

Big Data et IA

Monica-Juliana Pérez

Ordre du jour

- Accueillir
- Programme et thèmes du cours
- Règles et répartition des notes
- Prérequis et logiciels/outils recommandés :
- Introduction à l'IA et au Big Data

Accueillir

Objectifs d'apprentissage :

À la fin du cours, les étudiants seront capables de :

1. Comprendre les fondamentaux de l'IA et du Big Data et leurs applications en logistique.
2. Analysez de grands ensembles de données pour extraire des informations exploitables pour les opérations logistiques.
3. Concevez des solutions basées sur l'IA pour l'optimisation de la chaîne d'approvisionnement.
4. Évaluer les considérations éthiques et pratiques liées à l'utilisation de l'IA dans la logistique.

Monica-Juliana PÉREZ

Monica.PerezMorales@uphf.fr

Évaluation

- 50% Projet
- 40% Exercices (TP)
- 10% de participation aux cours

- **Prérequis :**

- Connaissances de base en programmation (Python de préférence) ou ChatGPT
- Connaissance des outils d'analyse de données (Excel, SQL).

- **Logiciels/outils recommandés :**

- Python (NumPy, Pandas, Scikit-learn, TensorFlow).



Google Collab



Comment nous prendre des décisions ?





A photograph of a woman with short dark hair and glasses, wearing a dark floral shirt and a white apron, cooking in a kitchen. She is standing at a stove, holding a blue lid and a wooden spoon, stirring a pot. The kitchen has white cabinets and a tiled backsplash. The text is overlaid on the image in a large, white, sans-serif font.

« intuition » ou « logique »

simplement savoir quand quelque chose est bien ou mal

Pratique

Observation

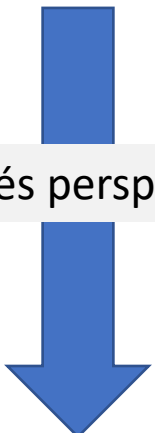
Apprentissage

Enseignement

Expérience

Réflexion

Apprentissage



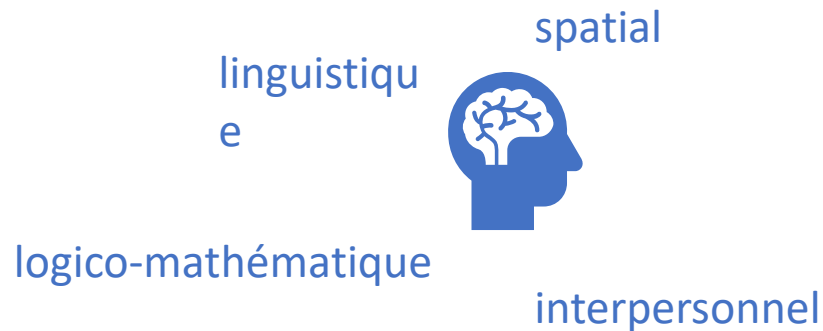
capacités perspicaces

Intelligence

Intelligence

« L'intelligence est la capacité globale de l'individu à agir avec détermination, à penser rationnellement et à gérer efficacement son environnement. »

David Wechsler



Intelligence artificielle (IA)

Le terme « intelligence artificielle » (IA) a été inventé par John McCarthy en 1956. De nombreuses définitions de l'IA ont été proposées par des scientifiques et des chercheurs, notamment :

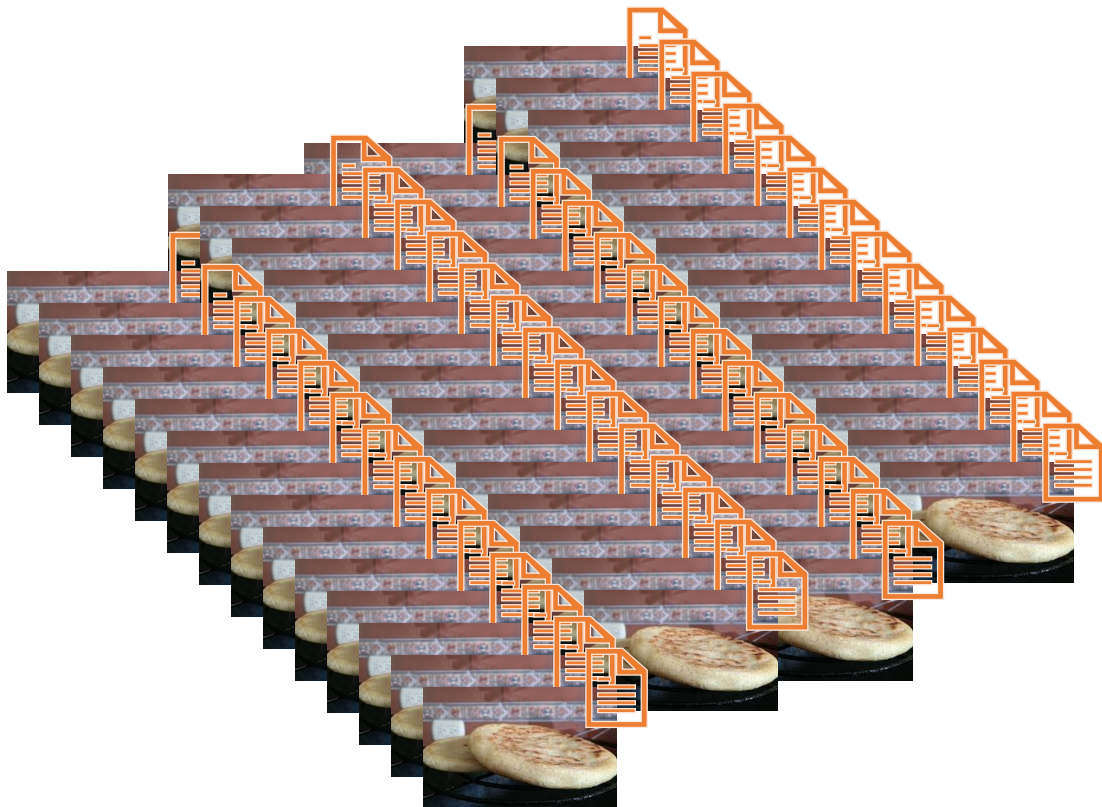
- L'IA est le domaine d'étude axé sur la création de programmes et de machines **capables d'effectuer des tâches qui nécessitent généralement l'intelligence humaine**, telles que le raisonnement, l'apprentissage, la perception et le langage.

Big Data

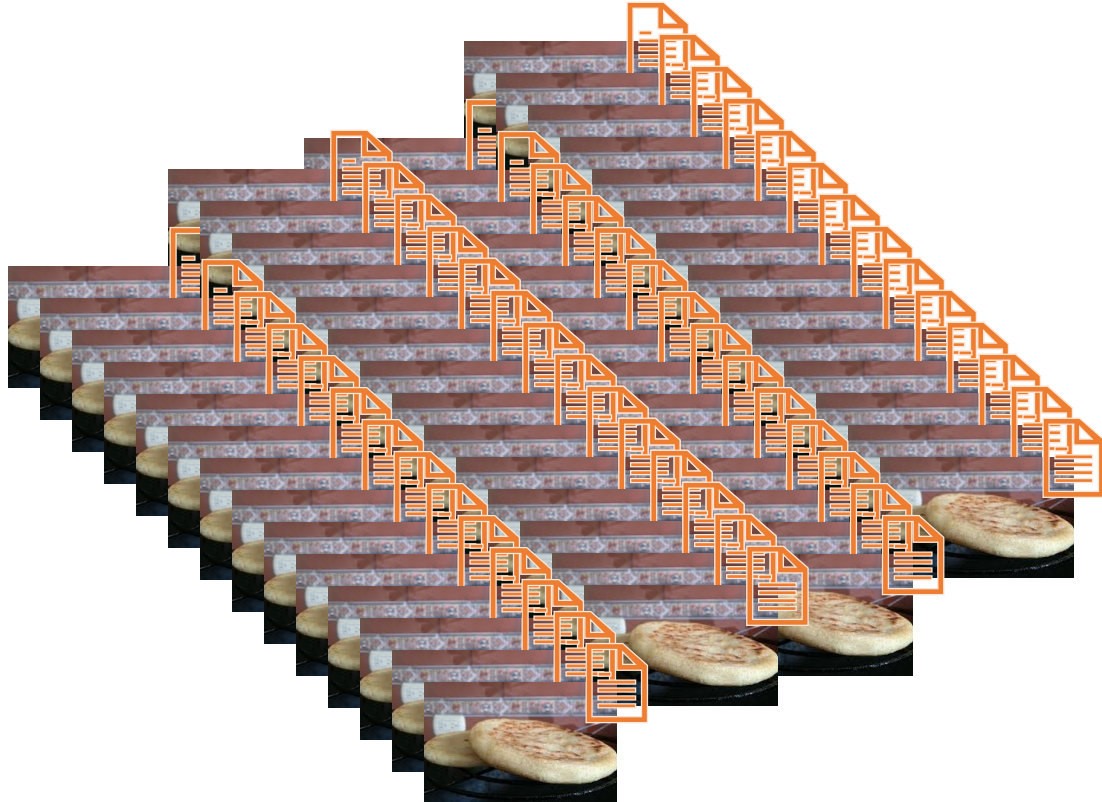
Le terme big data désigne spécifiquement les **grands ensembles de données dont la taille est si importante que la quantité ne peut plus tenir dans la mémoire utilisée par les ordinateurs pour le traitement**. Ces données peuvent être capturées, stockées, communiquées, agrégées et analysées. Il n'existe pas de définition spécifique de la taille du big data, comme le nombre de téraoctets ou de gigaoctets.



