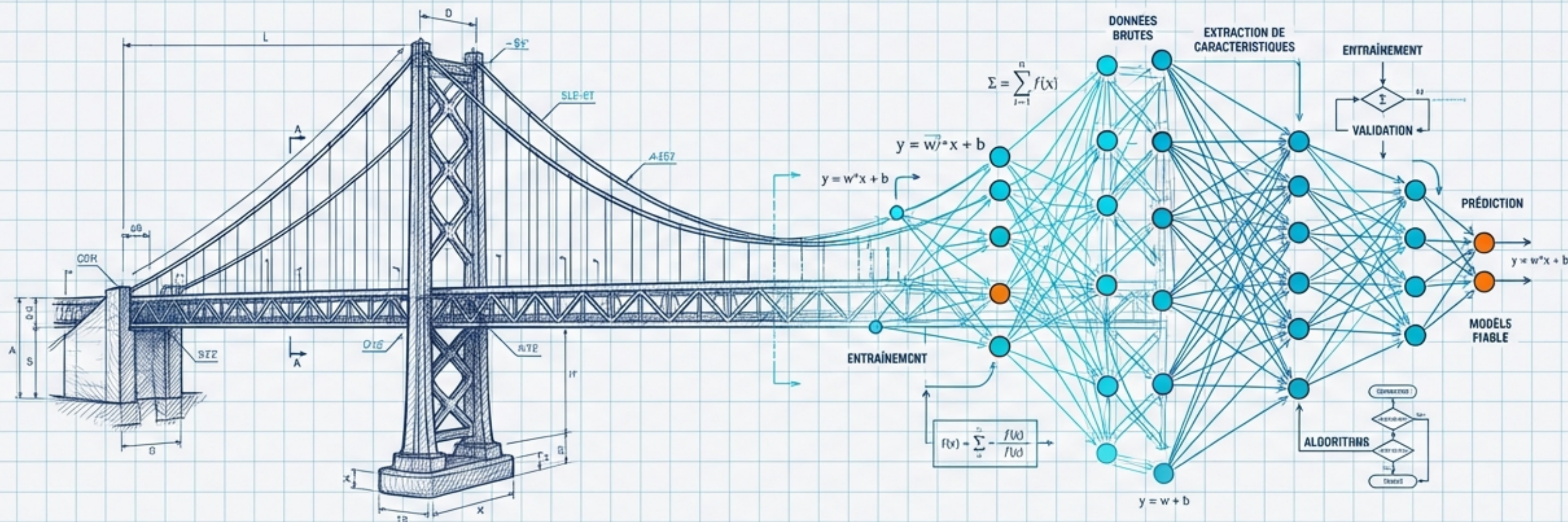


Dr. Lecheheb M



✓			
✓			
✓			
✓			
RE	REVIEWS HISTORY	100	100

**ACQUÉRIR UNE COMPRÉHENSION PROFONDE DU MACHINE LEARNING POUR
TIRER PROFIT DES DONNÉES EXPÉRIMENTALES ET CONSTRUIRE DES MODÈLES
FIABLES, AVANT LA MISE EN ŒUVRE PRATIQUE EN PYTHON.**

