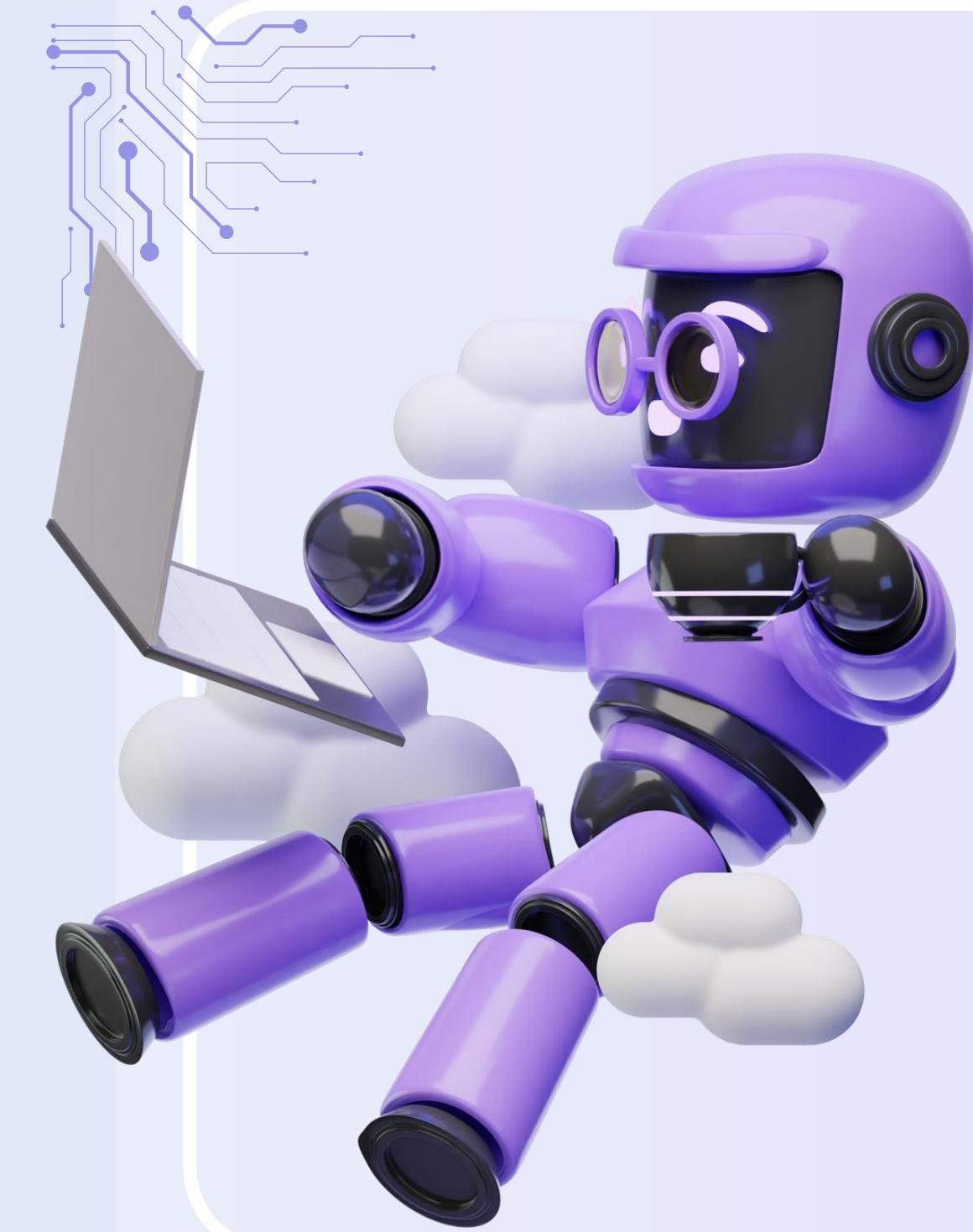




MACHINE LEARNING

PRESENTED BY:
PR. SANAA EL FILALI

ANNÉE UNIVERSITAIRE 2024-2025





80%
des emplois tech utiliseront l'IA
et le Machine Learning d'ici
2030.



LE MONDE CHANGE : SEREZ-VOUS PRÊT ?



Les entreprises recherchent des experts en Machine Learning pour prendre des décisions stratégiques basées sur les données.



Les IA remplacent déjà certaines tâches humaines dans la finance, la médecine et l'industrie.

Ignorer cette révolution, c'est prendre le risque d'être dépassé par l'IA au lieu de l'exploiter !



Automatiser des tâches complexes et améliorer l'efficacité.



Prendre des décisions basées sur les données plutôt que sur l'intuition.



Acquérir une compétence incontournable pour les métiers de demain.



savent ce que vous aimez regarder



prédict vos achats



détectent les fraudes en temps réel

Grâce au **Machine Learning !**

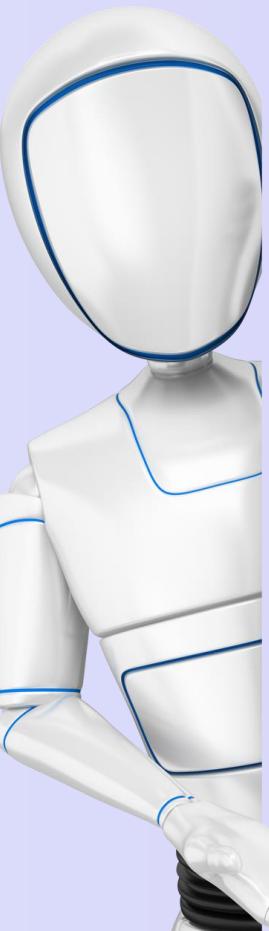


L'IA ne remplacera pas les humains, mais les humains qui savent utiliser l'IA remplaceront ceux qui ne le savent pas.



PLAN

- Séance 1 : Introduction au Machine Learning
- Séance 2 : Python pour ML (Algèbre Linéaire)
- Séance 3 : Préparation des Données - Encodage
- Séance 4 : Régression Linéaire
- Séance 5 : Régression Logistique
- Séance 6 : Arbres de Décision - PART I
- Séance 7 : Arbres de Décision - PART II
- Séance 8 : KNN (Métriques)
- Séance 9 : K-Means (Métriques)
- Séance 10 : DBSCAN - PART I
- Séance 11 : DBSCAN - PART II
- Séance 12 : Suivi de Projet





Séance 1

INTRODUCTION AU MACHINE LEARNING





QU'EST-CE QUE LE MACHINE LEARNING
?

