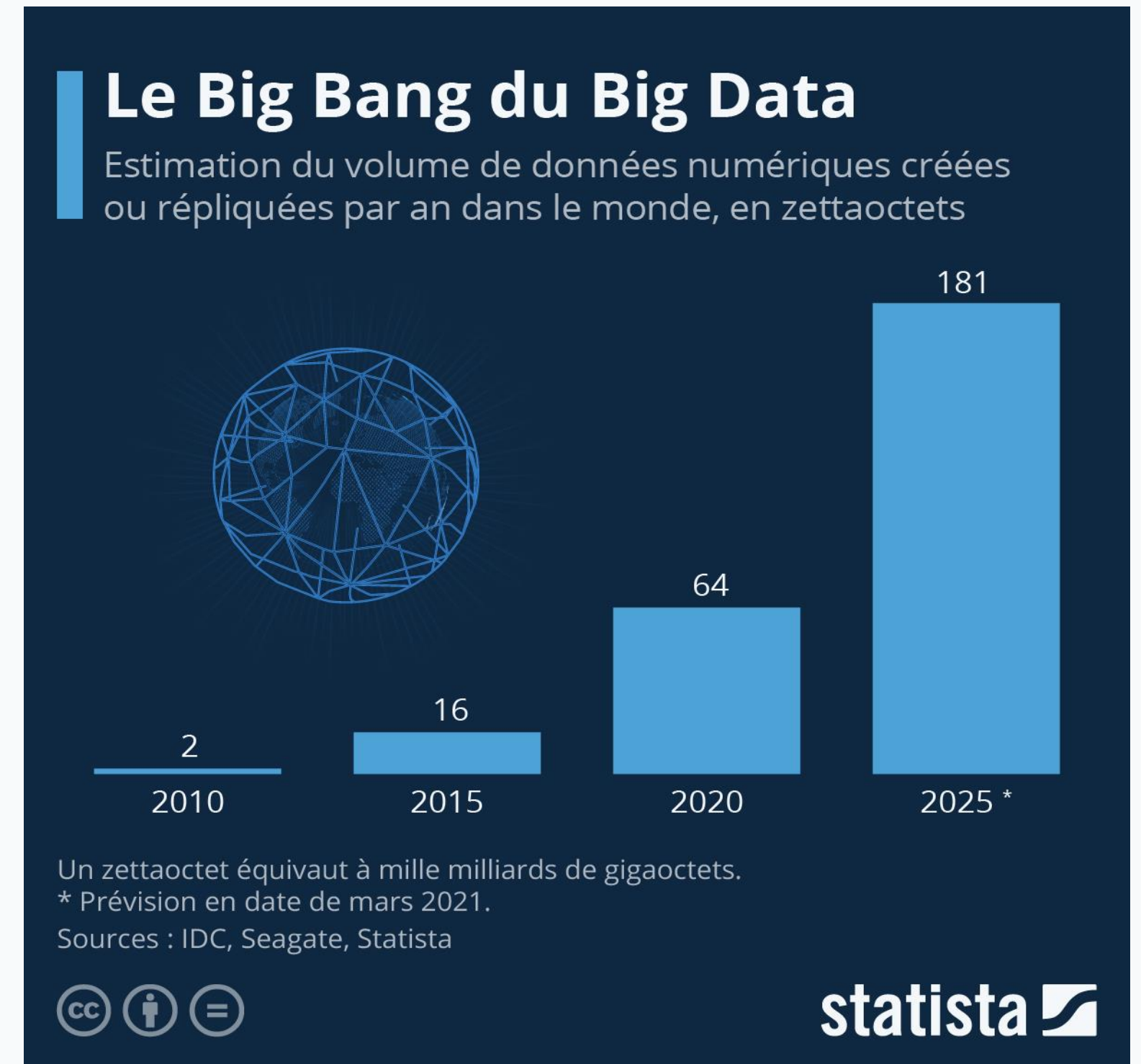
- Explosion des données numériques (utilisateurs, capteurs, réseaux sociaux).
- Gestion des données massives et complexes.

## 2. Progrès des infrastructures de calcul

- Puissance accrue des processeurs (GPU, CPU) et cloud computing.
- Plateformes de calcul distribué (Hadoop, Spark).

## 3. Avancées algorithmiques

- Nouveaux modèles (réseaux neuronaux, forêts aléatoires).
- Optimisation des méthodes d'apprentissage.



# Facteurs d'émergence du Machine Learning

## 4. Accès à des données labellisées

- Augmentation des bases de données annotées.
- Crowdsourcing pour l'étiquetage (ex. : Mechanical Turk).

## 5. Open source et collaboration

- Bibliothèques open-source (TensorFlow, PyTorch).
- Développement collaboratif par les communautés.