PROJECT	D&B CIVIL SOLUTIONS	
REFERENCE DMS	BRI 169CRIST	
LOCATION/DESCRIPTION	BARONE	DESIGNER FIRM
MANAGEMENT	FORUM	

QUAND LES MODÈLES ANALYTIQUES ATTEIGNENT LEURS LIMITES

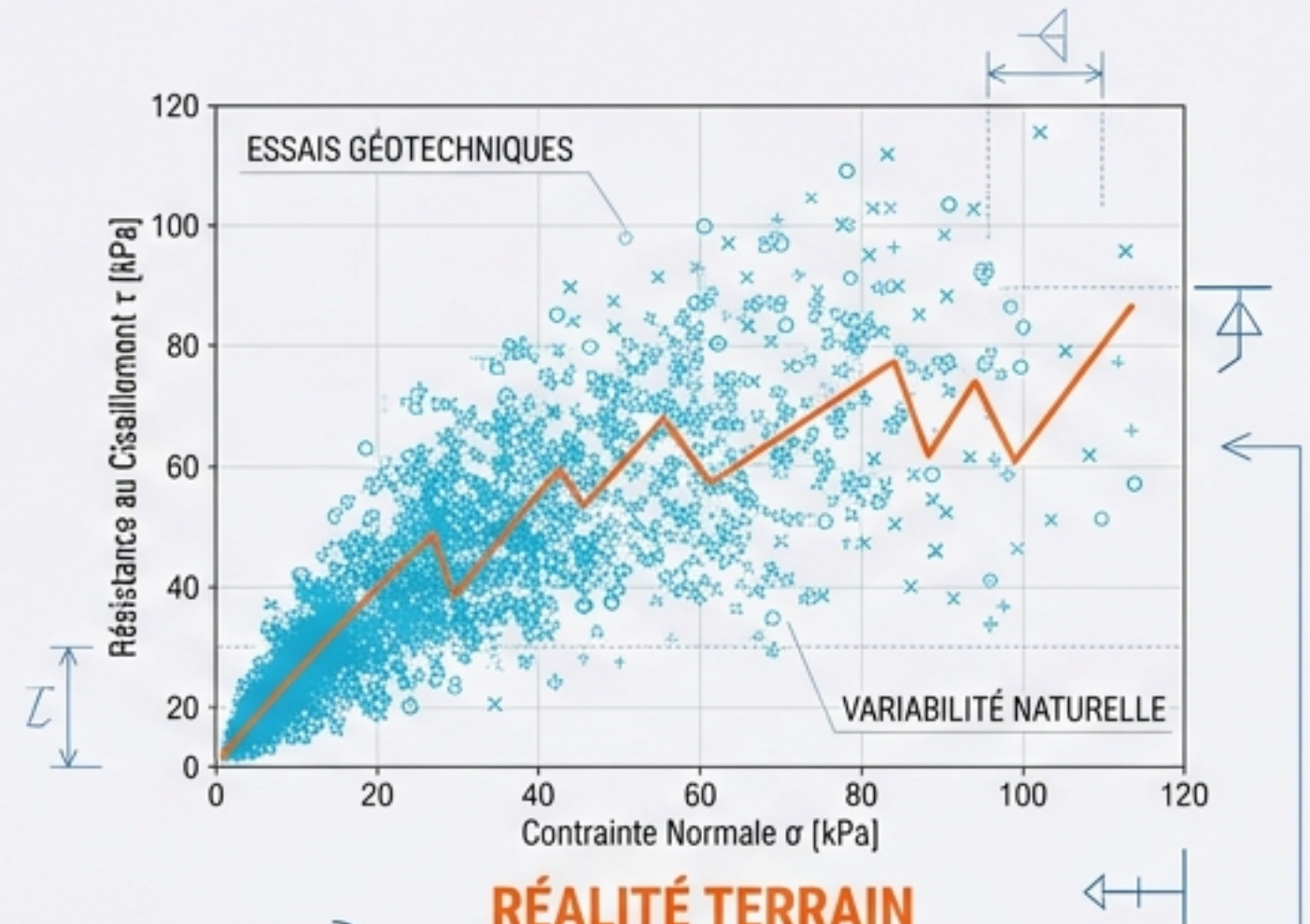
$$F = m \cdot a$$

MODÈLE ANALYTIQUE
(THÉORIE IDÉALE)

En génie civil, de nombreux problèmes ne disposent pas de modèles analytiques simples ou nécessitent des hypothèses trop lourdes pour être précis.

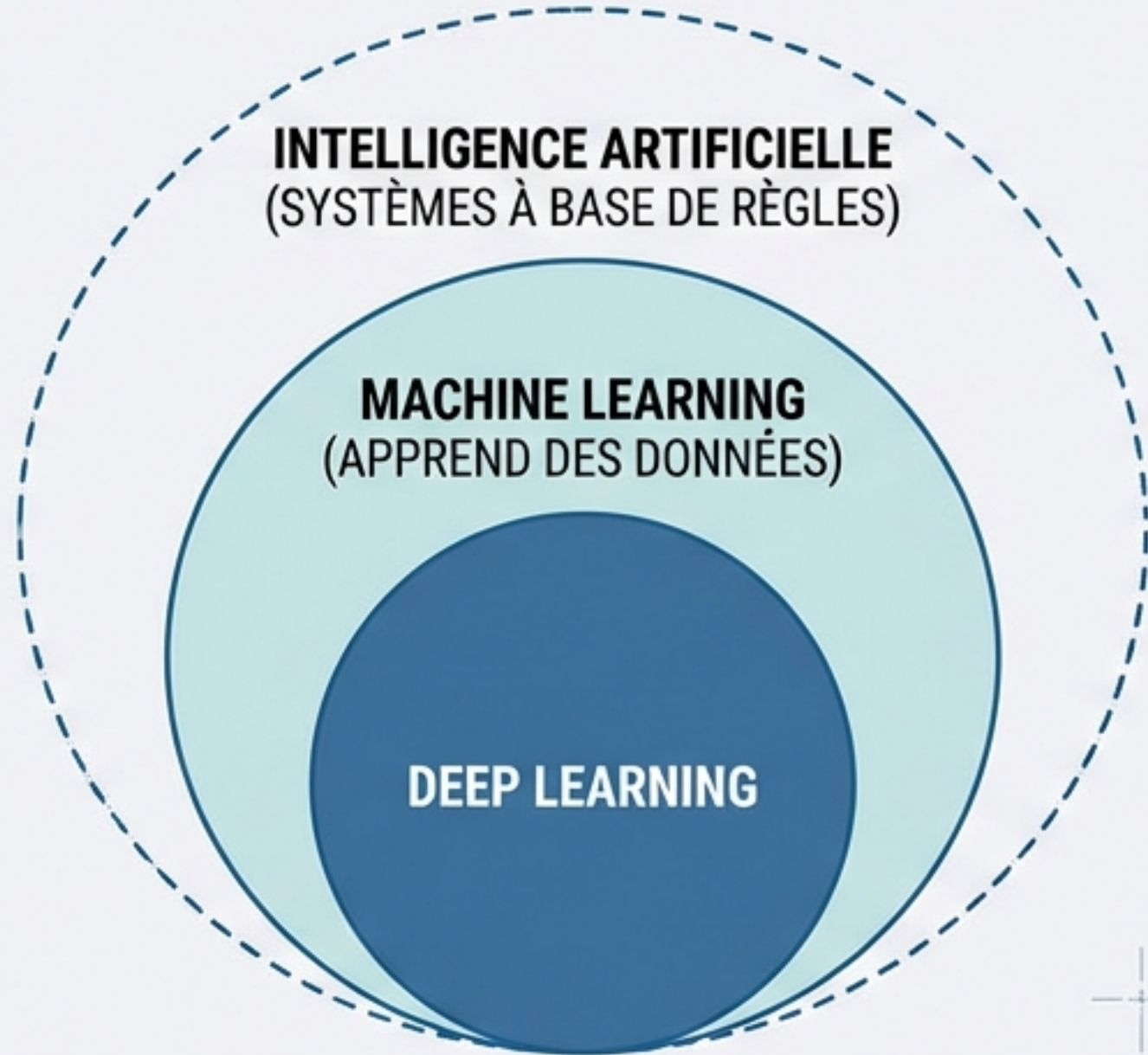
LA SOLUTION ML :