DÉFINITION GÉNÉRALE

- Le Machine Learning (ou apprentissage automatique) est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur la création de systèmes capables :
 - D'apprendre à partir de données,
 - D'identifier des motifs,
 - De prendre des décisions avec une intervention humaine minimale.





AUTRES DÉFINITIONS



ARTHUR SAMUEL (1959)

Le Machine Learning est le champ d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés



TOM MITCHELL (1997)

Un programme apprend à partir de l'expérience E, pour une tâche T et une mesure de performance P, si sa performance à la tâche T, mesurée par P, s'améliore avec l'expérience E.



EXEMPLE

.Arthur Samuel a développé un programme de jeu de dames capable d'apprendre et de s'améliorer en jouant contre lui-même, sans stratégie prédéfinie.



INTERPRÉTATION :

Cette définition met l'accent sur l'amélioration des performances en fonction des expériences accumulées, avec une évaluation mesurable.



ETHEM ALPAYDIN

Le Machine Learning est un ensemble de méthodes statistiques qui permettent de trouver des modèles dans des données.



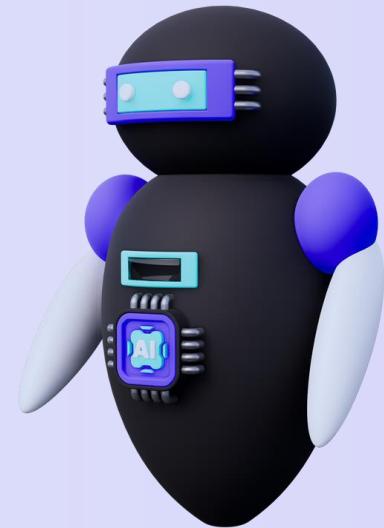
EXPLICATION

Il souligne l'importance des statistiques dans l'extraction d'informations utiles à partir de grands ensembles de données.



MACHINE LEARNING TIMELINE

Période	Événement Clé	Impact
1764	Théorème de Bayes 	Bases des probabilités conditionnelles.
1842	Ada Lovelace écrit le 1er algorithme 	Naissance du concept de programmation.
1952	Arthur Samuel crée un programme d'auto-apprentissage 	Premiers pas du Machine Learning.
1959	MADALINE, premier réseau neuronal 	Première IA utilisée en industrie.
2006	Deep Learning (Geoffrey Hinton) 	IA capable d'apprendre avec des grands volumes de données.
2016	AlphaGo bat un champion humain 	L'IA dépasse l'homme dans un jeu complexe.
2018	AlphaFold prédit la structure des protéines 	Révolution en biologie et santé.



MACHINE LEARNING TIMELINE

- Bayes' theorem is published.

$$P[A_n/B] = \frac{P[B/A_n] \cdot P[A_n]}{\sum P[B/A_i] \cdot P[A_i]}$$

1764

1842

1847

1936

1952

1959

1985

2006

2012

2016

2018

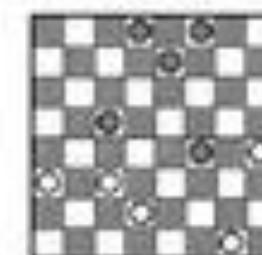
- Ada Lovelace lays the foundations of the first algorithm.



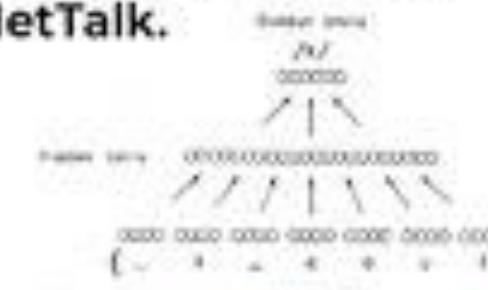
- George Boole creates Boolean Logic.



- Arthur Samuel creates the first computer program at IBM.



- Sejnowski and Rosenberg create the neural network, NetTalk.



- Google creates an unsupervised neural network.



- MADALINE, the first artificial neural network, is created.



- Geoffrey Hinton invents the term "Deep Learning".



- AlphaFold 1, a technology capable of predicting protein structures, is created.



MACHINE LEARNING TIMELINE

- Bayes' theorem is published.

$$P[A_n/B] = \frac{P[B/A_n] \cdot P[A_n]}{\sum P[B/A_i] \cdot P[A_i]}$$

1764

1842

1847

1936

1952

1959

1985

2006

2012

2016

2018

Ada Lovelace lays the foundations of the first algorithm.



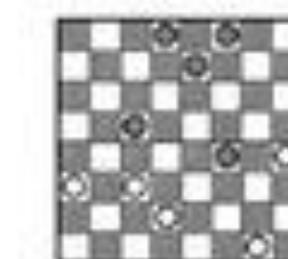
Alan Turing proposes a machine that can learn.



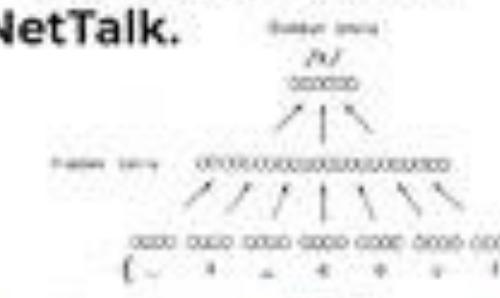
George Boole creates Boolean Logic.



Arthur Samuel creates the first computer program at IBM.



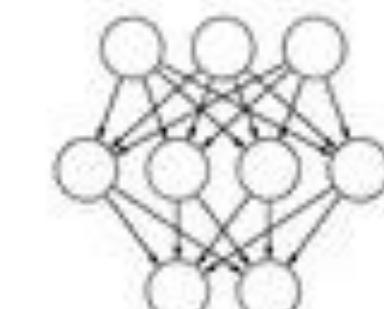
Sejnowski and Rosenberg create the neural network, NetTalk.



Google creates an unsupervised neural network.



Geoffrey Hinton invents the term "Deep Learning".