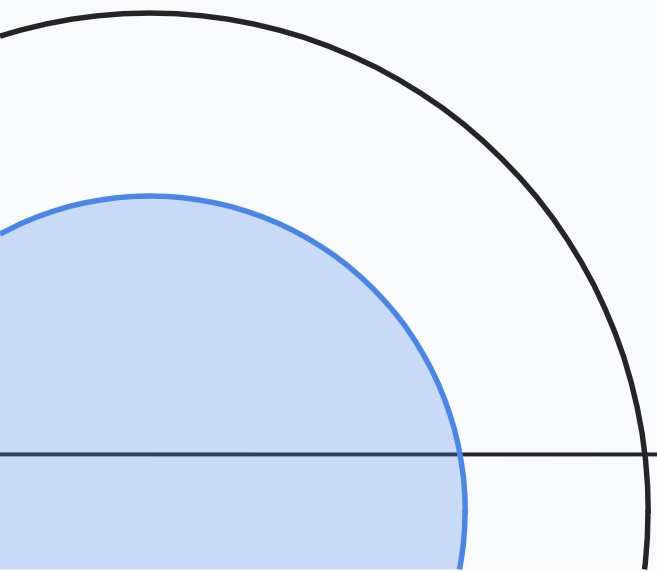
## 6. Investissements industriels

- Automatisation des processus décisionnels.
- Croissance des startups et investissements en IA.



MACHINE LEARNING

Problématique générale



# ML: Problématique générale

L'apprentissage automatique est:

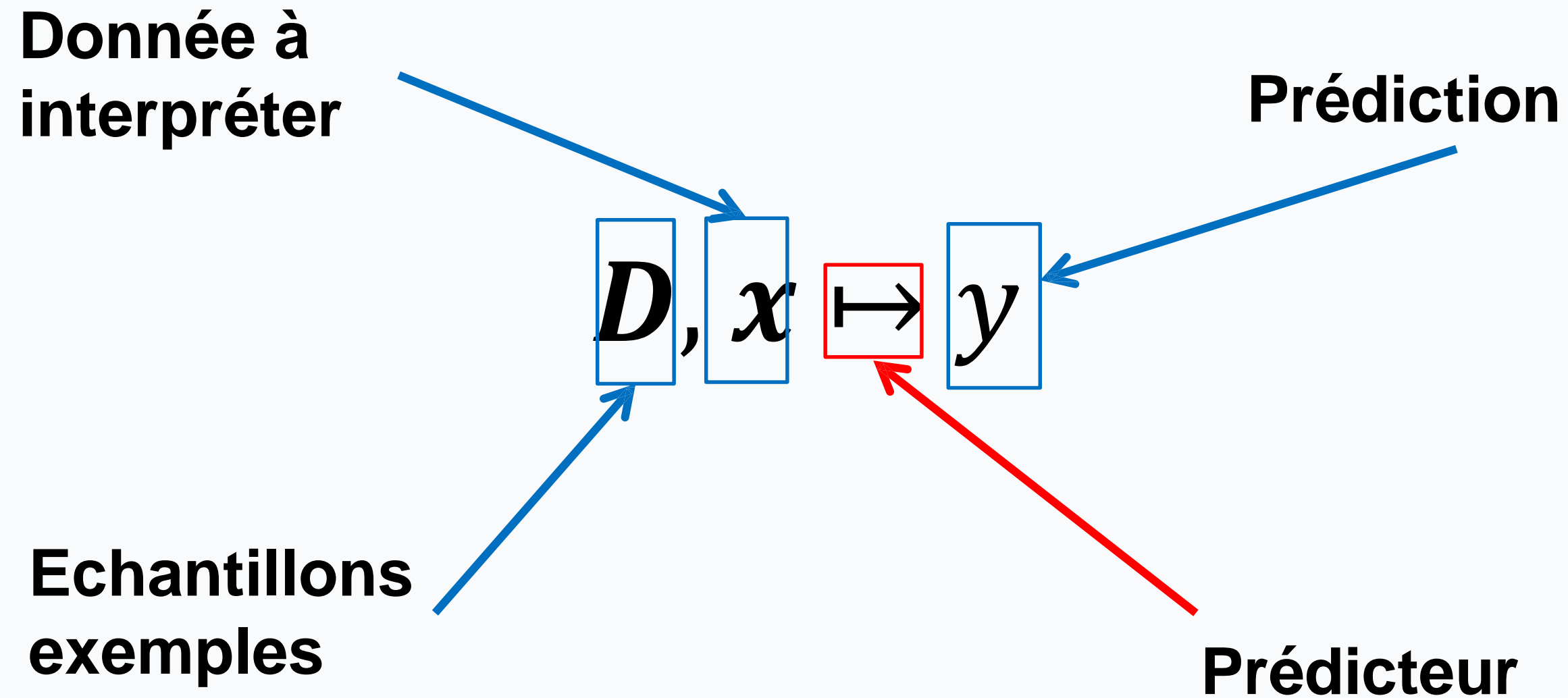
- Une démarche de **conception** d'une **fonction de prédiction**
- Une modélisation ou programmation **non explicite** à partir **d'exemples** (signaux, images, texte, mesures...)