Le Machine Learning permet d'exploiter les données de terrain et expérimentales pour construire des modèles prédictifs là où la théorie pure ne suffit pas.



RÉALITÉ TERRAIN
(DONNÉES COMPLEXES)

POSITIONNEMENT : TRANSFORMER LA DONNÉE EN CONNAISSANCE



LE MACHINE LEARNING SE DISTINGUE PAR SA CAPACITÉ À :

- Apprendre automatiquement à partir de données.
- S'adapter à de nouvelles situations sans être explicitement programmé.
- Améliorer ses performances avec l'expérience.

IDÉE CLÉ : Le ML n'est pas de la magie, c'est un outil statistique qui transforme les archives en outils de décision.

LES MATÉRIAUX DE CONSTRUCTION : NATURE DES DONNÉES



BASE DE DONNÉES CENTRALE
INVENTAIRE DES MATÉRIAUX



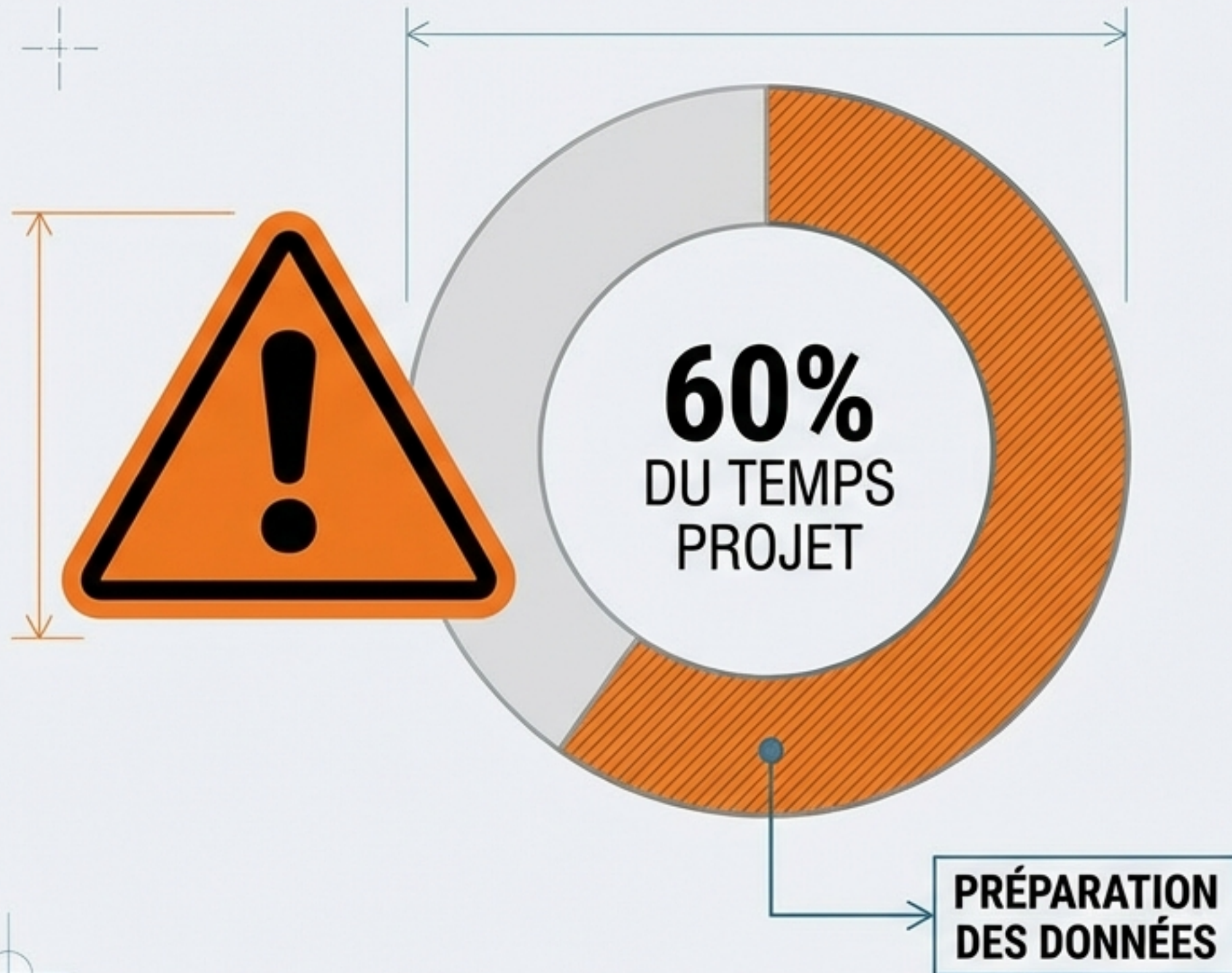
DONNÉES STRUCTURÉES (FOCUS DU COURS) :
Tableaux numériques, Excel, bases de données.
Accessibles et prêtes à l'emploi.