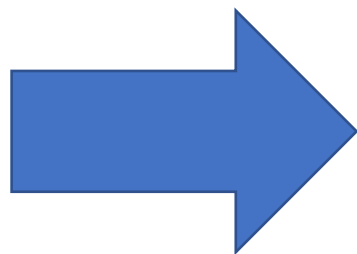
DONNÉES



GRAND
DONNEES



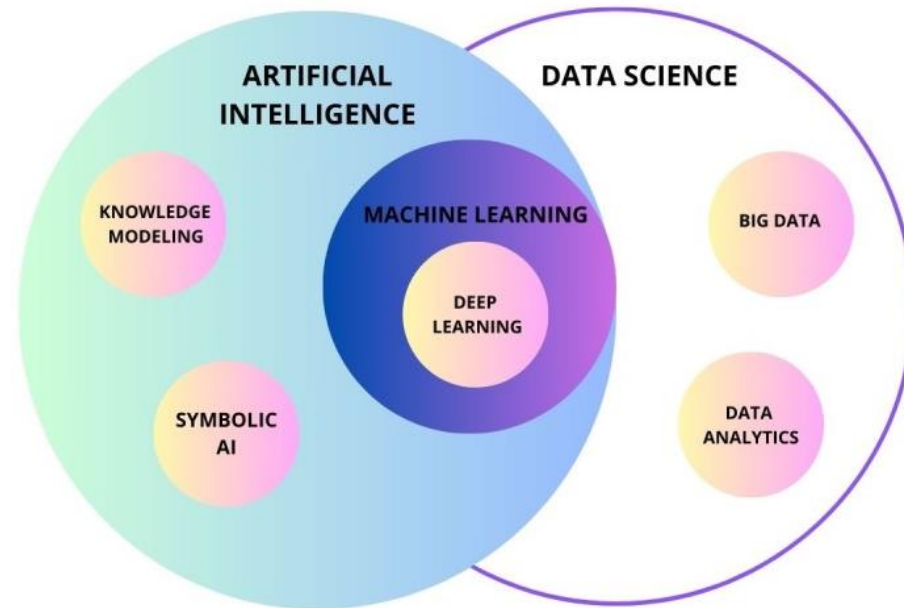
GRAND DONNÉES



ARTIFICIEL
Intelligence (IA)

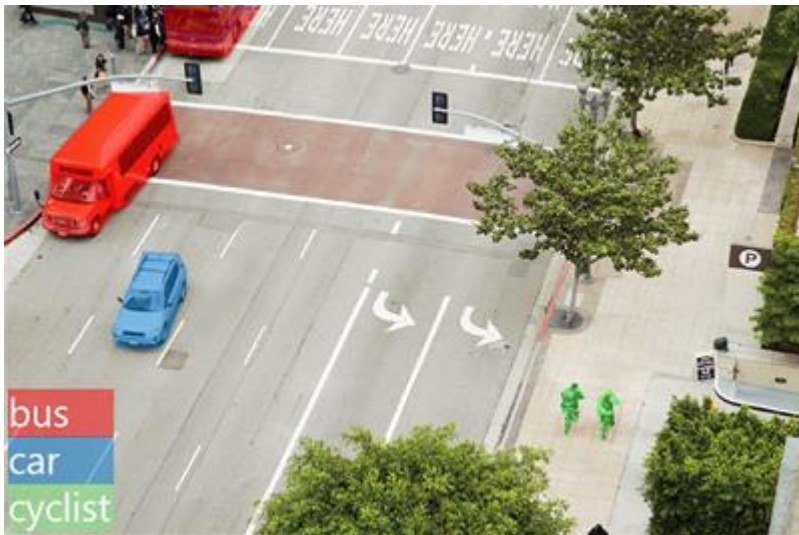
Différences entre **Big Data** et **Intelligence Artificielle**

Le **Big Data** et **l'IA** sont souvent utilisés en conjonction avec un autre, mais chaque remplir très des rôles différents, un est **informations** et la autre est un **traitement de que information**



INTELLIGENCE ARTIFICIELLE (IA)

L'IA est un logiciel qui imite les comportements et les capacités humaines .
Principales charges de travail inclure



Apprentissage automatique - Ceci est souvent le fondation pour un système d'IA , et est le chemin nous « enseignons » un ordinateur modèle à faire pronostics et tirage au sort conclusions à partir des données.

Ordinateur Vision - Capacités au sein de l'IA pour interpréter le monde visuellement à travers des caméras, des vidéos et des images .

Langage naturel Traitement - Capacités au sein de l'IA pour un ordinateur à interpréter écrit ou parlé langue , et répondre en nature

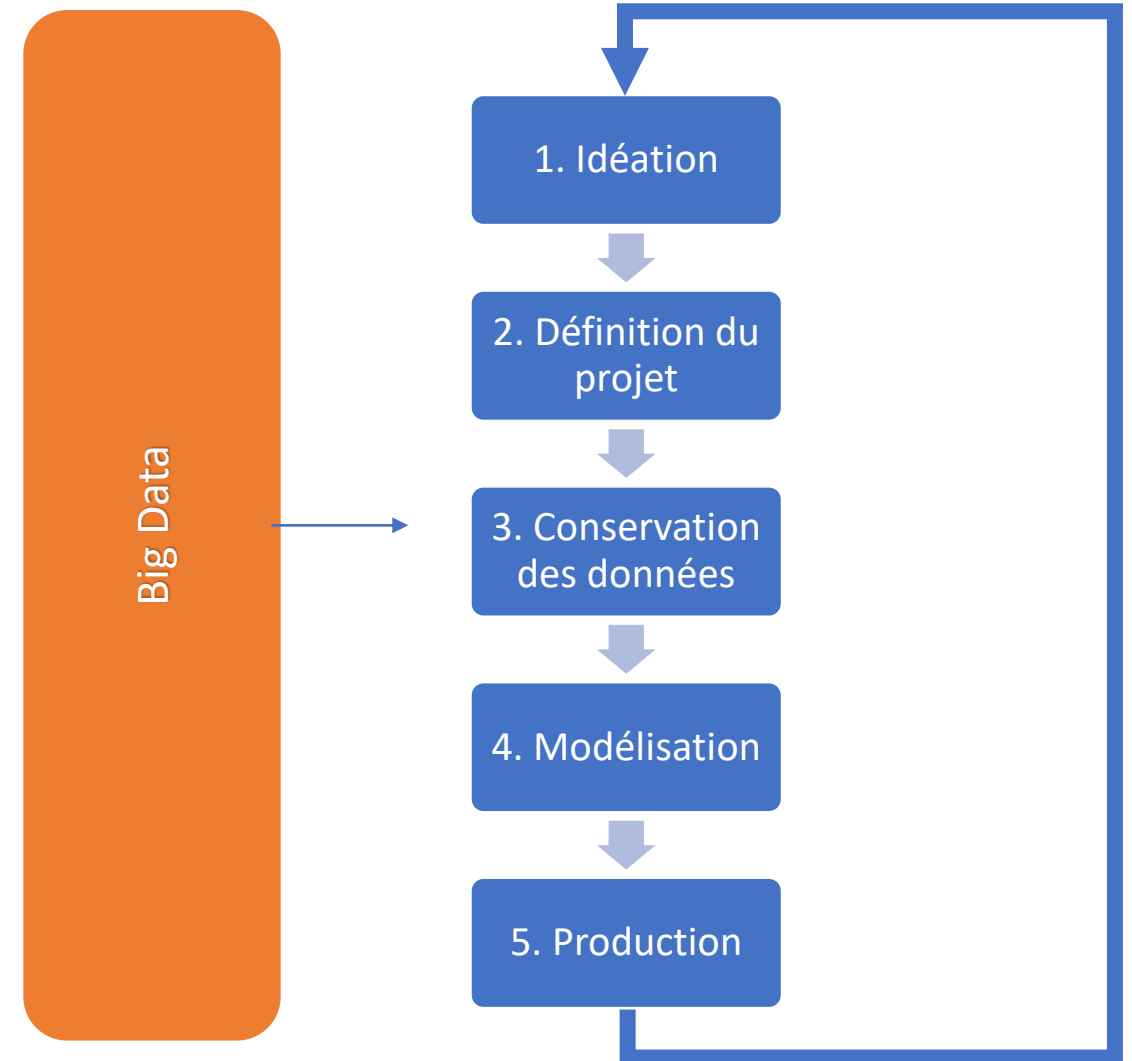
Document intelligence - Capacités au sein de l'IA qui traite de gestion , traitement et utilisation haut volumes des données trouvées dans les formulaires et les documents .