# Formalisation

- Donnée à interpréter ( $\mathbf{x}$ )
  - Mesures, texte, image, enregistrement, vidéo ou caractéristiques extraites de ...
- Prédiction ( $y$ )
  - Décision, choix, action, réponse, préférence, groupe, commande, valeur...
- Echantillons ( $\mathbf{D} = \{(\mathbf{x}_i, y_i)\}$ )
  - Exemples de données et de (bonnes) prédictions
  - « Base d'apprentissage »:  $\mathbf{D}$

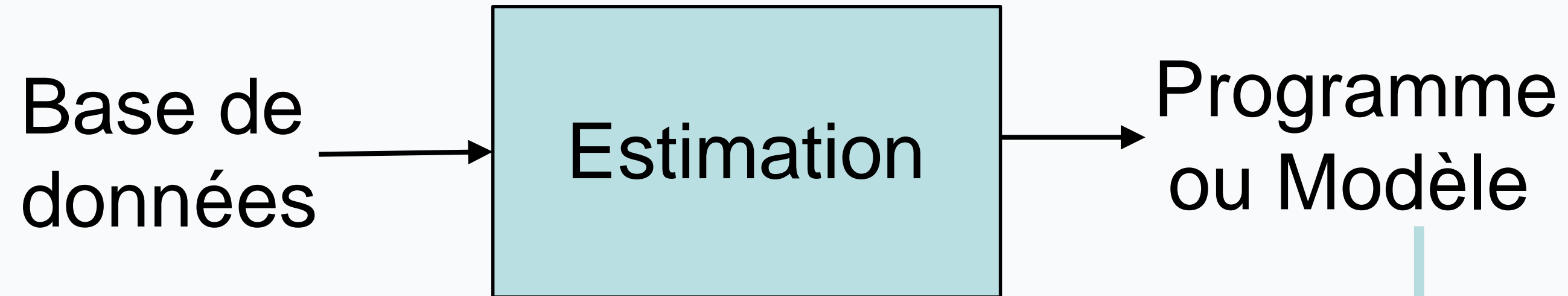
# Formalisation



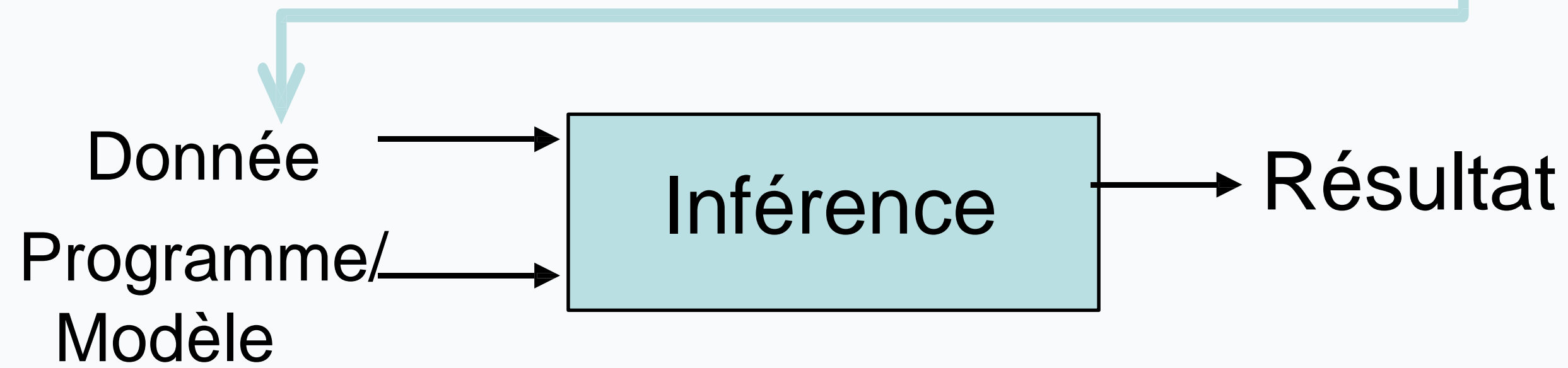
- Hypothèse forte: les échantillons contiennent toute l'information exploitable et utile
- Prédicteur = « interpolateur » à partir des données  $D$

# Deux phases

## Apprentissage



## Prédiction



# Formalisation

Donnée à interpréter

Prédiction

$$W, x \mapsto y$$

Deux phases :

Apprentissage

$$D \mapsto W$$

Échantillons  
exemples

Caractéristiques du  
prédicteur: Paramètres,  
Poids, Prototypes...



