



# AI4Diet

Apport de l'IA dans les modèles d'optimisation des rations pour soutenir la transition agro-écologique des systèmes d'élevage dans différents contextes

**INRAE**

AgroParisTech 

## WP3:

# AI contributions to the prediction of feed values and animal performance

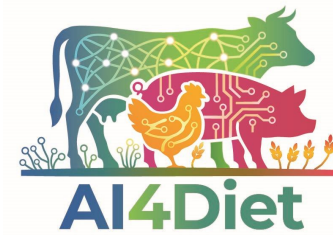
**Vincent Guigue**

Professeur d'informatique

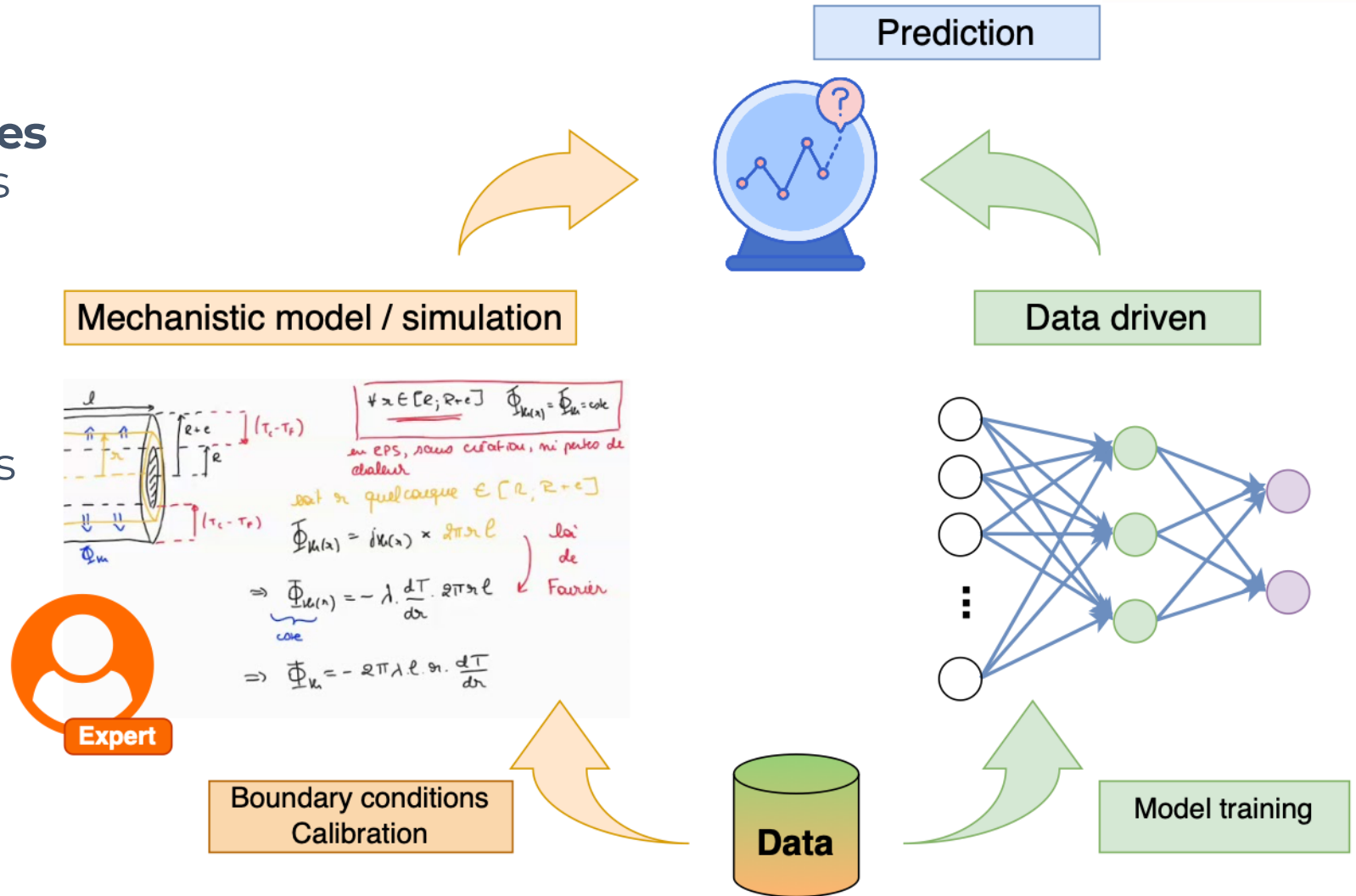


MIA  
PARIS-SACLAY  
EKINOCs

# Données & modélisation : différentes architectures



- Modélisation mécaniste