Connaissance exploitation minière - Capacités au sein de l'IA pour extrait information depuis grand volumes de souvent données non structurées à créer un élément consultable magasin de connaissances .

IA générative - Capacités au sein de l'IA qui créent du contenu original dans une variété de formats, notamment le langage naturel, l'image, le code, etc.

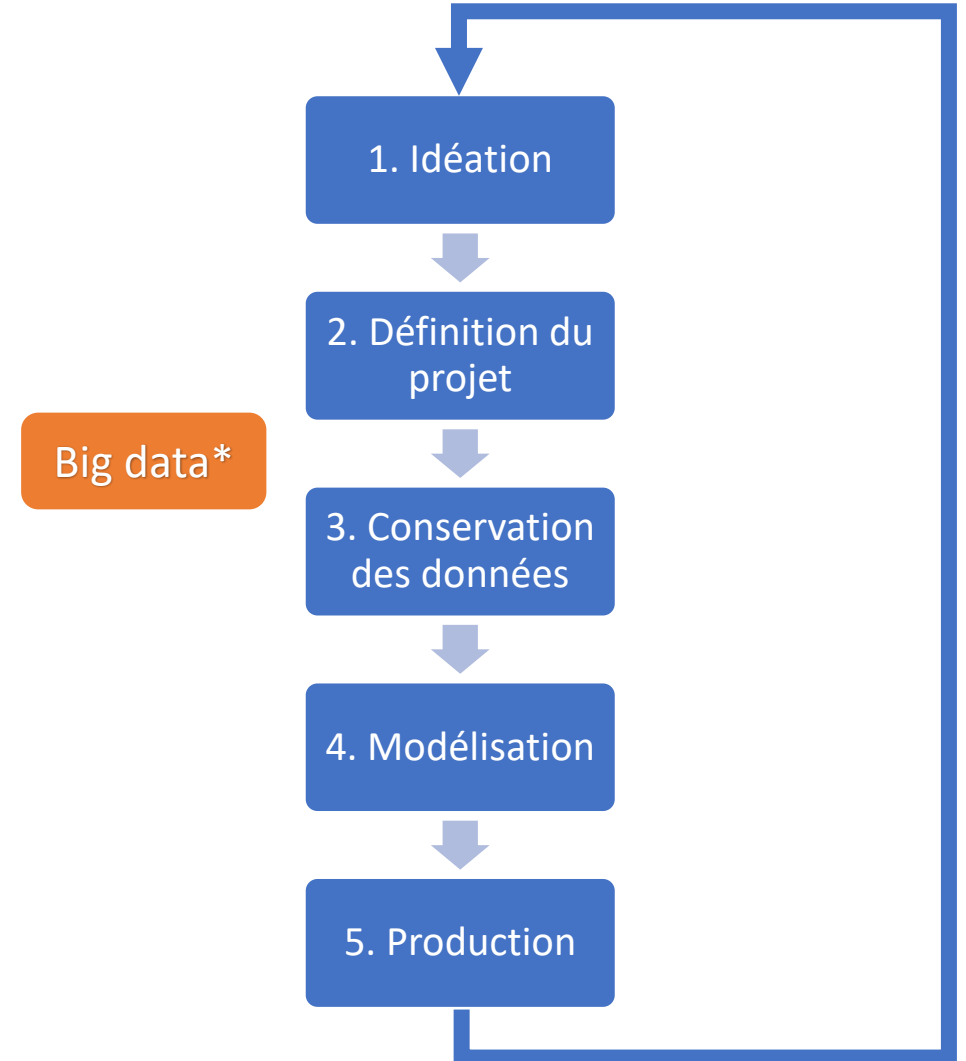
Comment commencer ?

L'IA en tant que projet et non en tant que méthode



Comment commencer ?

L'IA en tant que projet et non en tant que méthode



* Pas nécessaire mais voir il ici pour raison de ce cours

Idéation

Il s'agit de l'étape initiale où les idées pour le projet d'IA sont générées.

- Identifier les problèmes à résoudre
- Analyser les opportunités
- Définir l'objectif du projet.



Quelques méthodologies

- Remue-méninges
- Cadrage du problème
- Persona de l'utilisateur
- Cartographie de l'empathie
- Points douloureux
- Pipeline d'innovation
- Validation d'idée
- Atelier d'idéation

Quelques méthodologies

- Remue-méninges
- Cadrage du problème
- Persona de l'utilisateur
- Cartographie de l'empathie
- Points douloureux
- Pipeline d'innovation
- Validation d'idée
- Atelier d'idéation

Comprendre l'entreprise



Comprendre environnement

Idéation

- Identifier les problèmes à résoudre :
 - mon ami est végétarien ?
 - Des allergies ?
- Analyser les opportunités
 - Que puis-je cuisiner ?
- Définir l'objectif du projet.
 - Je vais cuisiner...

Exemple : J'ai un dîner à la maison.



2. Définition du projet

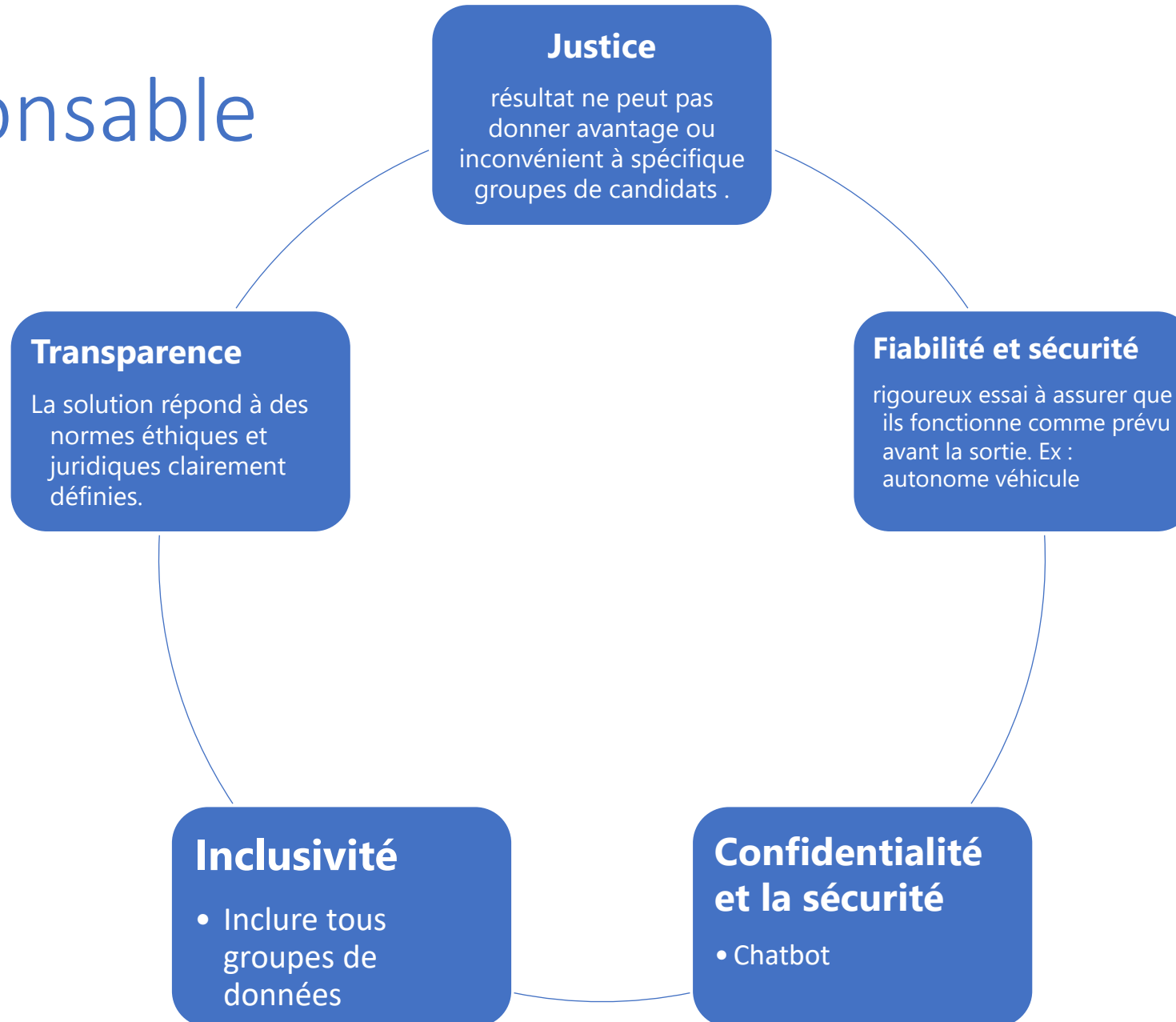
- À ce stade, la portée du projet, les objectifs, les buts spécifiques et les mesures de réussite sont clairement définis.
- Les ressources nécessaires sont également identifiées.
 - Temps
 - Technologie
 - membres de l'équipe