# Formalisation

## 1. Jeu de données

- **L'objectif** est de construire un modèle qui **apprend** à partir d'un ensemble de données d'entraînement
- $D = \{(\mathbf{x}_i, y_i)\}$  où :
- $\mathbf{x}_i$  : les **entrées** (features),
- $y_i$  : les **sorties** ().
- Le modèle cherche à prédire  $y$  (labels ou cibles) à partir de nouvelles entrées  $\mathbf{x}$ .

# Formalisation

## 2. Fonction paramétrique de prédiction :

- Le modèle est défini par une fonction paramétrique  $F$ :

$$y = F(\mathbf{x}; \mathbf{W})$$

- $\mathbf{x}$  : entrée (vecteur de caractéristiques),
- $\mathbf{W}$ : ensemble des paramètres du modèle,
- $F$  : fonction de prédiction (par exemple : régression linéaire, réseau de neurones).

# Formalisation

## 3. Apprentissage du modèle :

- L'apprentissage consiste à trouver les paramètres  $W$  qui **minimisent une fonction de perte**  $L$ , définie pour mesurer l'erreur de prédiction sur les données d'entraînement.

*(Apprentissage = trouver le  $W$  qui optimise un critère  $L$ )*

$$W = \arg \min_{W'} L(D, W')$$

- $L$  : fonction de perte (par exemple : l'erreur quadratique moyenne),
- $D$  : ensemble de données d'entraînement,
- $W'$  : paramètres candidats du modèle à optimiser.

# Formalisation

## Rappel : Fonction Arg Min

- La fonction **argmin**(ou opérateur argmin) est utilisée en mathématiques et en optimisation pour désigner l'argument qui minimise une fonction donnée.
- Plus précisément, lorsqu'on applique un **argmin** sur une fonction, cela signifie qu'on cherche la valeur des variables d'entrée qui rendent cette fonction aussi **petite** que possible.
- Soit une fonction  $f$ , l'argmin de cette fonction est défini comme:
- Cela signifie : trouver la valeur de  $x$  qui minimise  $f(x)$ .

$$\operatorname{argmin}_x f(x)$$



# Formalisation

## Rappel : Fonction Arg Min

### Exemple concret :

Prenons une fonction simple  $f(x) = (x - 2)^2$ .

- **Étape 1** : On cherche la valeur de  $x$  qui minimise  $f(x)$ .
- **Étape 2** : La fonction  $f(x)$  atteint son minimum quand  $x = 2$  parce que  $(2 - 2)^2 = 0$ , qui est la plus petite valeur possible pour cette fonction.
- **Résultat** : L'argmin de  $f(x)$  est donc  $x = 2$ , car c'est la valeur de  $x$  qui rend  $f(x)$  minimale.

$$\operatorname{argmin}_x (x - 2)^2 = 2$$

# Formalisation

Fonction paramétrique de prédiction

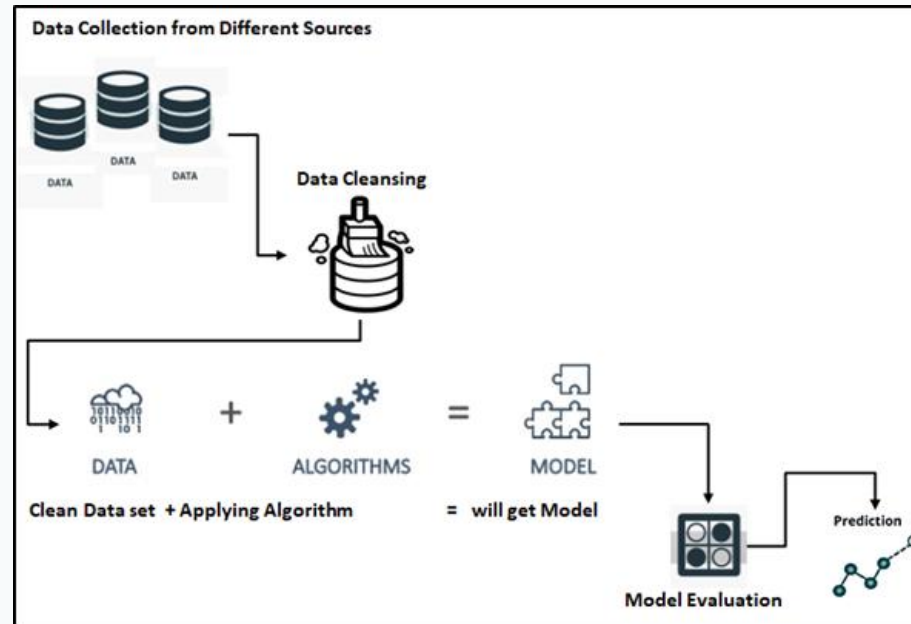
$$y = F(\boldsymbol{x}; \boldsymbol{W})$$