  - Traduire les **connaissances des experts** en équations
  - Données = calibration, conditions limites, ...
  
- IA / Science des données
  - Extraire les connaissances des **observations**
  - Données = sources des connaissances

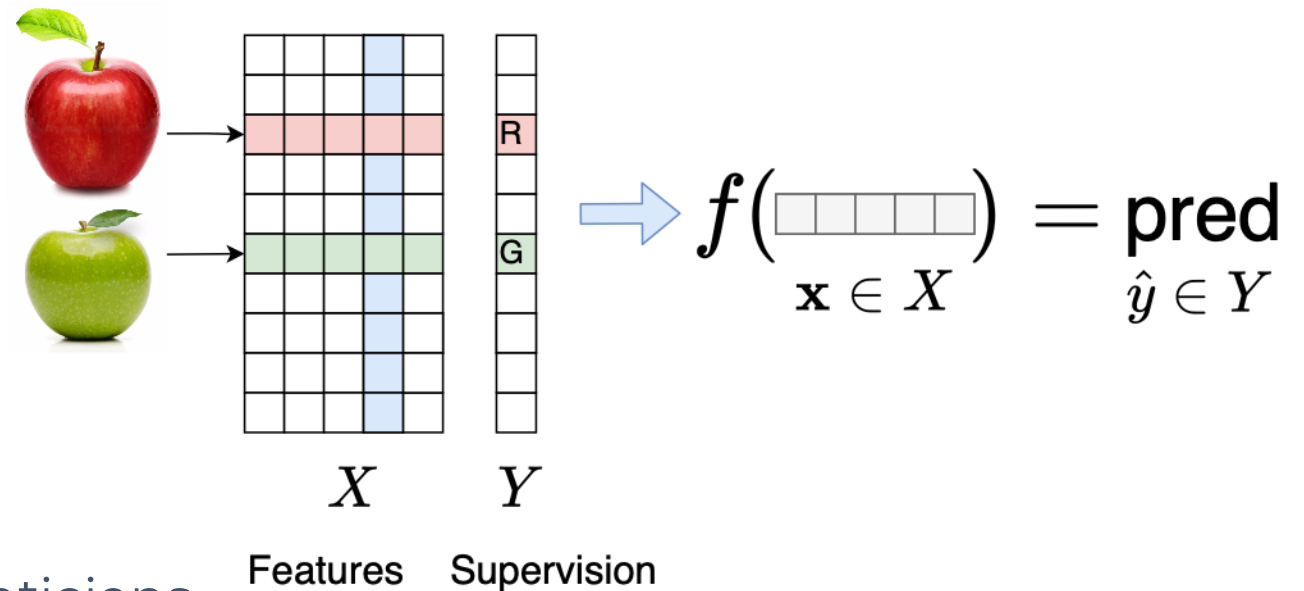


## □ Données

- Entrées (=observations)
- Sorties (=prédictions)

## □ Problématiques usuelles

- Supervision (données de sortie)
- Volume de données disponibles
- Périmètre des données
- Périmètre de validité du modèle
- Incompréhensions entre les informaticiens et le terrain