DONNÉES NON-STRUCTURÉES :
Images, rapports PDF, vidéos.
Nécessitent un traitement complexe.



CONTRÔLE QUALITÉ : GARBAGE IN, GARBAGE OUT



Un modèle de Machine Learning est aussi bon que les données utilisées pour l'entraîner.

DÉFAUTS MAJEURS À SURVEILLER :

- Données manquantes (trous dans l'historique).
- Erreurs de mesure (capteurs défectueux).
- Valeurs aberrantes (outliers).
- Données non représentatives du problème physique.

LE PLAN TECHNIQUE : VOCABULAIRE ESSENTIEL

FEATURES (x) - LES ENTRÉES

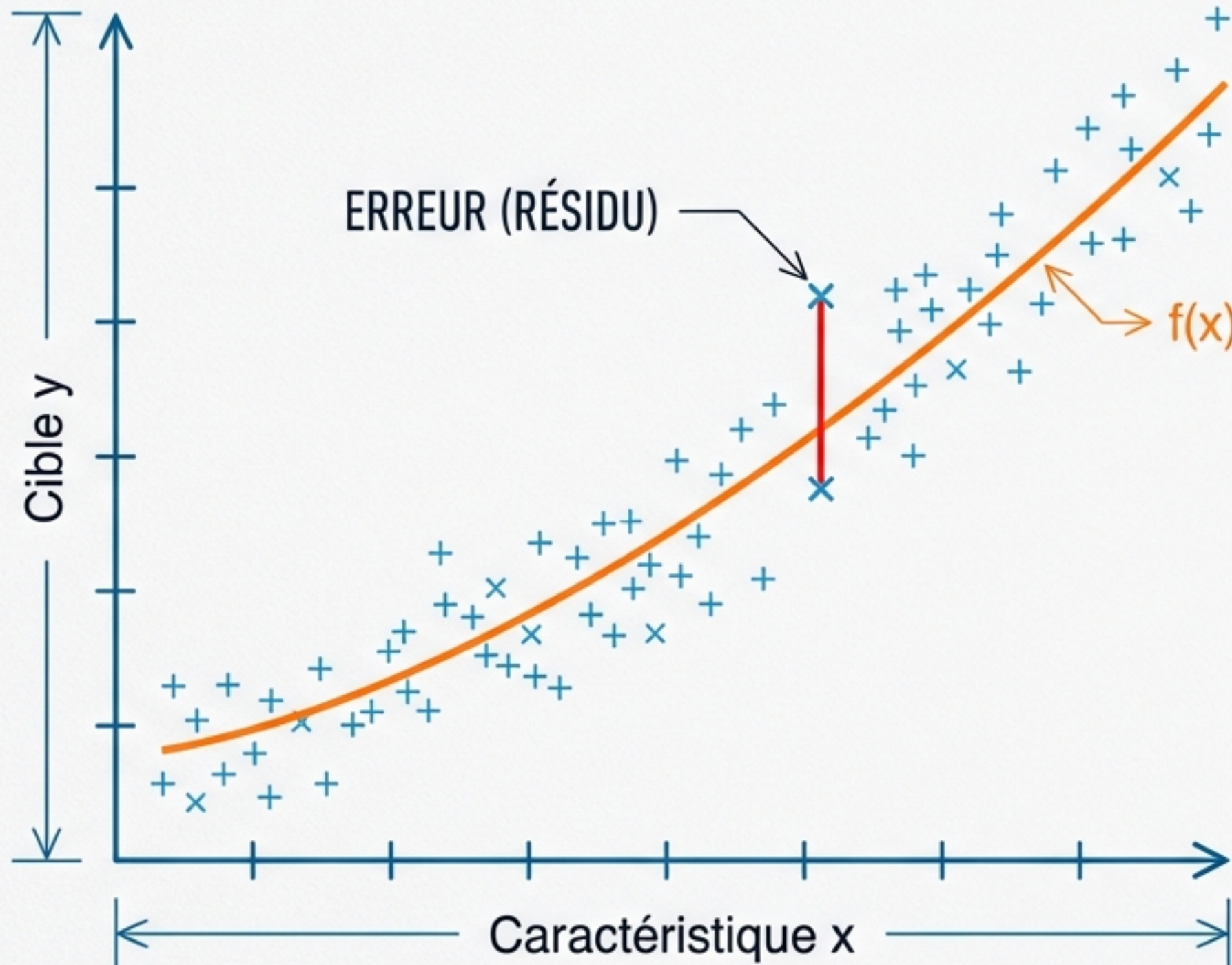
LABEL (y) - LA CIBLE À PRÉDIRE

DATASET
(JEU DE
DONNÉES) :
Matrice
d'observations

ID	MASSE VOLUMIQUE (kg/m ³)	ÂGE (Jours)	RÉSISTANCE (MPa)
C-101	2300	28	35.5
C-102	2320	21	31.2
C-103	2290	14	25.0
C-104	2310	7	18.3
C-105	2305	28	36.0