AlphaFold 1, a technology capable of predicting protein structures, is created.



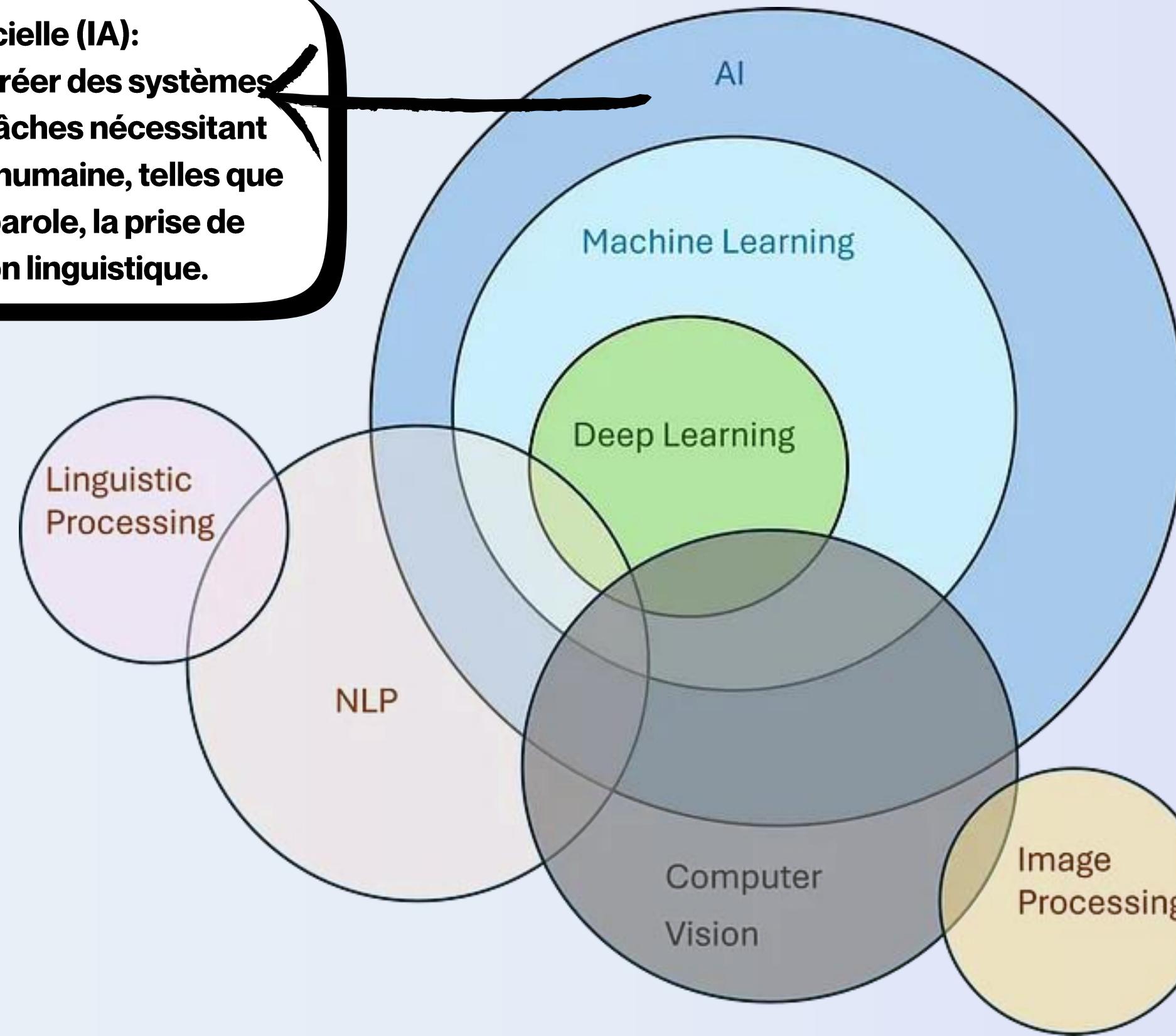
AlphaGo beats the first human player.



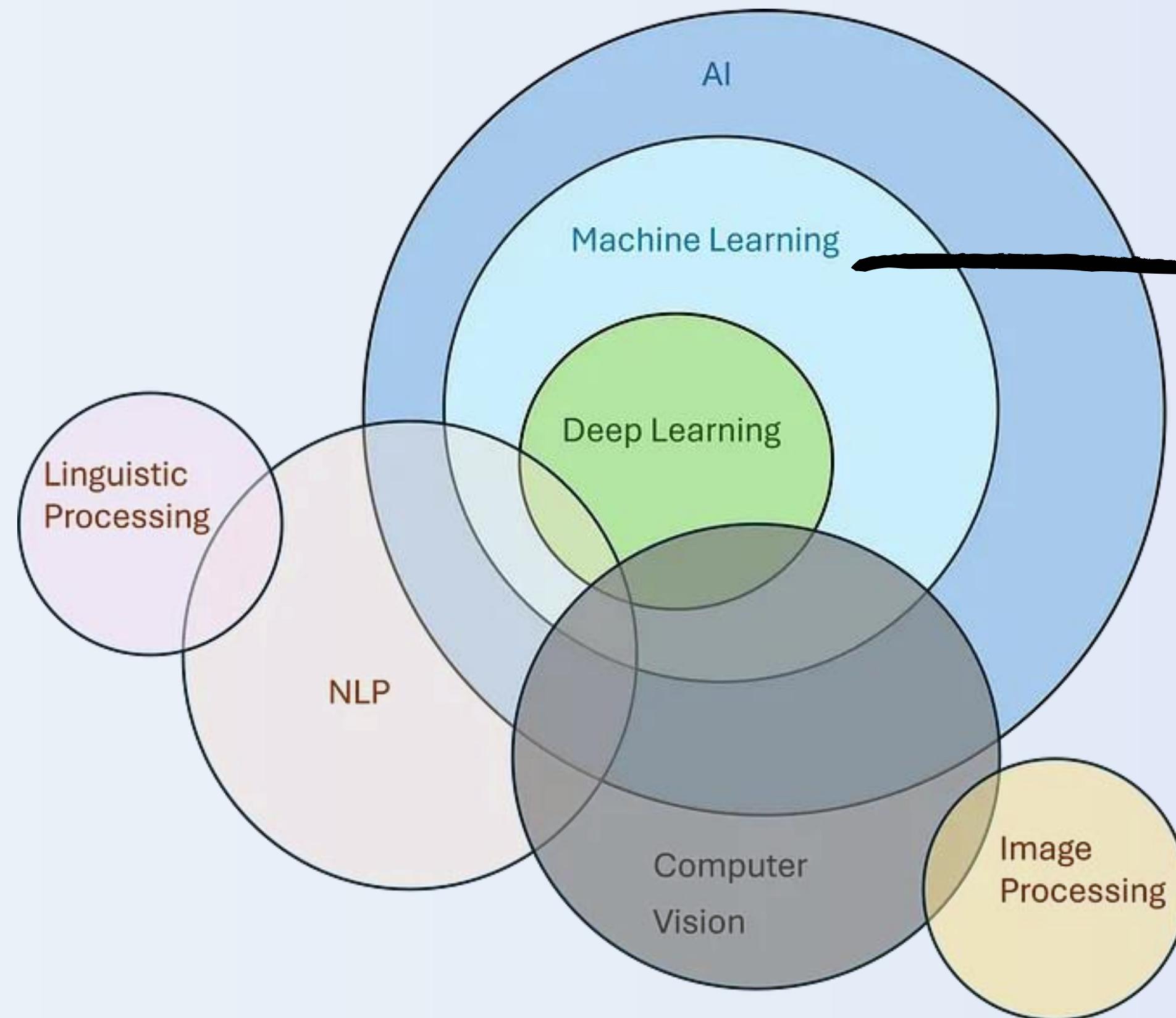
Différence entre IA, Machine Learning et Deep Learning :