Besoin d'expertise limité

Librairies matures et faciles d'accès

Collecte des données de plus en plus facile

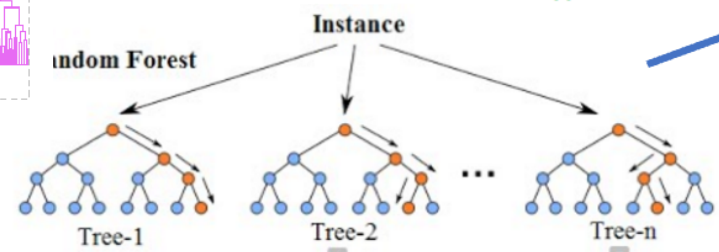
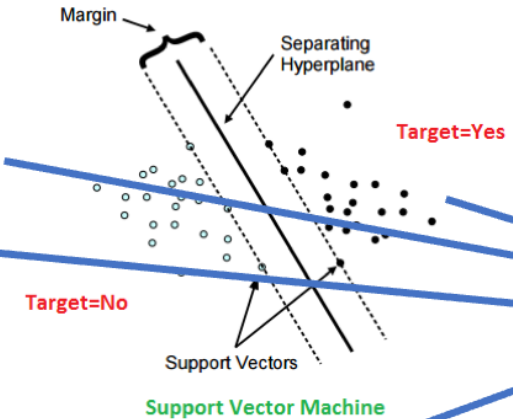
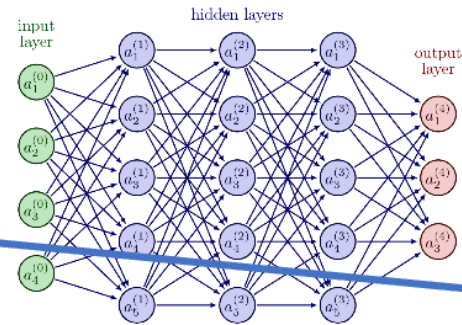
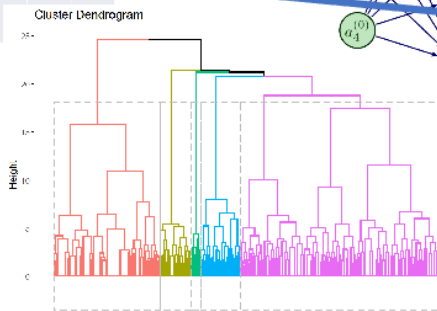
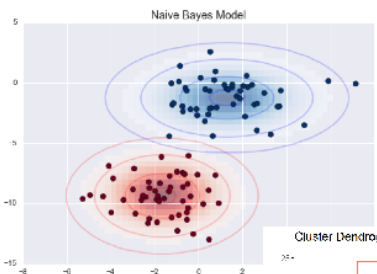
Traitement de données hétérogènes (texte, nombres, ...)

Performances non garanties

Explicabilité / boîtes noires

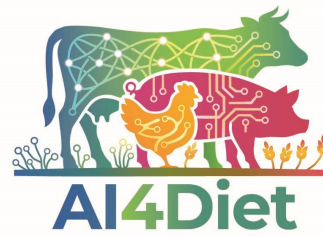
# De nombreux modèles disponibles

- De nombreux modèles sont disponibles
  - Différentes propriétés adaptées à différents types de problèmes & applications
  - Approches bayésienne, arbres de décision, forêts aléatoires, ...