Défis et risques liés à l'IA

Défi ou Risque	Exemple
Les préjugés peuvent affecter résultats	Un modèle d'approbation de prêt discrimine selon le sexe en raison d'un biais dans les données avec lesquelles il a été formé
Les erreurs peuvent causer des dommages	Un véhicule autonome subit une défaillance du système et provoque une collision
Les données pourraient être exposées	Un robot de diagnostic médical est formé à l'aide de données sensibles sur les patients, qui sont stockées de manière non sécurisée
Les solutions ne fonctionnent peut-être pas pour tout le monde	Une domotique assistant ne fournit aucune sortie audio pour visuellement déficient utilisateurs
Les utilisateurs doivent faire confiance à un système complexe	Un outil financier basé sur l'IA fait des recommandations d'investissement : sur quoi sont-elles basées ?
Qui est responsable pour l'IA pilotée décisions ?	Un innocent personne est condamné d' un crime basé sur preuve de la reconnaissance faciale – qui est responsable ?

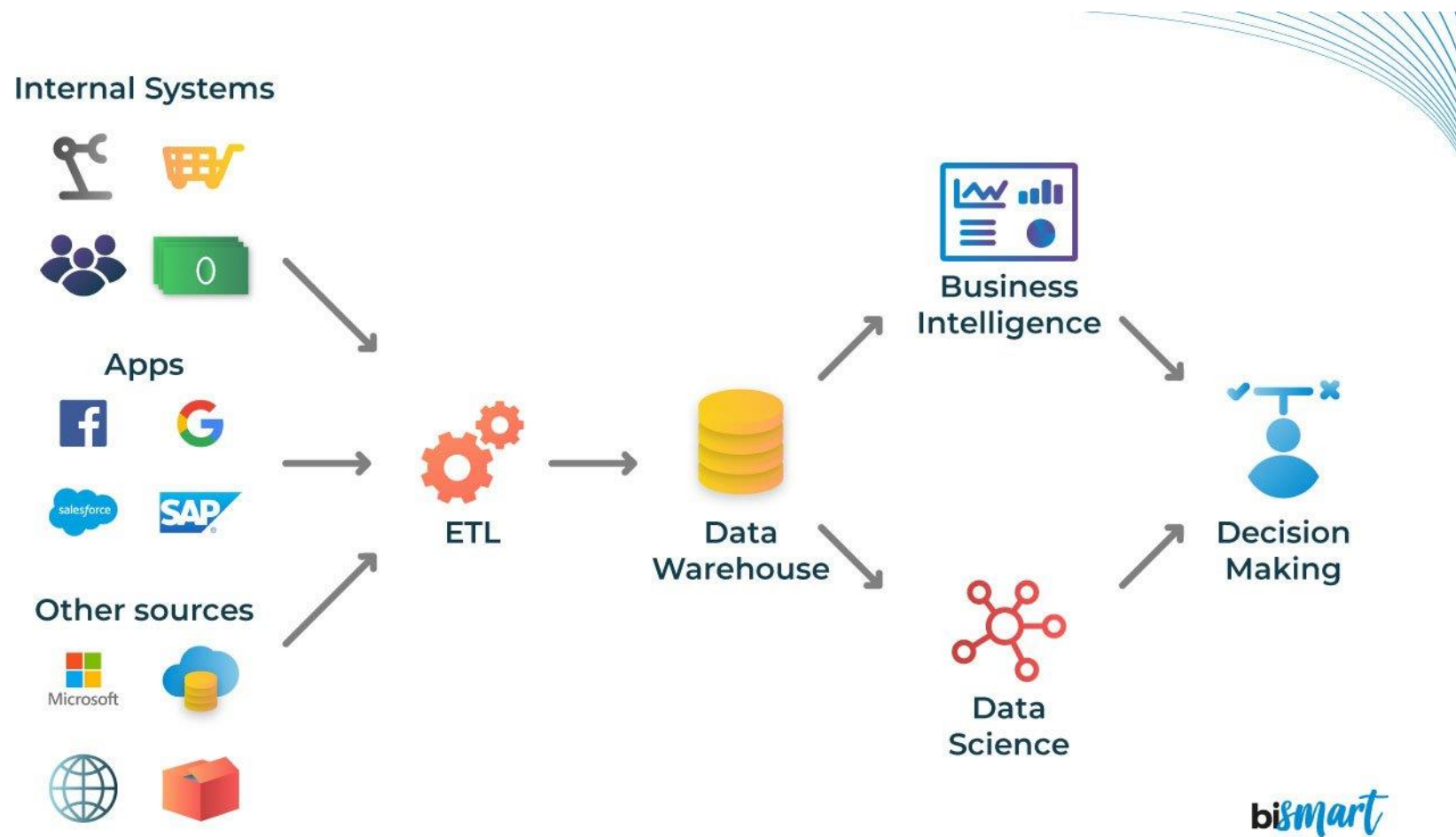
IA responsable



3. Conservation des données

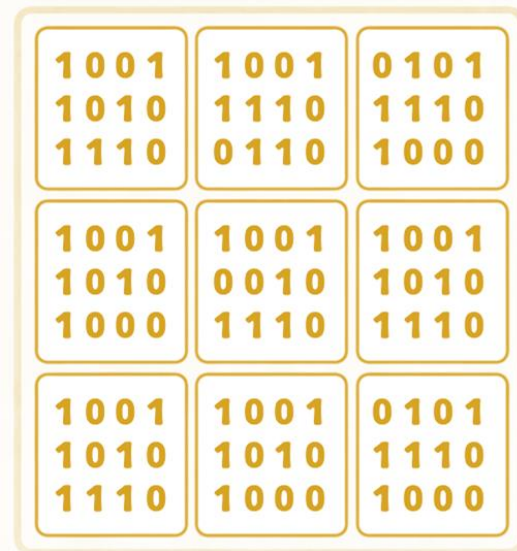
- Cette étape se concentre sur la collecte, le nettoyage, l'organisation et la préparation des **données** à utiliser dans le projet.
- Cela implique :
 - identifier les sources de données pertinentes
 - suppression des doublons ou des données incorrectes
 - formater les données pour les rendre utilisables.

Les données font référence à des faits bruts, des chiffres ou des éléments d'information qui sont collectés et enregistrés à des fins d'analyse, de traitement ou de stockage.