Intelligence Artificielle (IA):

Domaine général visant à créer des systèmes capables d'effectuer des tâches nécessitant normalement l'intelligence humaine, telles que la reconnaissance de la parole, la prise de décision et la traduction linguistique.



Différence entre IA, Machine Learning et Deep Learning :

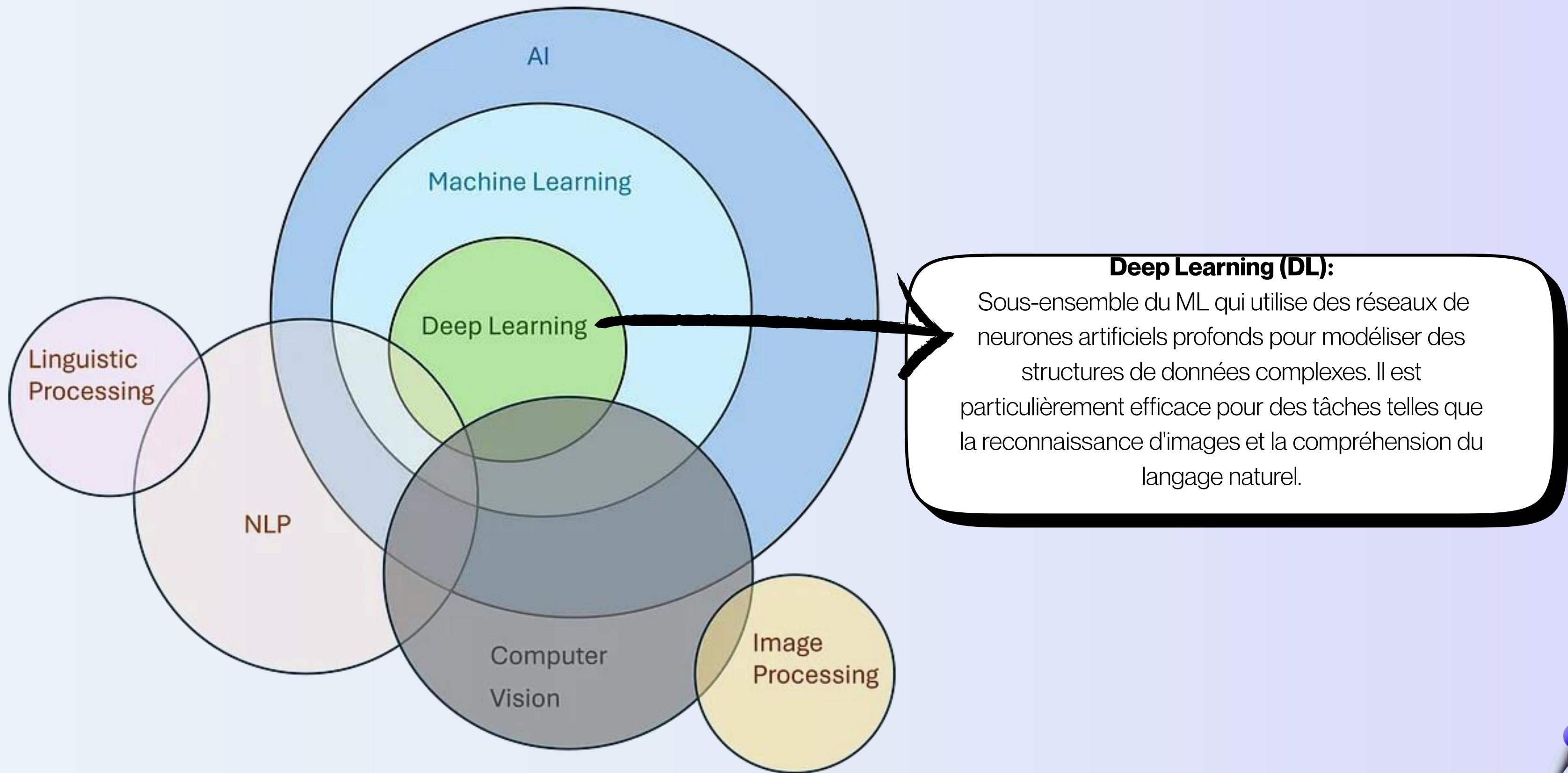


Machine Learning (ML):

Sous-ensemble de l'IA qui se concentre sur le développement d'algorithmes permettant aux machines d'apprendre à partir de données et de faire des prédictions ou des décisions sans être explicitement programmées pour chaque tâche.