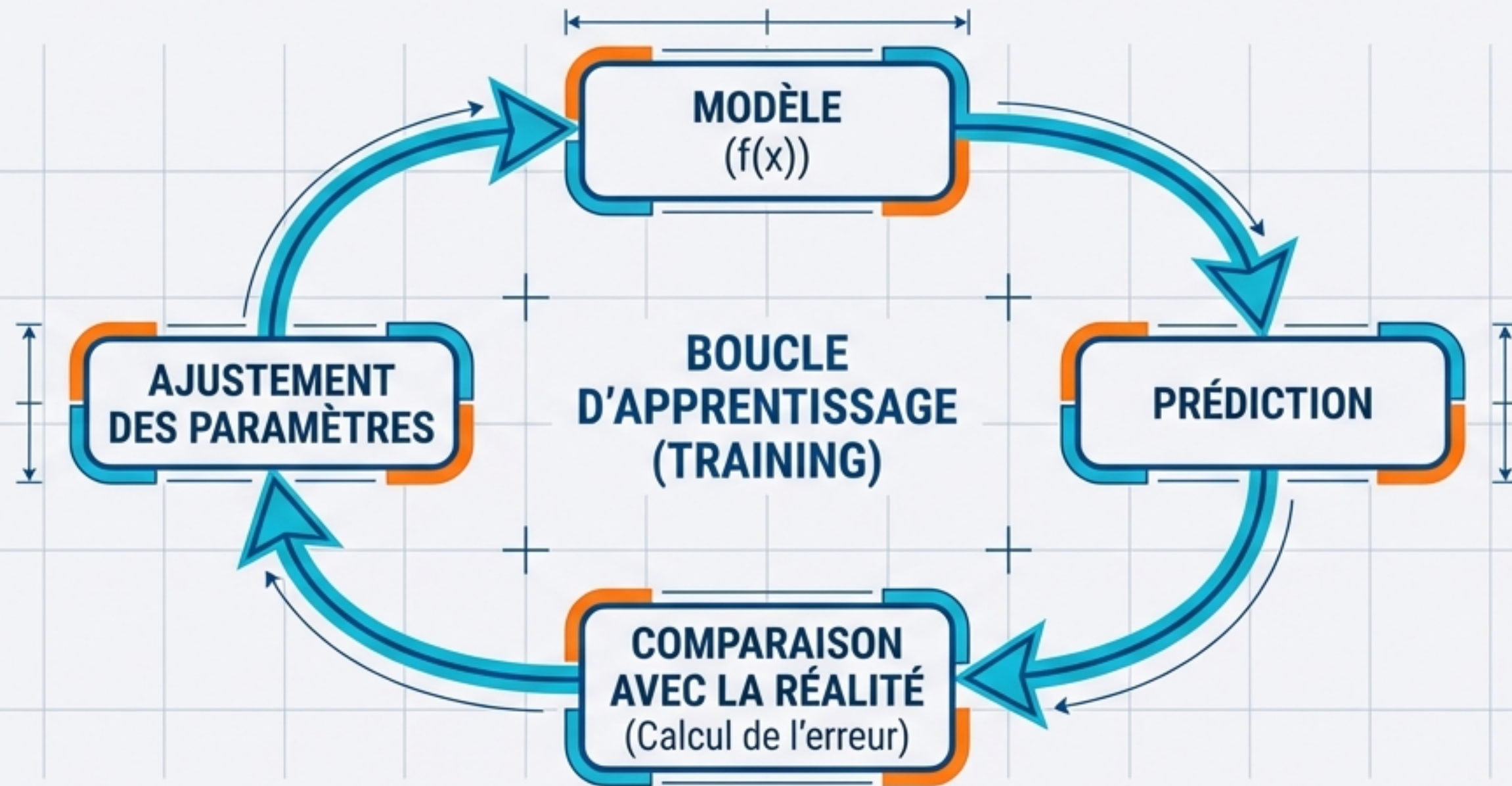
NOTE : Il est crucial de normaliser les grandeurs de nature différente pour éviter de biaiser le modèle.

LA STRUCTURE : QU'EST-CE QU'UN MODÈLE ?



- Un **modèle** est une fonction mathématique $y = f(x)$ qui approxime la relation entre les entrées et la sortie.
 - L'objectif est de trouver la fonction f qui minimise l'écart entre la prédiction et la réalité observée.
- ⚠ ATTENTION : Ce modèle n'est pas une vérité physique absolue. C'est une approximation issue des données.

MÉTHODE DE CONSTRUCTION : APPRENTISSAGE ET GÉNÉRALISATION



LE BUT ULTIME : LA GÉNÉRALISATION

La capacité du modèle à fonctionner sur des données NOUVELLES, qu'il n'a jamais vues auparavant.
Un bon modèle n'est ni trop simple, ni trop complexe.