Apprentissage = trouver le  $\boldsymbol{W}$  qui optimise un critère  $L$

$$\boldsymbol{W} = \arg \min_{\boldsymbol{W}'} L(\boldsymbol{D}, \boldsymbol{W}')$$

A partir d'une base d'apprentissage

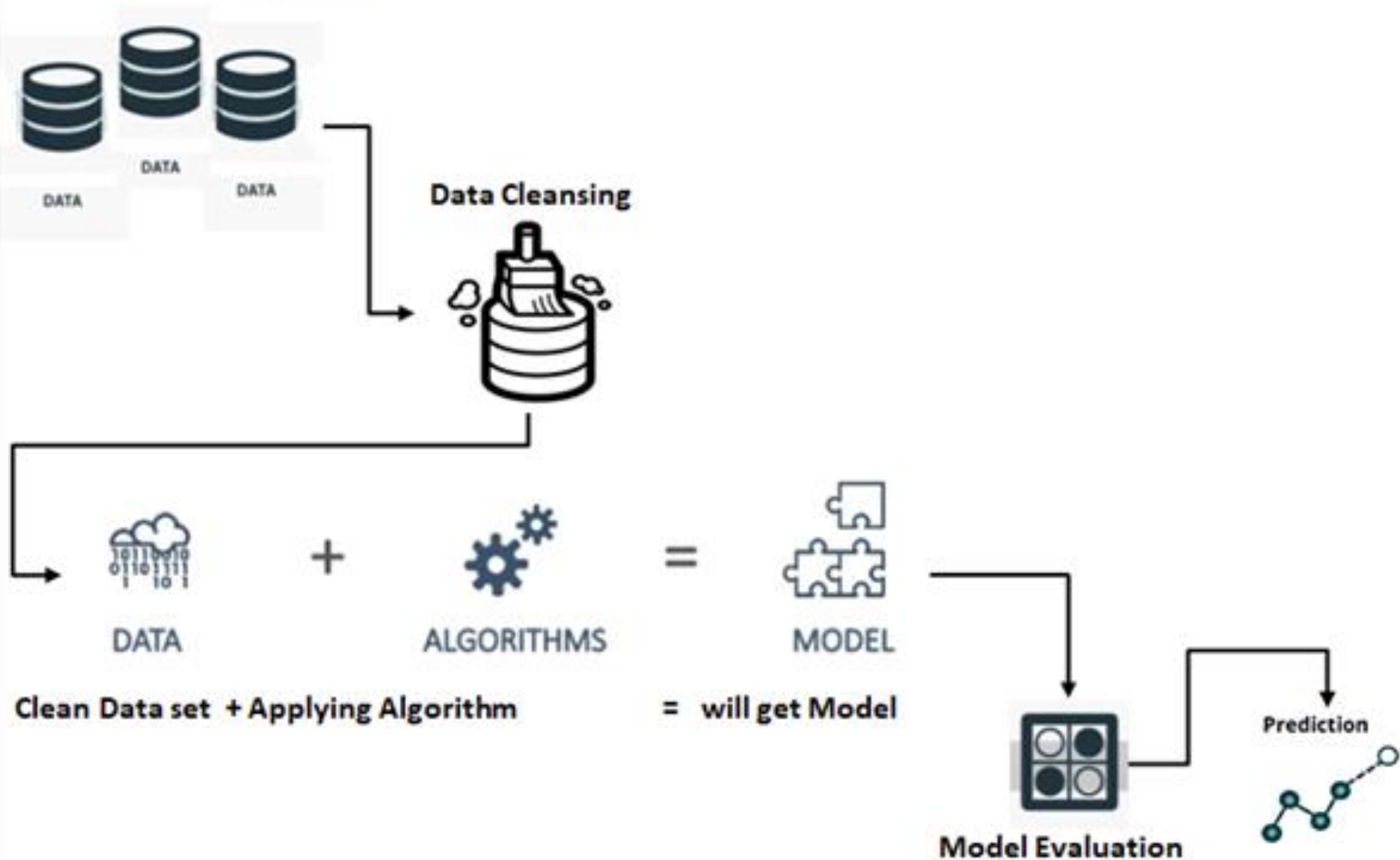
$$\boldsymbol{D} = \{(\boldsymbol{x}_i, y_i)\}$$



# Pipeline général du Machine Learning

1. **Collecte et préparation des données** :  $D = \{ (x_i, y_i) \}$
2. **Préparation et nettoyage des données** : Gestion des données manquantes, normalisation, encodage des variables catégorielles, etc.
3. **Séparation des données** : Division en ensemble d'entraînement et ensemble de test (ex. 80/20 ou 70/30).
4. **Choix d'un modèle d'apprentissage** et d'une fonction de perte  $L$ .
5. **Entraînement** du modèle (optimisation des paramètres  $W$ ).
6. **Évaluation** du modèle sur l'ensemble de test.
7. **Ajustement** des hyperparamètres (validation croisée).