Types de données

Données structurées



informations hautement organisées, facilement consultables et généralement stockées dans des bases de données relationnelles ou des feuilles de calcul

Données semi-structurées



Bien qu'il ne réside pas dans une base de données relationnelle, il contient des balises ou d'autres marqueurs pour séparer les éléments sémantiques et appliquer des hiérarchies d'enregistrements et de champs dans les données

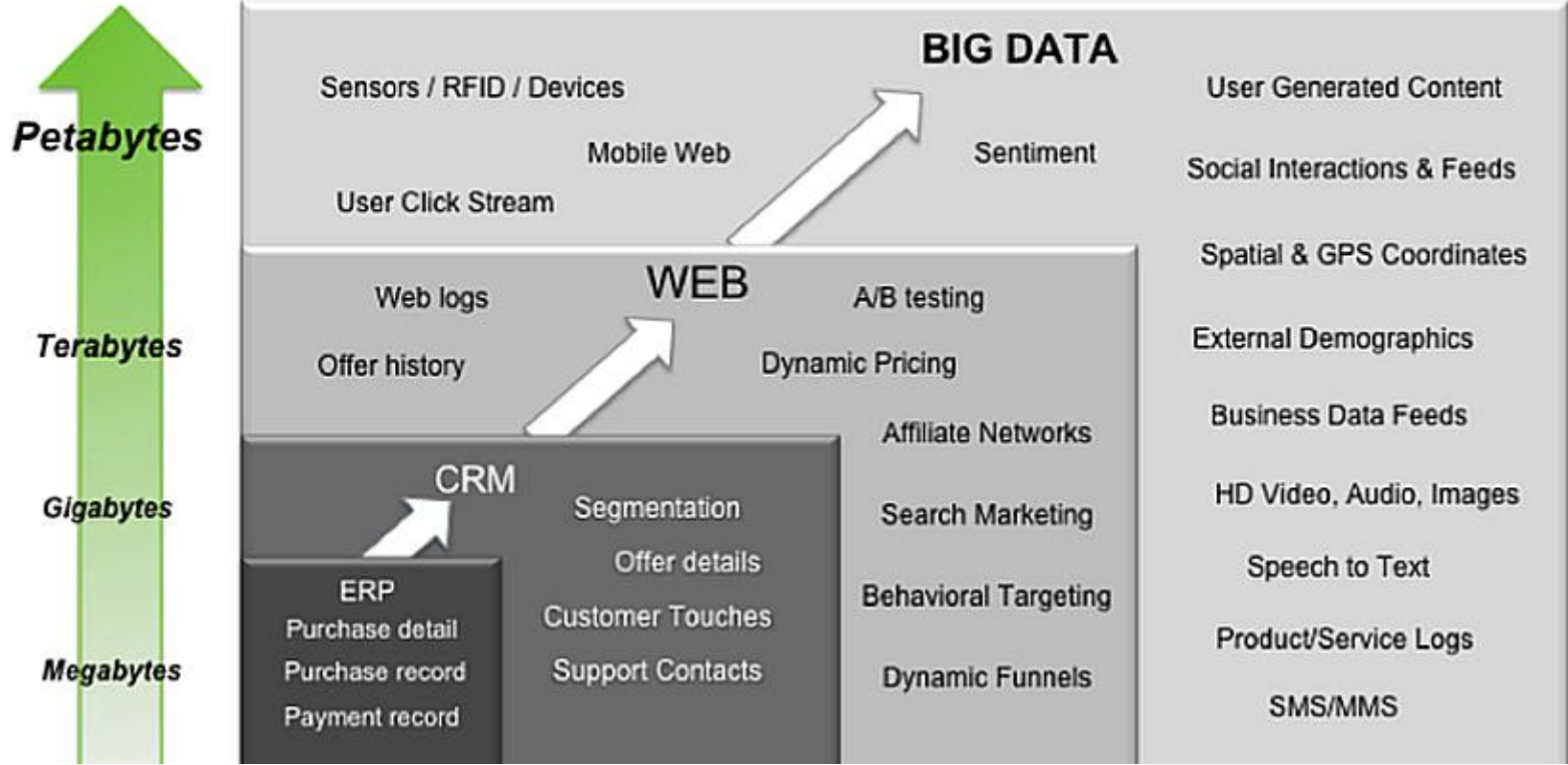
Données non structurées

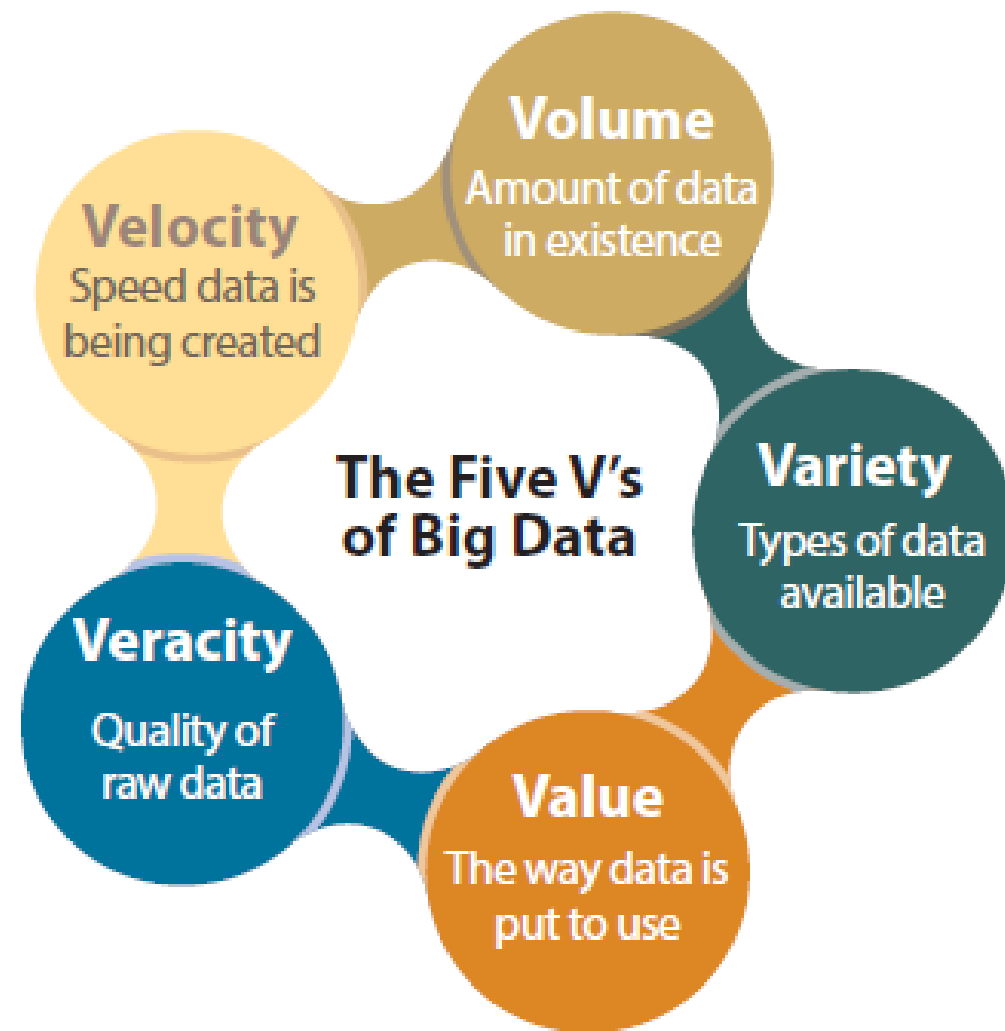


Les données non structurées manquent d'un modèle de données prédéfini, ce qui les rend plus difficiles à collecter, à traiter et à analyser

Big Data

Big Data = Transactions + Interactions + Observations





Nettoyage des données processus

- **Comprendre les données** : examiner la structure, les attributs clés et les métadonnées.
- **Gérer les données manquantes** : identifier, supprimer ou imputer les valeurs manquantes.
- **Supprimer les doublons** : détectez et supprimez les enregistrements en double.
- **Résoudre les incohérences** : normaliser les formats, unifier la casse du texte et corriger les fautes de frappe.
- **Gérer les valeurs aberrantes** : détecter, analyser et décider de conserver, de transformer ou de supprimer.
- **Valider les types de données** : garantir les types de données corrects et analyser les dates.
- **Corriger les données non valides** : identifier et corriger les valeurs erronées ou hors limites.
- **Normaliser et transformer** : mettre à l'échelle les données numériques et encoder les variables catégorielles.
- **Supprimer les fonctionnalités non pertinentes** : supprimez les colonnes redondantes ou non pertinentes.