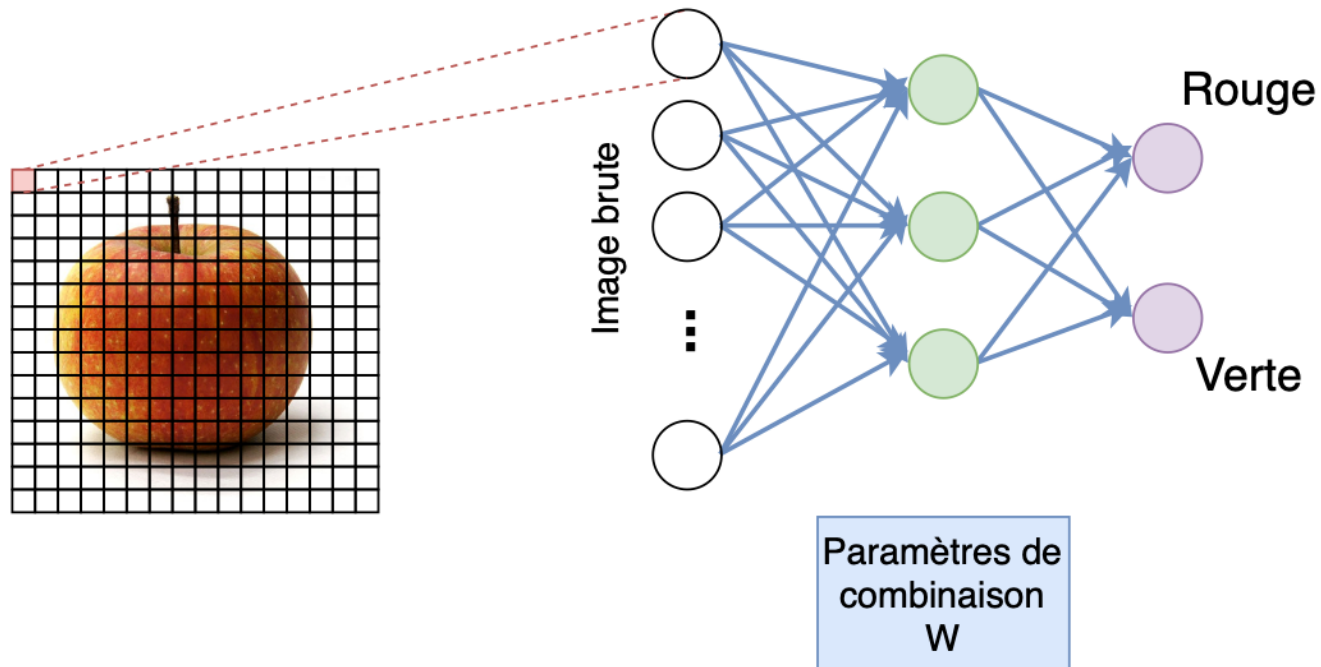


$$f\left(\begin{matrix} \square & \square & \square & \square \\ \mathbf{x} \in X \end{matrix}\right) = \text{pred} \hat{y} \in Y$$