## Data Collection from Different Sources





## Exemple: Reconnaissance de chiffres manuscrits



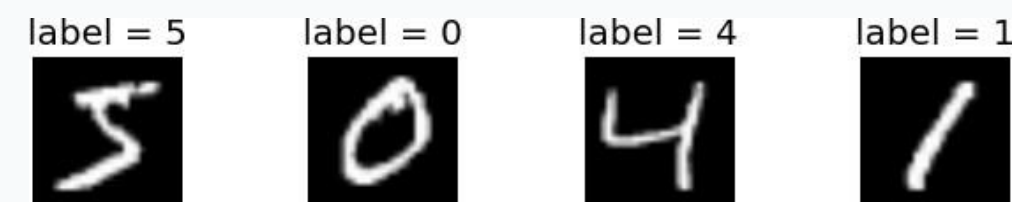
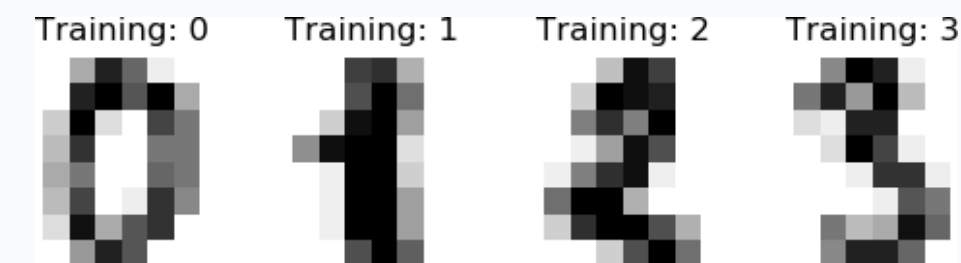
- Comment définir les éléments ?  
 $D, W, x, y$
- Les fonctions d'apprentissage et de prédiction?

$$D \mapsto W$$

$$W, x \mapsto y$$

# Etape 1: choix de la base de données

- Elle existe:
  - Scikit-learn:
  - MNIST:
  - SVHN:



- Il faut la construire:
  - Recueil de données existantes
  - Expérimentations (photos, mesures...)

# Etape 2: mise en forme des données

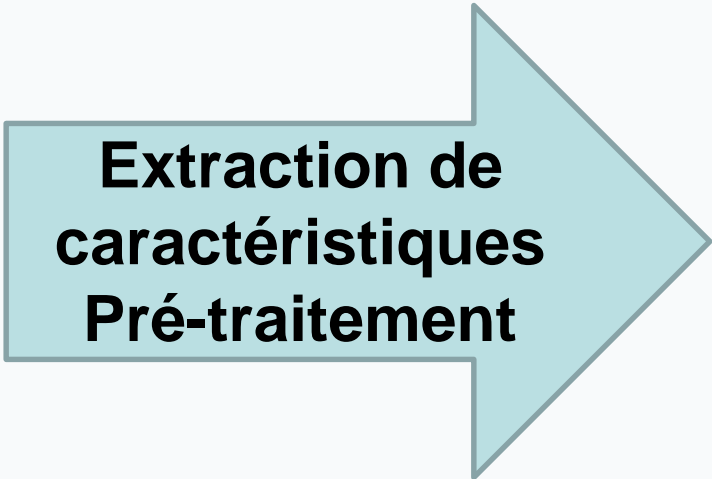
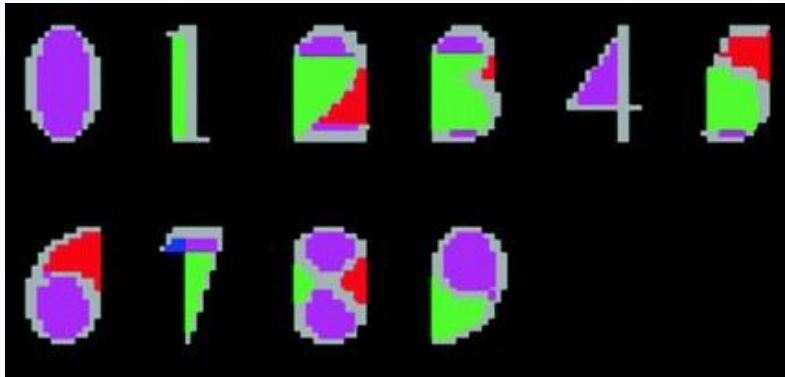