Analyse exploratoire des données

- L'analyse exploratoire des données ou (EDA) consiste à comprendre les ensembles de données en résumant leurs principales caractéristiques, souvent en les traçant visuellement.
- 1) EDA Niveau 0 — Compréhension pure des données originales
 - 2) EDA Niveau 1 — Transformation des données originales
 - 3) EDA Niveau 2 — Compréhension des données transformées

Modélisation

- Au cours de cette phase, un modèle ou une solution initiale est développé pour tester la faisabilité du projet.
- Cela peut inclure la formation de modèles d'IA préliminaires et la réalisation de tests pour valider leurs performances.

Créer modèle

Modèle

Test

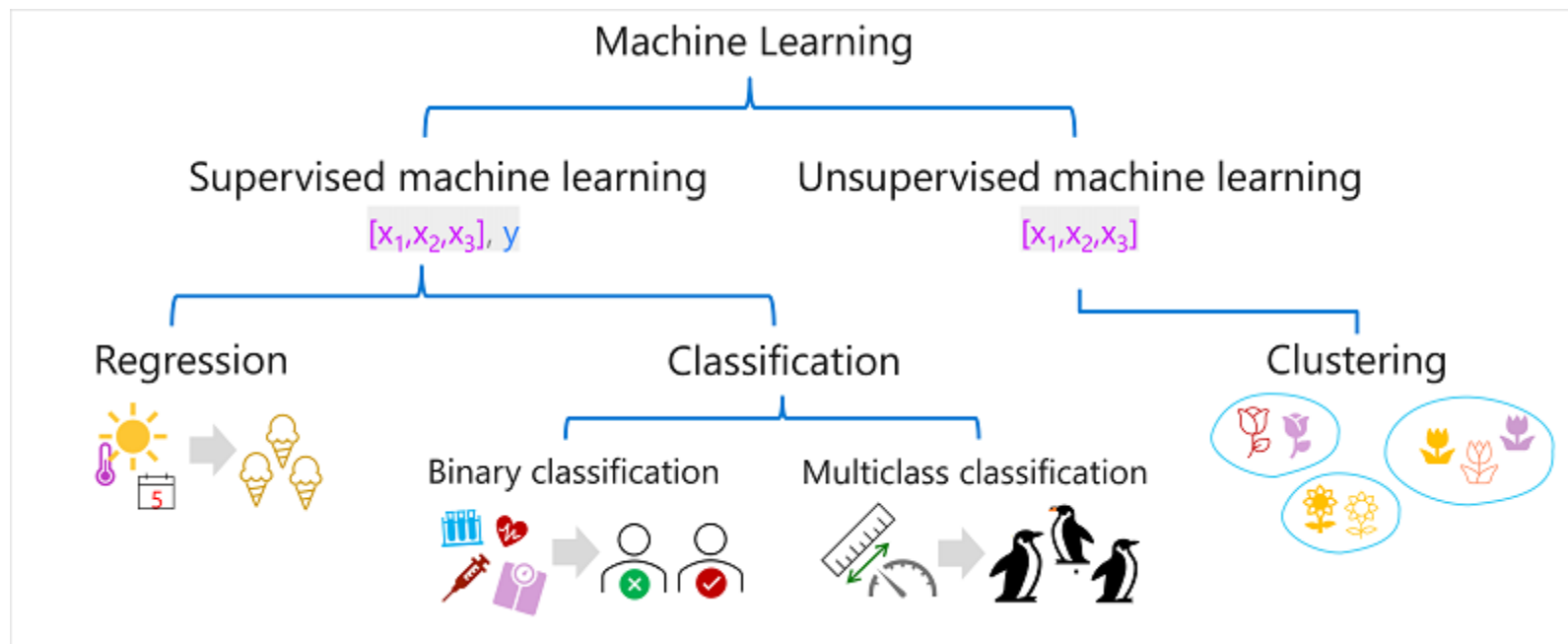
Améliorer

INTELLIGENCE ARTIFICIELLE (IA)

L'IA est un logiciel qui imite les comportements et les capacités humaines .
Principales charges de travail inclure

Apprentissage automatique –

Ceci est souvent le fondation pour un système d'IA , et est le chemin nous « enseignons » un ordinateur modèle à faire pronostics et tirage au sort conclusions à partir des données.



apprentissage supervisé

étant donné un ensemble de données de paires d'entrées-sorties, apprendre une fonction pour mapper les entrées aux sorties



Date	Humidité (humidité relative)	Pression (niveau de la mer, mb)	Pluie

Date	Humidité (humidité relative)	Pression (niveau de la mer, mb)	Pluie
1er janvier	93 %	999,7	Pluie
2 janvier	49 %	1015,5	Pas de pluie
3 janvier	79 %	1031,1	Pas de pluie
4 janvier	65 %	984,9	Pluie
5 janvier	90 %	975,2	Pluie

Date	Humidité (humidité relative)	Pression (niveau de la mer, mb)	Pluie
1er janvier	93 %	999,7	Pluie
2 janvier	49 %	1015,5	Pas de pluie
3 janvier	79 %	1031,1	Pas de pluie
4 janvier	65 %	984,9	Pluie
5 janvier	90 %	975,2	Pluie

x1

x2

y

f (humidité , pression)

f (93, 999,7) = Pluie

f (49, 1015,5) = Pas de pluie

f (79, 1031,1) = Pas de pluie

h(humidité, pression)