En Python : `model.fit(X_train, y_train)`

LE PIPELINE (1/3) : PRÉPARATION DU CHANTIER

PRÉPARATION DU CHANTIER

CONSTRUCTION DU MODÈLE

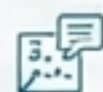
DÉPLOIEMENT ET SUIVI



ÉTAPE 1 : COMPRÉHENSION DU PROBLÈME



☒ Définir l'objectif et identifier la variable à prédire (y).



☒ Comprendre le contexte physique avant de lancer des calculs.



ÉTAPE 2 : PRÉPARATION DES DONNÉES



☒ Nettoyage (suppression des erreurs).

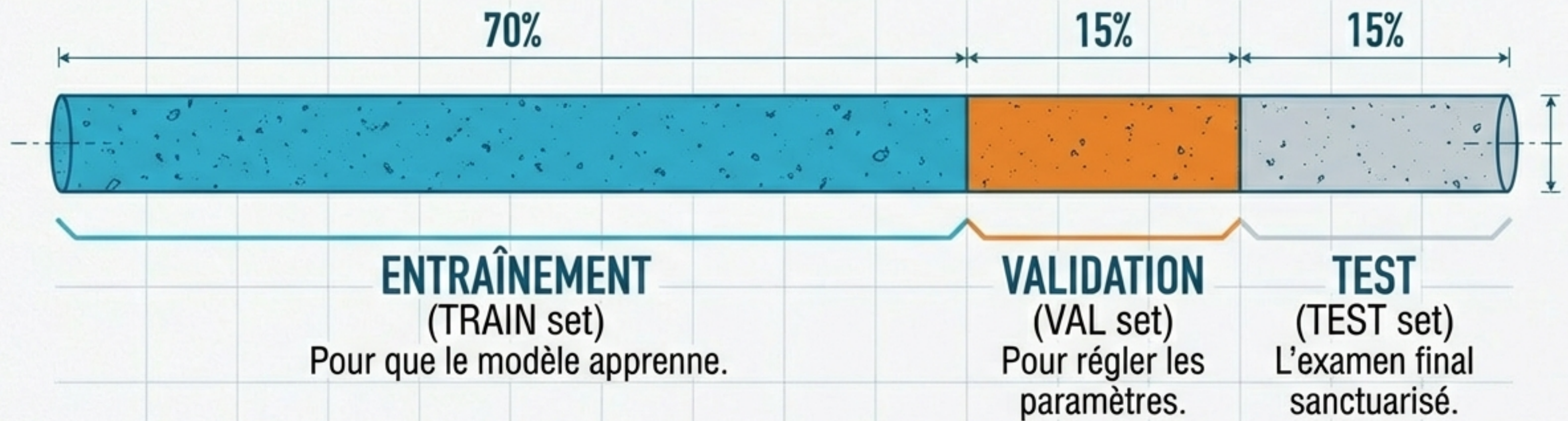


☒ Normalisation (mise à l'échelle).



☒ Sélection des variables pertinentes.

LE PIPELINE (2/3) : SÉPARATION DES ZONES



ÉTAPE 3 : SÉPARATION DES DONNÉES (SPLIT)

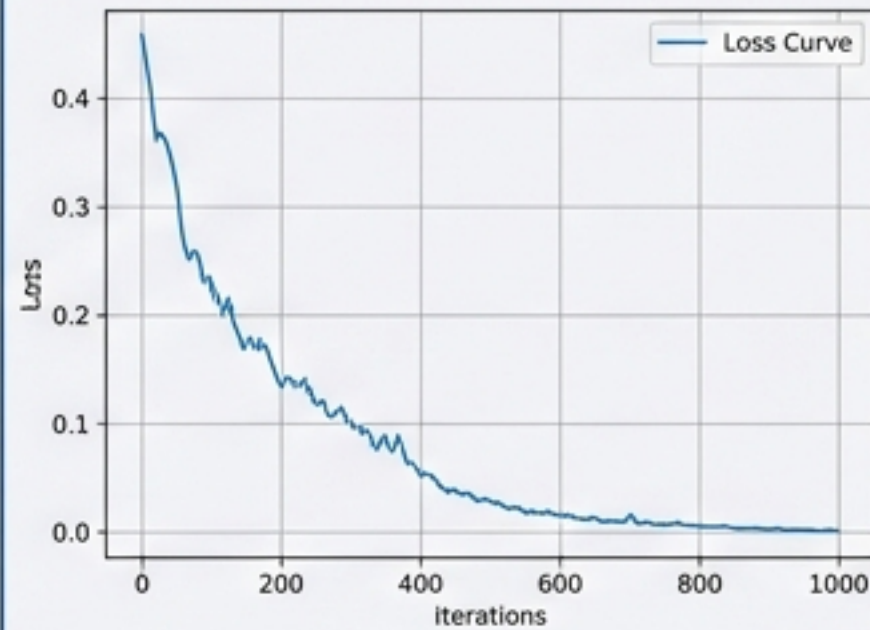


RÈGLE D'OR : Ne jamais tester le modèle sur les données qu'il a déjà vues. Ce serait comme donner les réponses avant l'examen.