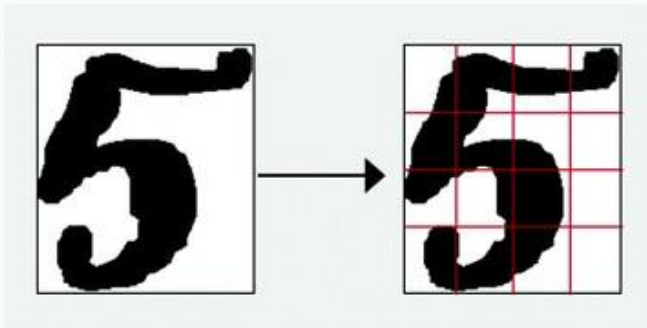
Image (2D) *grande dimension, bruitée, hétérogène*



Profils



Cavités

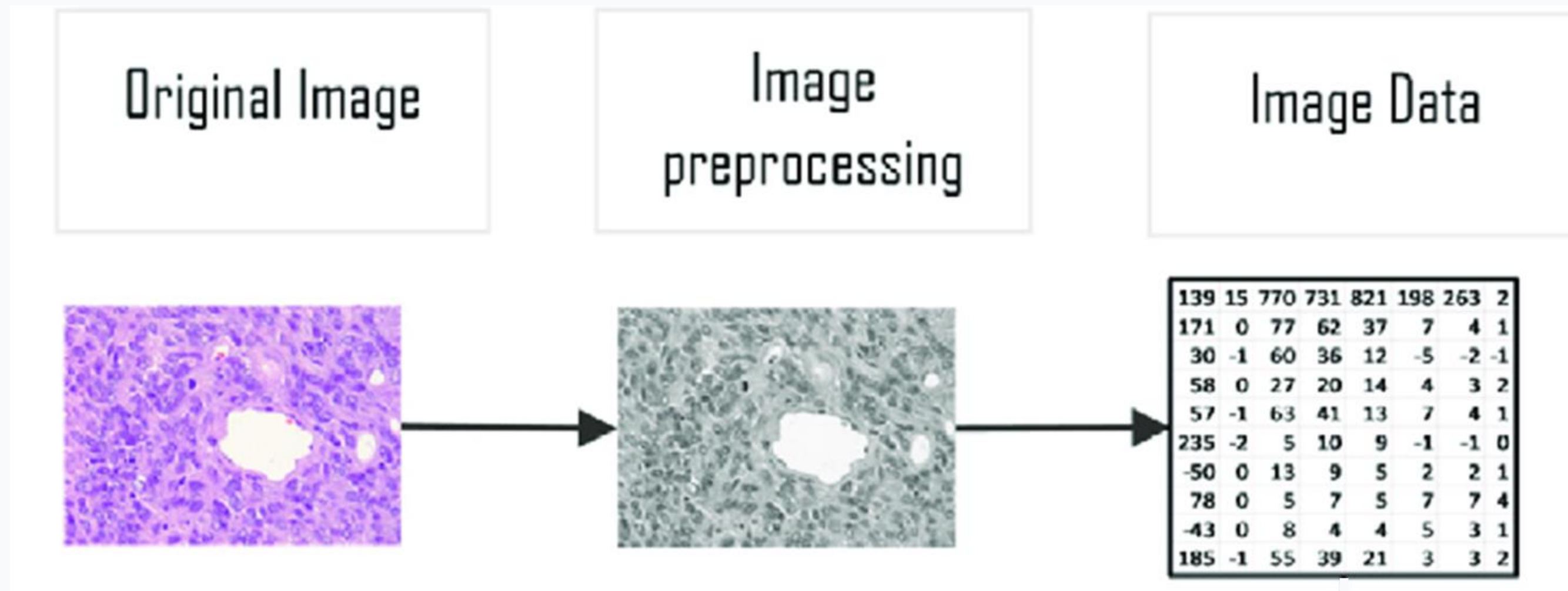


Occupation de zones

0
0
0
0
0
1
1
0
0
0
1
0
0

$x$  = Vecteur  
*petite dimension, homogène, « propre »*

## Etape 2: mise en forme des données





## Etape 3: choix de l'approche

- Quel type de fonction et de problème d'apprentissage?
  - Classification
  - On connaît les classes cibles → Apprentissage supervisé
- Nature des données?
  - Vecteurs de taille fixe mais grands → algorithmes avec bon contrôle de la régularisation
- Taille de la base de données?
  - Grande (> 10000 exemples) → optimisation efficace
- Nature fonctionnelle des prédicteurs?
  - Arbres de décision, SVM, Réseaux de neurones...