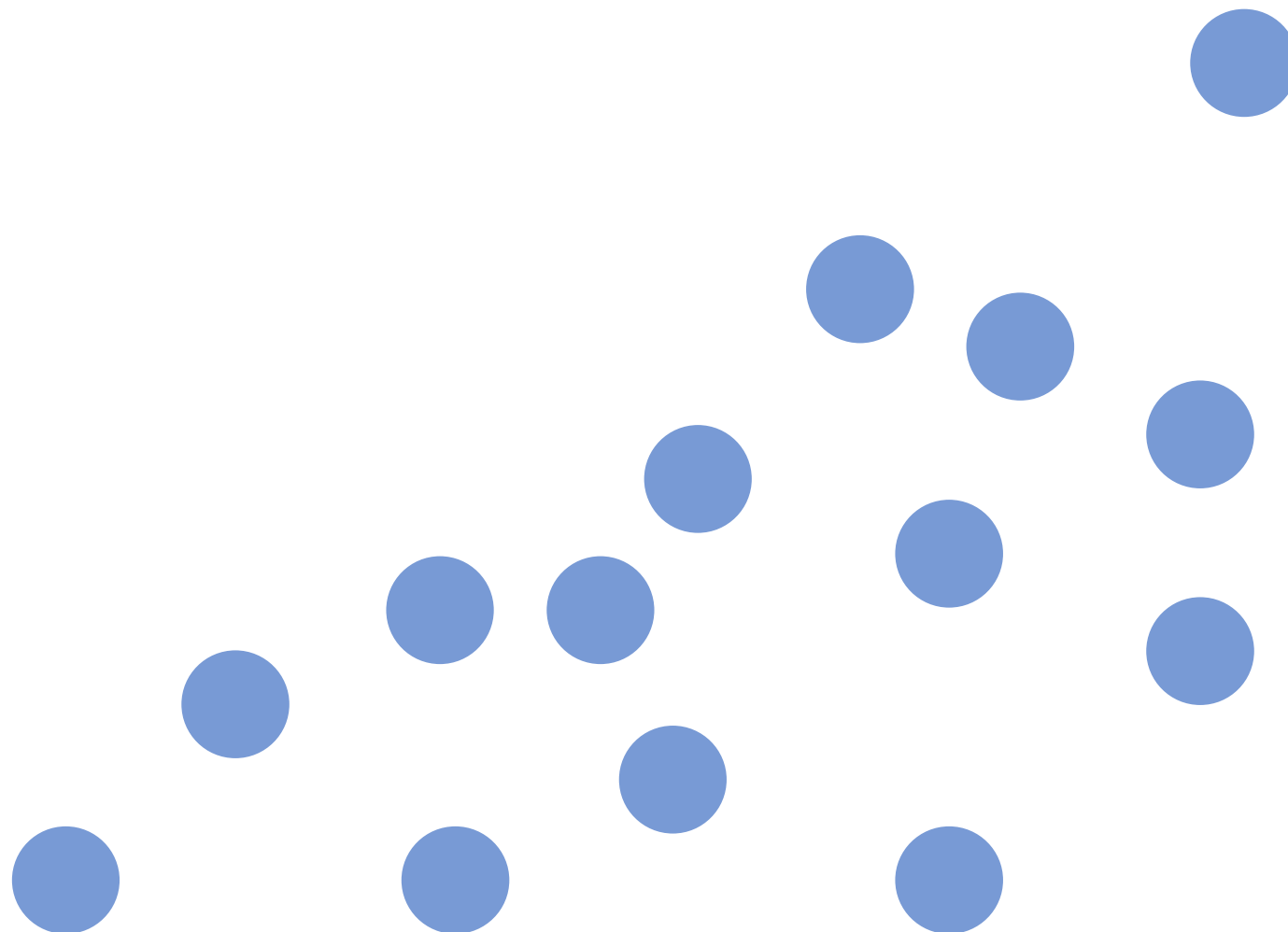
pression

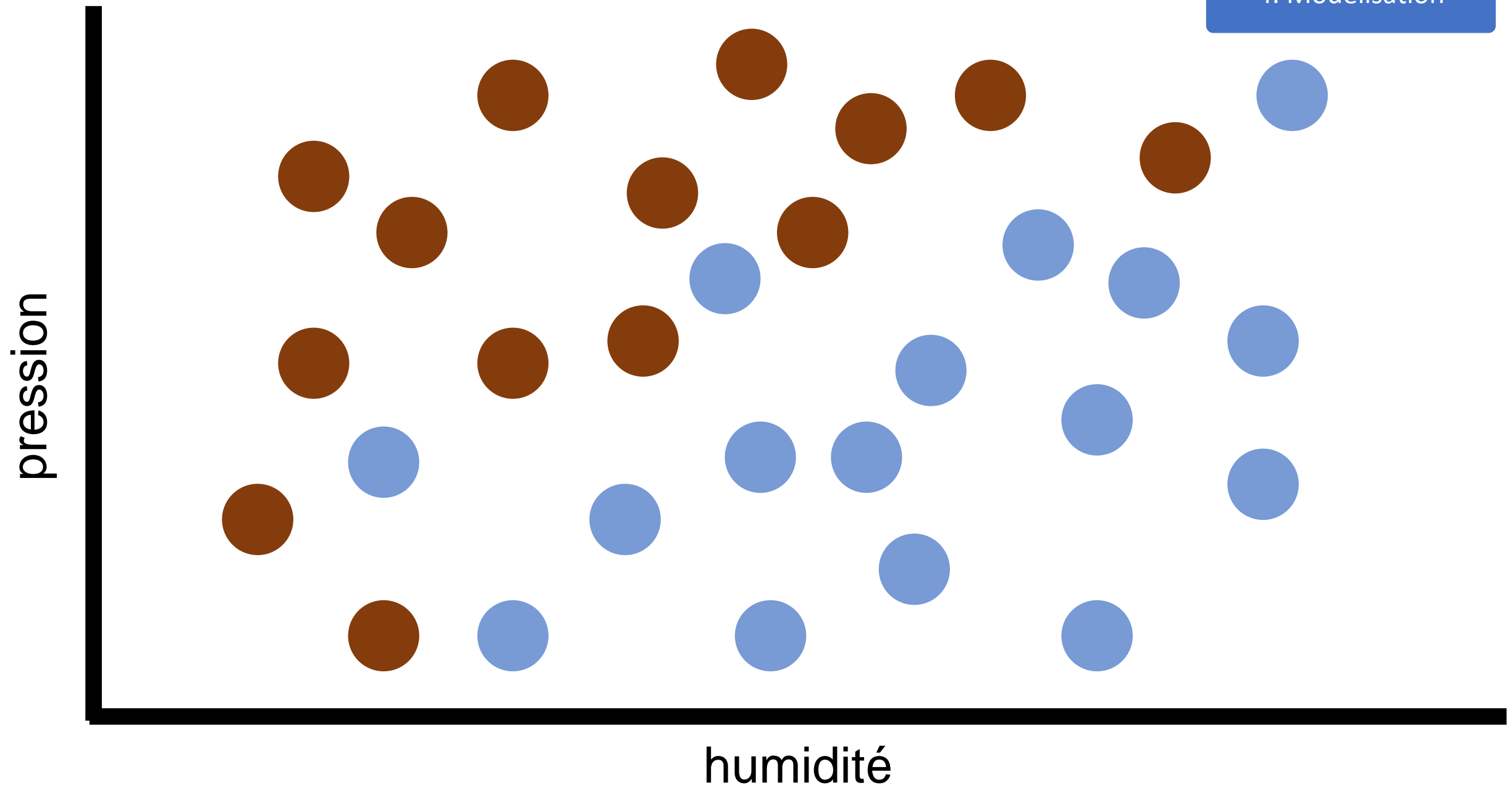
humidité

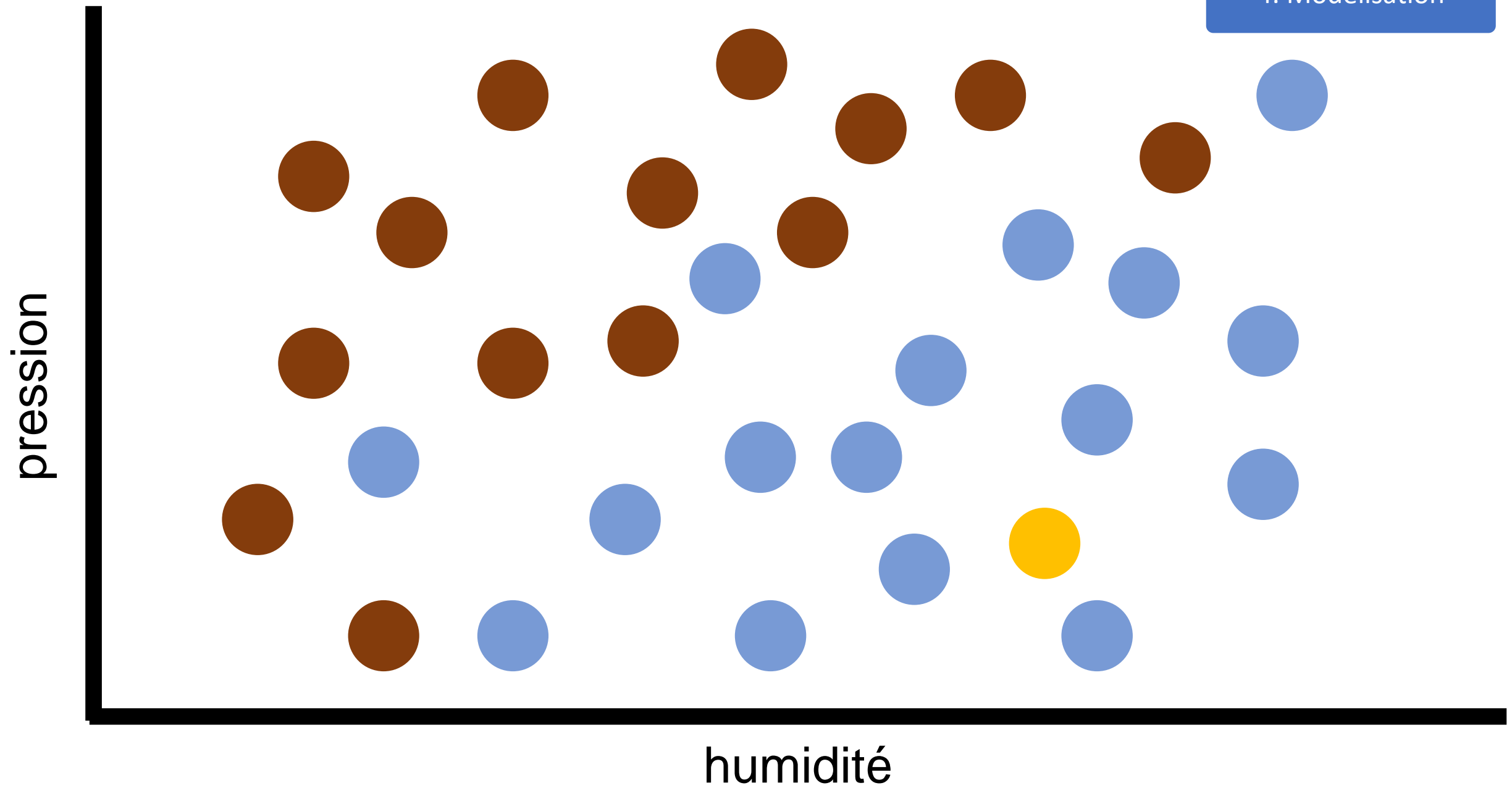
pression

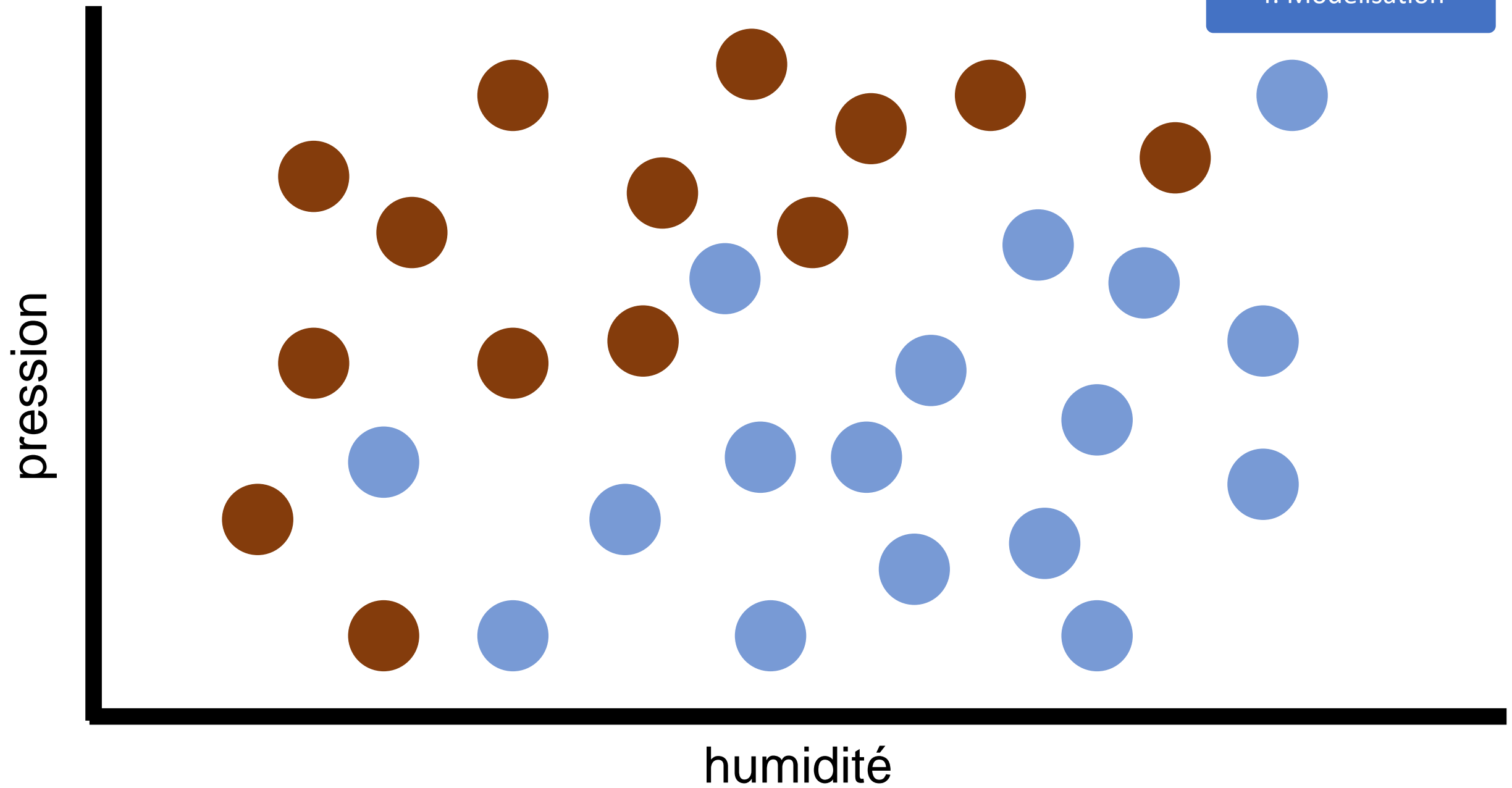
humidité

4. Modélisation









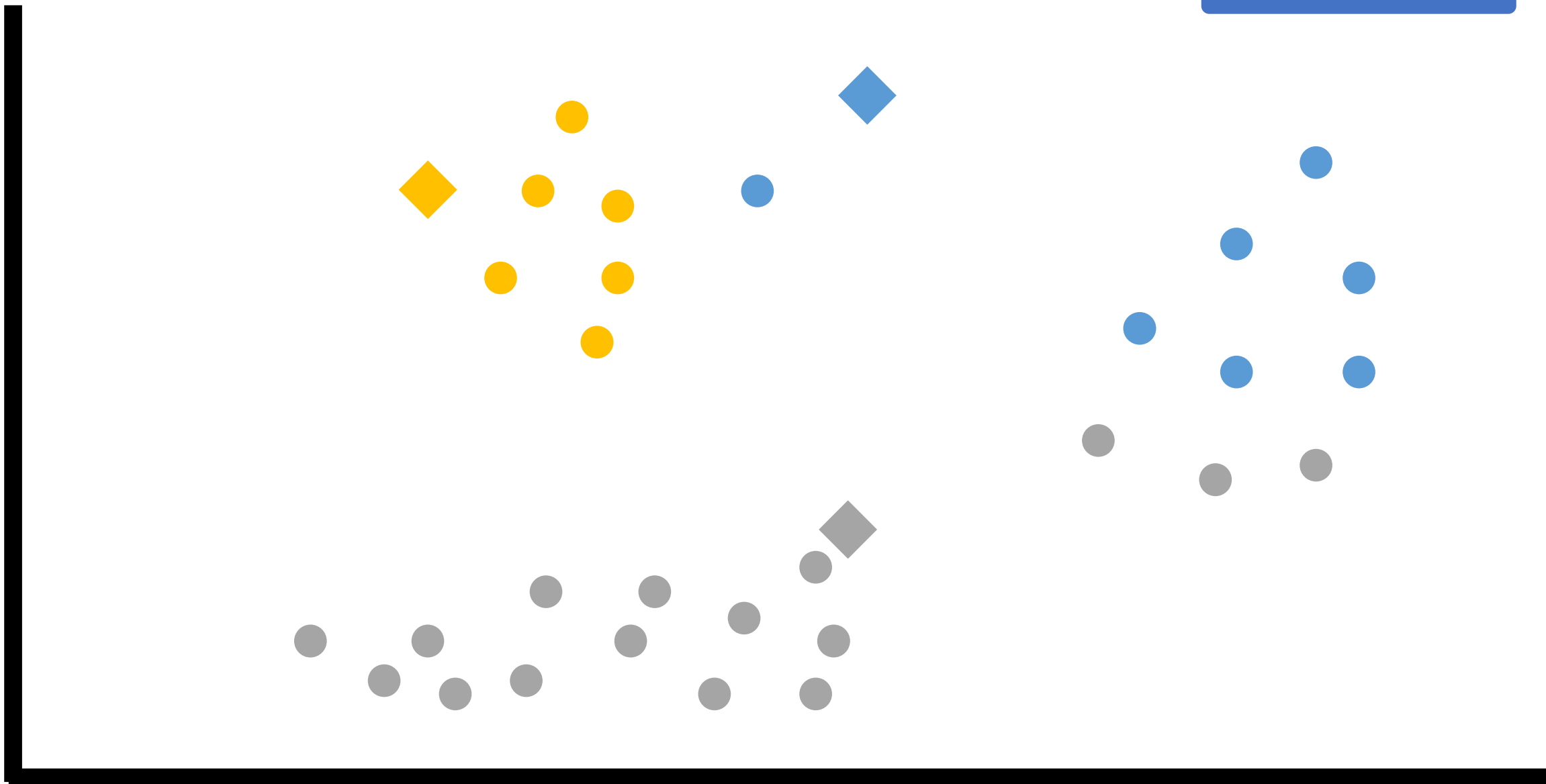
unsupervised learning

à partir de données d'entrée sans aucun retour supplémentaire, apprenez les modèles

clustering

organiser un ensemble d'objets en groupes de telle manière que les objets similaires tendent à se trouver dans le même groupe

4. Modelling



4. Modelling



Production

- Au cours de cette étape, la solution finale est déployée dans un environnement réel. Cela comprend l'optimisation du modèle, l'intégration du système dans l'infrastructure existante et la surveillance de ses performances en temps réel.