LE PIPELINE (3/3) : EXÉCUTION ET INSPECTION



ÉTAPE 4 : ENTRAÎNEMENT



ÉTAPE 5 : ÉVALUATION



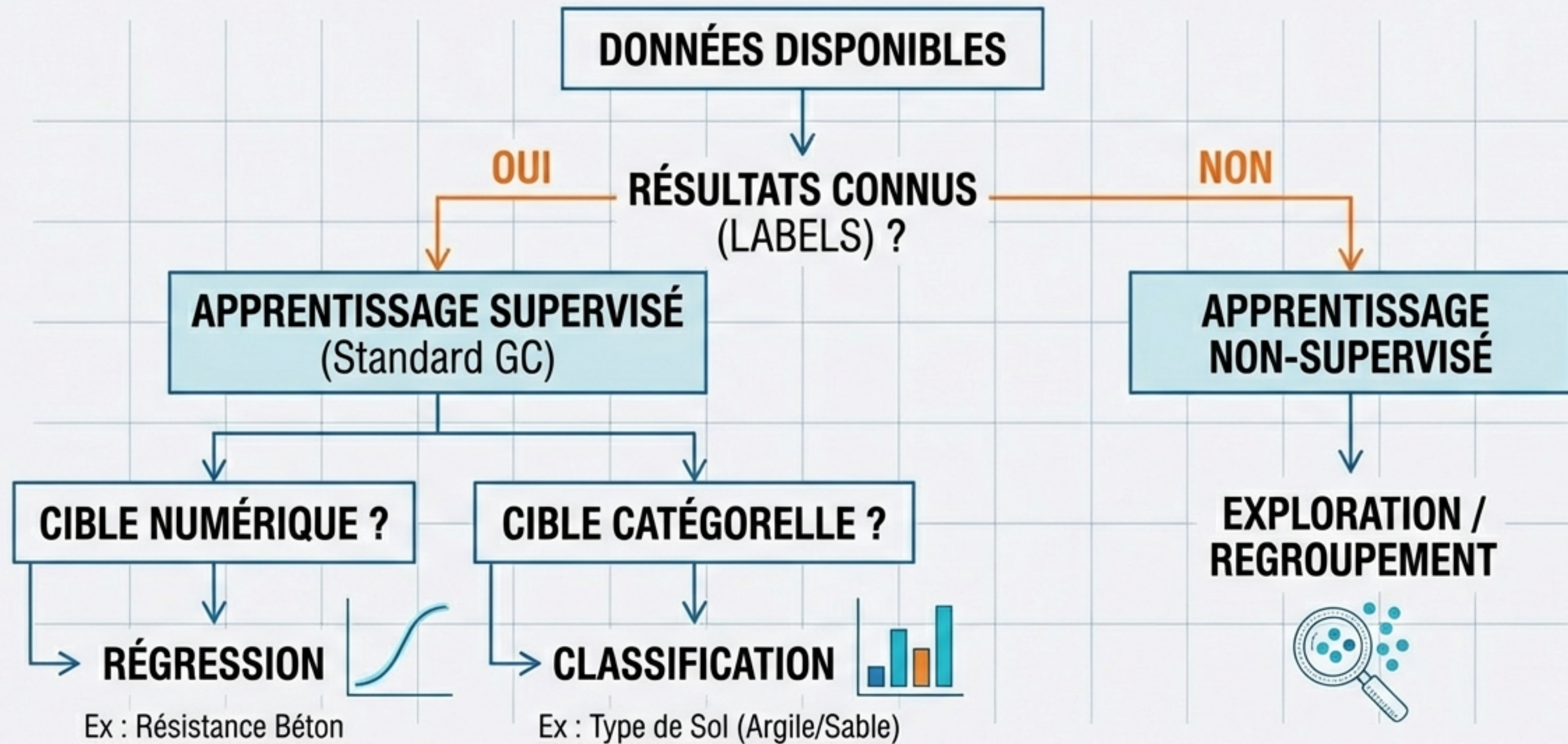
ÉTAPE 6 : PRÉDICTION ET INTERPRÉTATION



Le résultat mathématique doit toujours être **VALIDÉ PHYSIQUEMENT**.
Est-ce que la prédiction a du **sens pour un ingénieur** ?