# Types d'apprentissage

- **Apprentissage supervisé**
  - Les données d'apprentissage contiennent les objectifs de prédiction (annotations)
- **Apprentissage non supervisé**
  - Les données d'apprentissage sont brutes
- **Apprentissage semi-supervisé**
  - Les données d'apprentissage sont partiellement annotées
- **Apprentissage par transfert**
  - Les données d'apprentissage sont proches du problème visé
- **Apprentissage par renforcement**
  - Les prédictions sont issues d'une séquence d'actions et sont caractérisées par une mesure de qualité (« reward »)



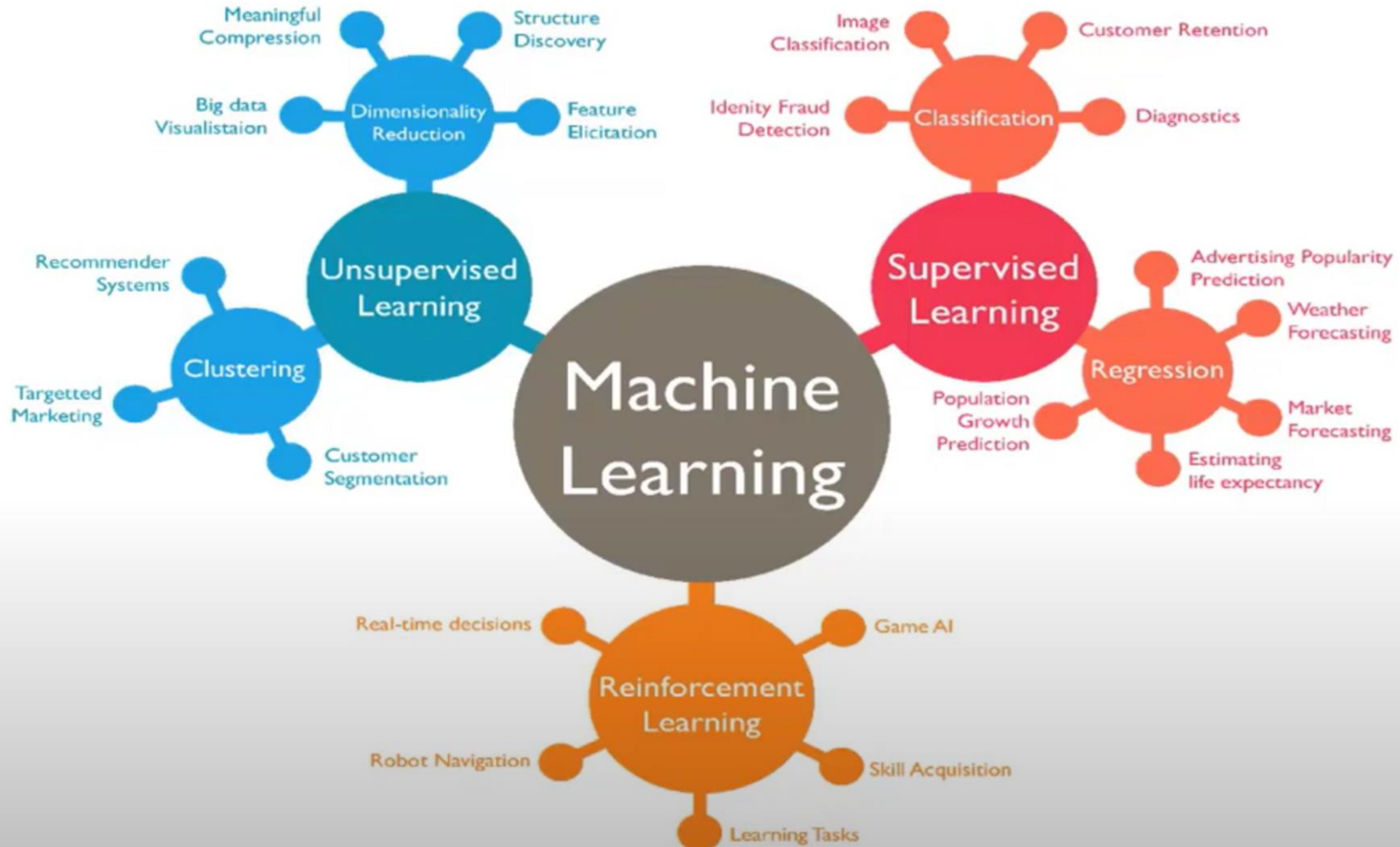
**Labelled** Dataset  
(données étiquetées)  
**Supervised Learning**



**Unlabelled** Dataset  
(données non-étiquetées)



# Machine Learning : Types





# Types de prédictions

- **Classification**
  - Binaire: spam / non spam
  - Identification: « tata Monique »
- **Régression**
  - Prédiction de température, de cours de bourse
  - Localisation d'objet dans image
  - Commande
- **Structure**
  - Graphe des articulations d'une personne
- **Regroupement**
  - Photos dans base de données personnelle
- **Texte**
  - « C'est un chat qui saute sur une table. »