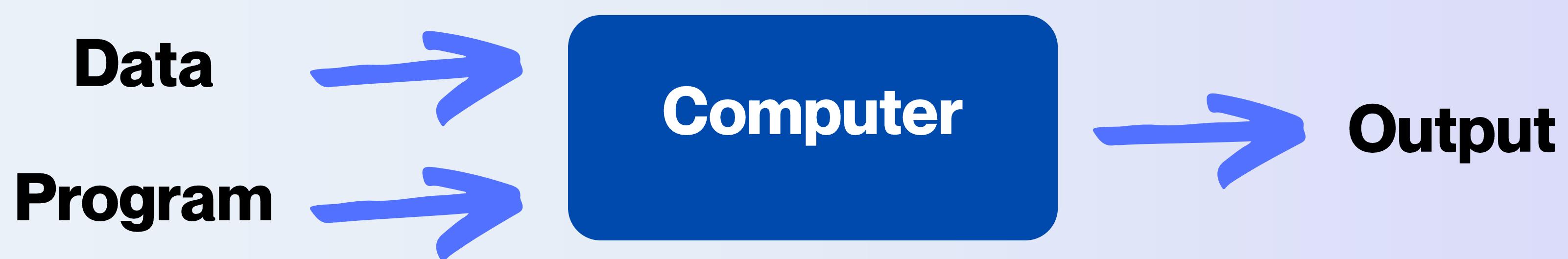
Différence entre IA, Machine Learning et Deep Learning :



Le Changement de Paradigme en

Programmation

De la Programmation Traditionnelle au Machine Learning



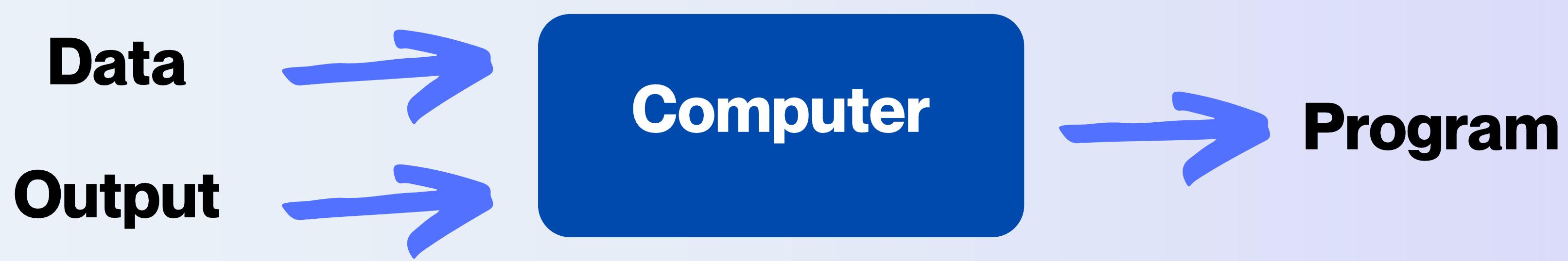
➡ Programmation traditionnelle

- ➡ Les règles et instructions (programme) sont codées manuellement
- ➡ Traite les données pour produire un résultat spécifique



Le Changement de Paradigme en Programmation

De la Programmation Traditionnelle au Machine Learning

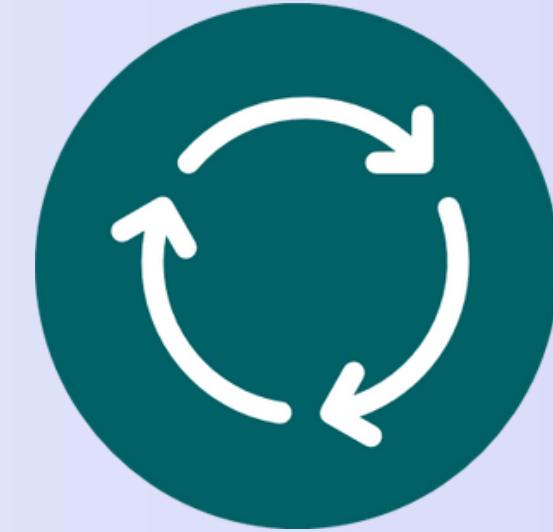
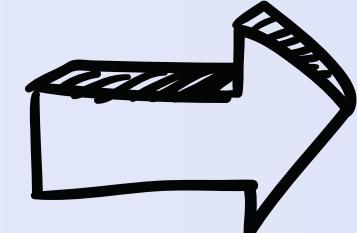
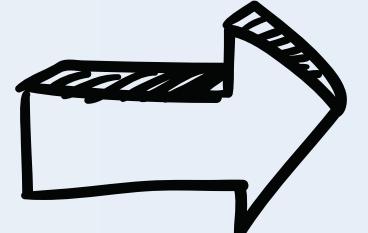


➡ Inverse le processus : les données et les résultats souhaités servent à apprendre les règles (modèle)

➡ Permet aux machines de s'adapter à de nouvelles données sans recodage



Étapes clés d'un projet ML



1 Compréhension du domaine

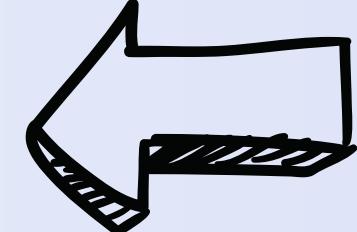
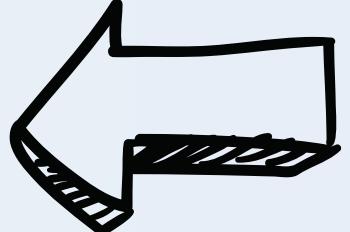
Identifier le problème et les objectifs du projet

2 Collecte des données

Rassembler des données pertinentes et fiables.

3 Traitement des données

Nettoyer, transformer et préparer les données.



6 Déploiement

Intégrer le modèle en production pour une utilisation réelle. Construire et valider le modèle de machine learning.

4 Analyse exploratoire des données

Déetecter des tendances et visualiser les données.





1 - Compréhension du domaine

⚡ Pourquoi c'est important ?

Une bonne compréhension du domaine permet de créer un modèle précis et efficace.