LA BOÎTE À OUTILS : TYPES D'APPRENTISSAGE

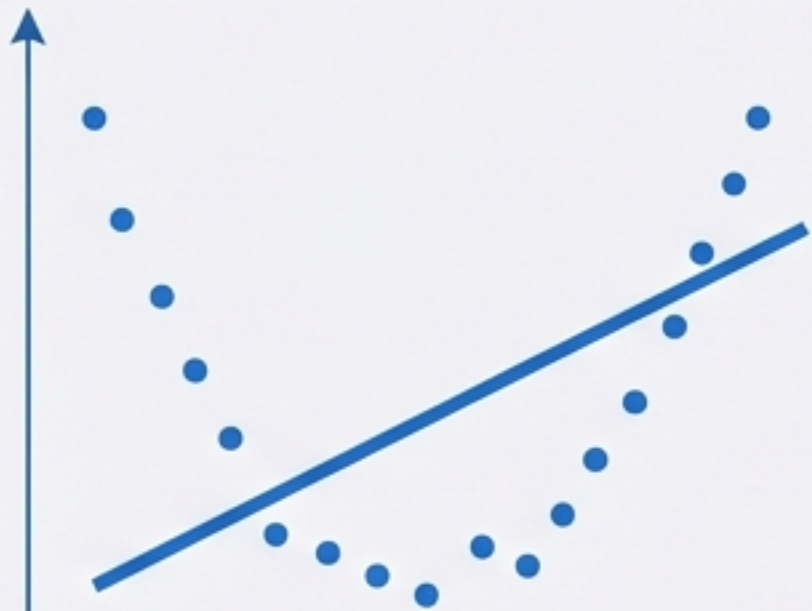




RISQUES STRUCTURELS : BIAIS ET VARIANCE



SOUS-APPRENTISSAGE (UNDERFITTING)



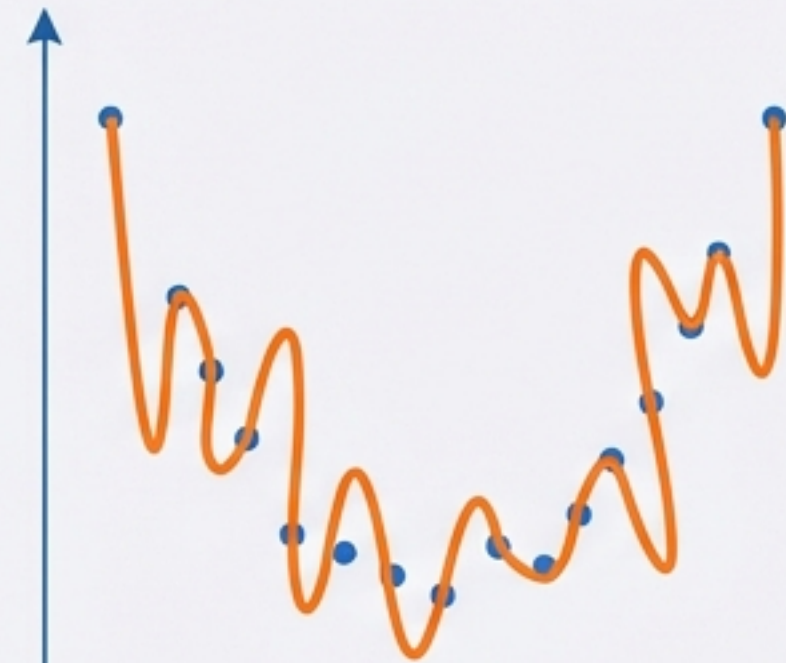
Trop simple / Biais élevé

BON ÉQUILIBRE



Généralisation optimale

SURAPPRENTISSAGE (OVERFITTING)



Trop complexe / Variance élevée

Conséquence du surapprentissage : Le modèle apprend le 'bruit' et échoue sur le terrain.



LE RÔLE DE L'INGÉNIEUR CIVIL

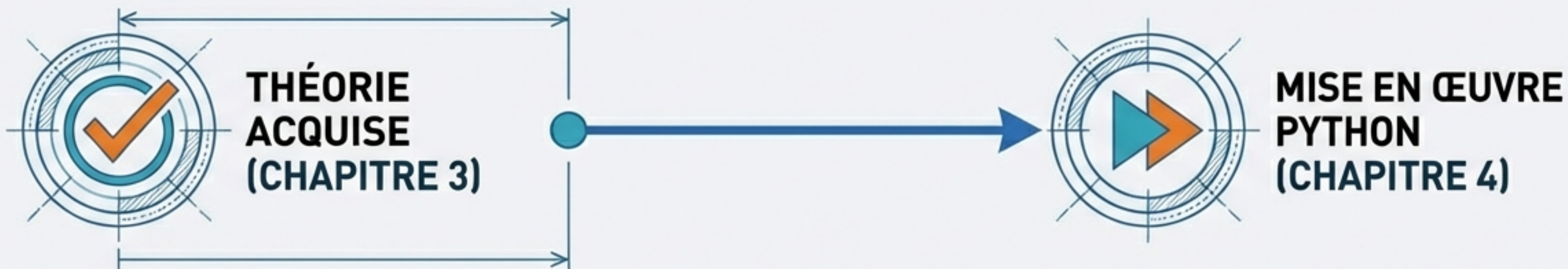


Le Machine Learning n'est pas autonome et ne remplace pas l'ingénieur.

VOS RESPONSABILITÉS :

1. **CHOIX DES DONNÉES** : Sélectionner les variables pertinentes.
2. **VALIDATION** : Juger si le modèle respecte la physique.
3. **DÉCISION** : L'algorithme propose, l'ingénieur dispose.

CONCLUSION ET PROCHAINES ÉTAPES



CE QU'IL FAUT RETENIR :

- Le ML est un outil puissant pour valoriser les données complexes.
- La qualité des données (60% du travail) prime sur la complexité de l'algorithme.
- L'ingénieur reste le garant de la validité physique.

**PROCHAINE ÉTAPE : MISE EN ŒUVRE PRATIQUE
DE CES CONCEPTS AVEC PYTHON.**