PRÉDIRE LE PRIX DES VOITURES

- Comprendre les spécifications des véhicules
- Analyser les tendances du marché
- Identifier les facteurs clés influençant le prix



DÉTECTION DE FRAUDES BANCAIRES :

- Étudier les comportements normaux et suspects des transactions
- Analyser les schémas de paiement (montant, fréquence, localisation...)
- Repérer les anomalies par rapport aux habitudes des utilisateurs



PRÉDICTION DES VENTES D'UN MAGASIN :

- Étudier la saisonnalité et les périodes de forte demande
- Analyser l'impact des promotions et des campagnes marketing
- Identifier les facteurs externes influençant les ventes (météo, jours fériés...)



DIAGNOSTIC MÉDICAL ASSISTÉ PAR IA :

- Connaître les symptômes et pathologies les plus courants
- Analyser les données médicales des patients (antécédents, examens...)
- Assister les médecins dans la prise de décision grâce aux modèles prédictifs

Objectif

Mieux choisir, collecter et analyser les données pour obtenir un modèle performant.





2 - Collection des données

- ❖ **Définir les besoins – Identifier quelles données sont nécessaires pour le modèle.**
- ❖ **Sources de données – Bases de données, capteurs, API, web scraping, etc.**
- ❖ **Qualité des données – Vérifier la fiabilité, la pertinence et l'exactitude.**



collecte → stockage

Avant :

Nous devons collecter des expériences qui serviront à enseigner au modèle comment accomplir la tâche

Après :

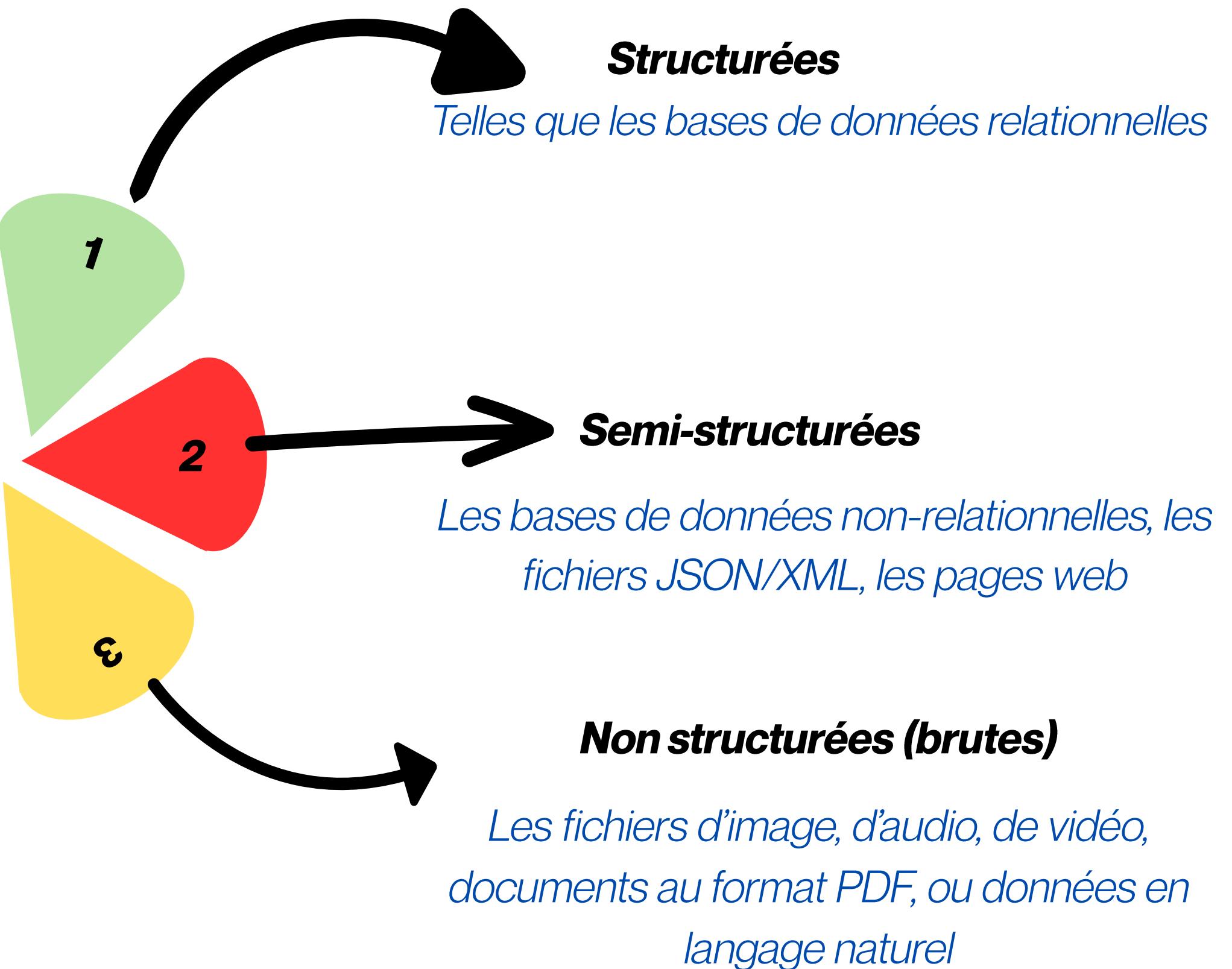
Nous devons rassembler des données pertinentes à partir de différentes sources, puis les organiser et les nettoyer afin d'entraîner efficacement notre modèle.

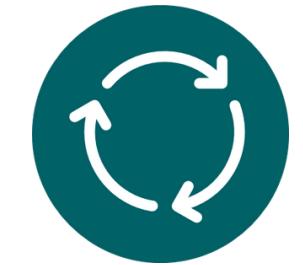




2 - Collection des données

Types de données





3 - Traitement des données

DÉFINITION

Le traitement des données est l'ensemble des étapes permettant de **préparer les données** avant leur utilisation par un modèle de Machine Learning. Cela inclut le nettoyage, la transformation et la structuration des données pour améliorer leur qualité et leur pertinence.



3 - Traitement des données

Problèmes liés aux données structurées



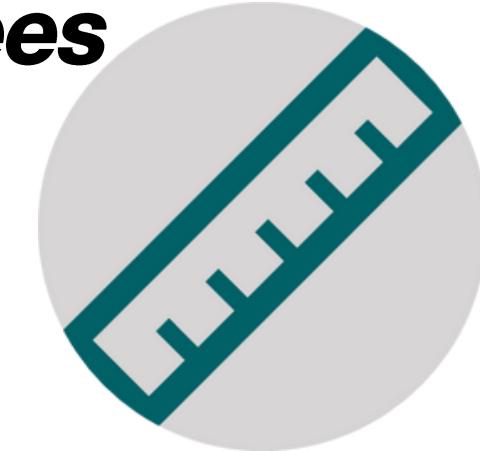
Données manquantes

*Impact sur la qualité des prédictions,
solutions : suppression ou imputation.*



Valeurs aberrantes

*Problèmes de biais, détection
par statistique ou algorithmes.*



Normalisation/Mise à l'échelle

*Harmoniser les plages de valeurs
pour de meilleures comparaisons.*



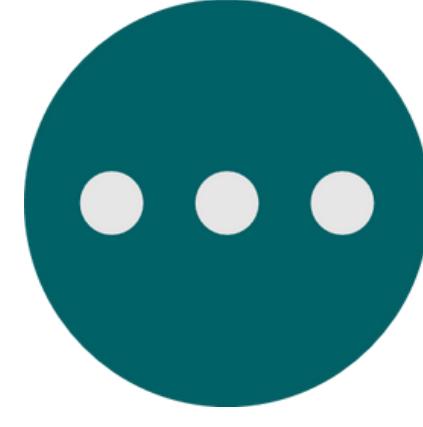
Doublons

*Affectent la diversité des données et
faussent les modèles.*



Types de données incorrects

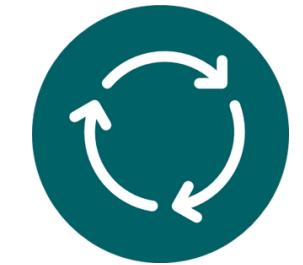
*Incohérences (textes au lieu de
nombres), correction nécessaire.*



Autres erreurs

*Mauvais encodage, données
obsolètes...*

Bien que structurées, les données doivent être traitées pour résoudre plusieurs problèmes qui pourraient entraver le processus d'entraînement du modèle

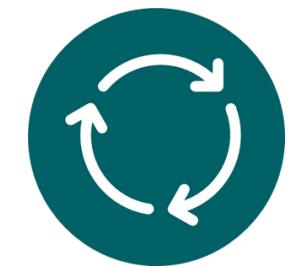


3 - Traitement des données

Avant d'utiliser un jeu de données pour entraîner un modèle, il est essentiel de :

- Vérifier la qualité des données (nettoyage, normalisation, gestion des valeurs aberrantes).
- Vérifier la représentativité de l'échantillon (éviter les biais de population).
- Effectuer une analyse exploratoire pour détecter d'éventuelles anomalies.





4 - Analyse exploratoire des données

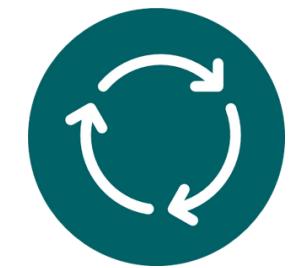
DÉFINITION

L'Analyse Exploratoire des Données (EDA) est une approche utilisée pour examiner, résumer et comprendre un jeu de données avant de l'utiliser dans un modèle de Machine Learning.

Objectifs de l'EDA :

- Déetecter les erreurs et incohérences dans les données.**
- Comprendre la distribution des variables et leurs relations.**
- Identifier les valeurs aberrantes et les corrélations importantes.**
- Guider les choix de transformation et de modélisation des données.**





4 - Analyse exploratoire des données

TECHNIQUES

1 Statistiques

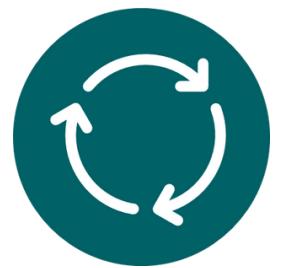
Analyse des statistiques descriptives (moyenne, médiane, écart-type), détection des valeurs aberrantes et étude des corrélations entre variables.

2 Visuelles

Utilisation de histogrammes, boxplots, heatmaps et nuages de points pour visualiser distributions, anomalies et relations entre variables.

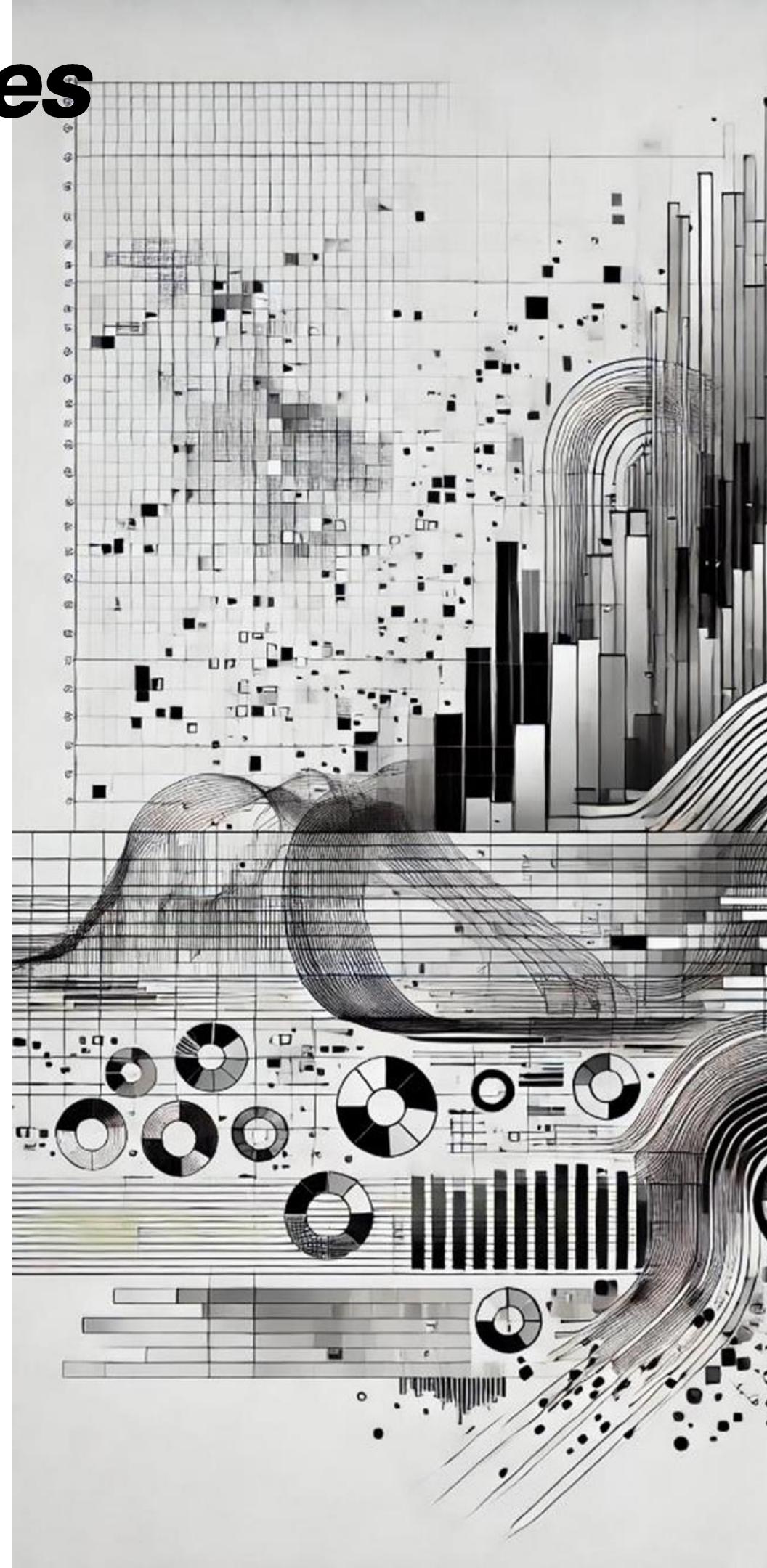


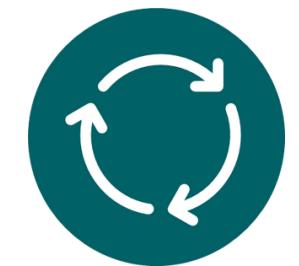
4- Analyse exploratoire des données



EXEMPLES

- Comprendre la distribution du prix des maisons
 - Tracer un histogramme des prix pour voir s'ils sont concentrés autour d'une valeur ou s'ils sont dispersés.
 - Analyser la corrélation entre la surface habitable et le prix
 - Utiliser un nuage de points pour voir si une relation linéaire existe.
 - Déetecter les valeurs aberrantes
 - Avec un boxplot, repérer des maisons dont le prix est anormalement élevé ou bas par rapport aux autres.
 - Traiter les valeurs manquantes
 - Vérifier si des données importantes sont absentes (ex : certaines maisons sans surface indiquée) et choisir une méthode pour les compléter (moyenne, médiane, interpolation...).
-  Conclusion :
- L'EDA est une étape essentielle pour éviter les erreurs, maximiser l'efficacité du modèle et prendre des décisions basées sur une bonne compréhension des données.





5 - Entraînement et tests

DÉFINITION

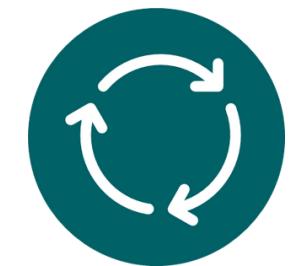
L'entraînement et les tests sont deux étapes clés du Machine Learning permettant de construire et d'évaluer la performance d'un modèle.

Objectifs

Entraîner le modèle sur des données connues pour apprendre des motifs.

Tester le modèle sur des données inconnues pour évaluer sa capacité de généralisation.





5 - Entraînement et tests

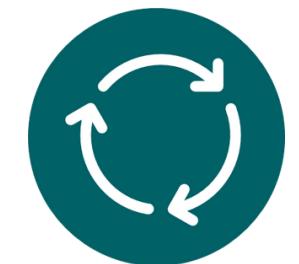
DÉFINITION

L'entraînement et les tests sont deux étapes clés du Machine Learning permettant de construire et d'évaluer la performance d'un modèle.

Objectifs

- Entraîner le modèle sur des données connues pour apprendre des motifs.**
- Tester le modèle sur des données inconnues pour évaluer sa capacité de généralisation.**



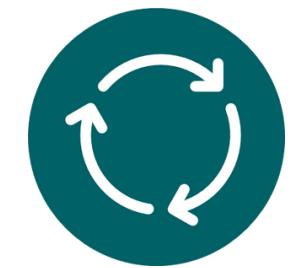


5 - Entraînement et tests

TÉCHNIQUES

- Entraînement supervisé : Avec des données labellisées (ex. : prix d'une maison).
- Entraînement non supervisé : Sans labels, pour découvrir des motifs cachés (ex. : clustering).
- Apprentissage par renforcement : Basé sur un système de récompense pour améliorer la prise de décision





6 - Déploiement

DÉFINITION

Le déploiement d'un modèle de Machine Learning consiste à le rendre opérationnel dans un environnement réel, où il peut traiter de nouvelles données et produire des prédictions exploitables.

Objectifs du déploiement

- Intégrer le modèle dans une application, un site web ou un système d'information.**
- Assurer une scalabilité et une robustesse pour gérer des données en continu.**
- Faciliter la mise à jour et le suivi des performances en production.**

Types de déploiement

- ◊ **Déploiement local : Sur un ordinateur ou un serveur privé.**
- ◊ **Déploiement cloud : Sur AWS, Azure, Google Cloud...**
- ◊ **Déploiement embarqué : Sur un appareil IoT ou un mobile.**





1 Questions générales



- Qu'est-ce que le Machine Learning et en quoi se distingue-t-il de la programmation traditionnelle ?
- Quels sont les principaux types d'apprentissage en Machine Learning ? Donnez un exemple pour chacun.
- Pourquoi est-il important de bien comprendre le domaine avant de développer un modèle de Machine Learning ?



EXERCICE 1



Nous voulons créer une IA qui prédit les prix des maisons

Rédigez un paragraphe où vous identifiez et expliquez les différents paramètres qui, selon vous, influencent le prix d'une maison.
Justifiez pourquoi ces paramètres sont importants et comment ils peuvent affecter la valeur de la propriété