# Réseaux de neurones : architectures complexes...

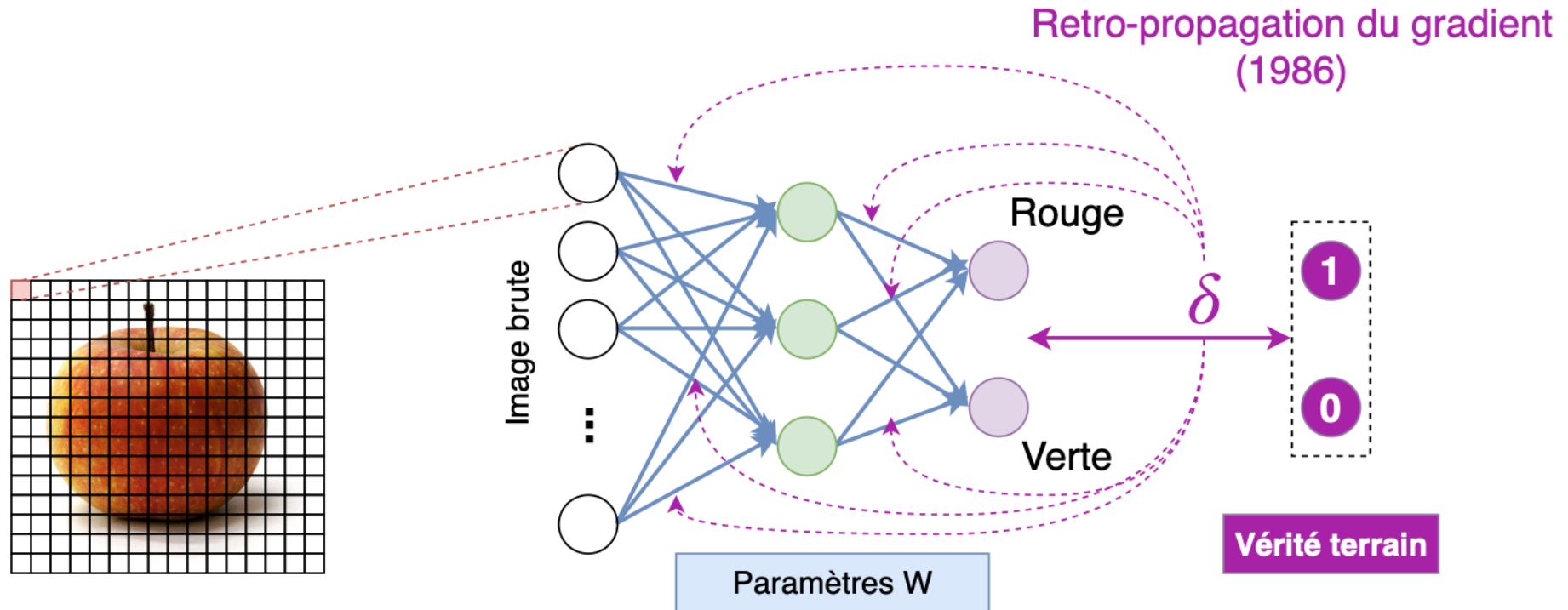


- Opérateur mathématique hyper-complexe
  - Gestion de données brutes riches (texte, image, ...)
- Système de correction (progressive) des erreurs

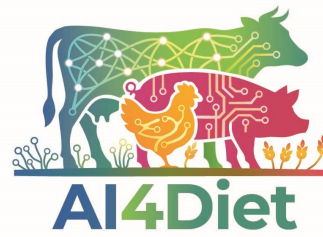


# Réseaux de neurones : architectures complexes...

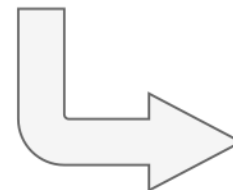
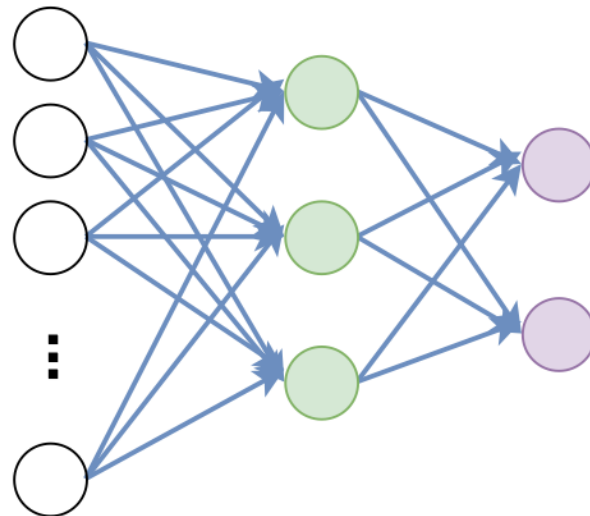
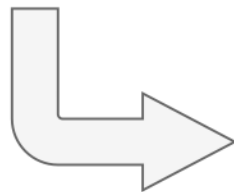
- Opérateur mathématique hyper-complexe
  - Gestion de données brutes riches (texte, image, ...)
- Système de correction (progressive) des erreurs



# Réseaux de neurones : architectures complexes...



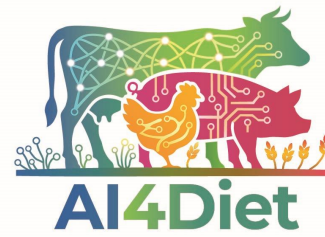
- Opérateur mathématique hyper-complexe
  - Gestion de données brutes riches (texte, image, ...)
- Système de correction (progressive) des erreurs



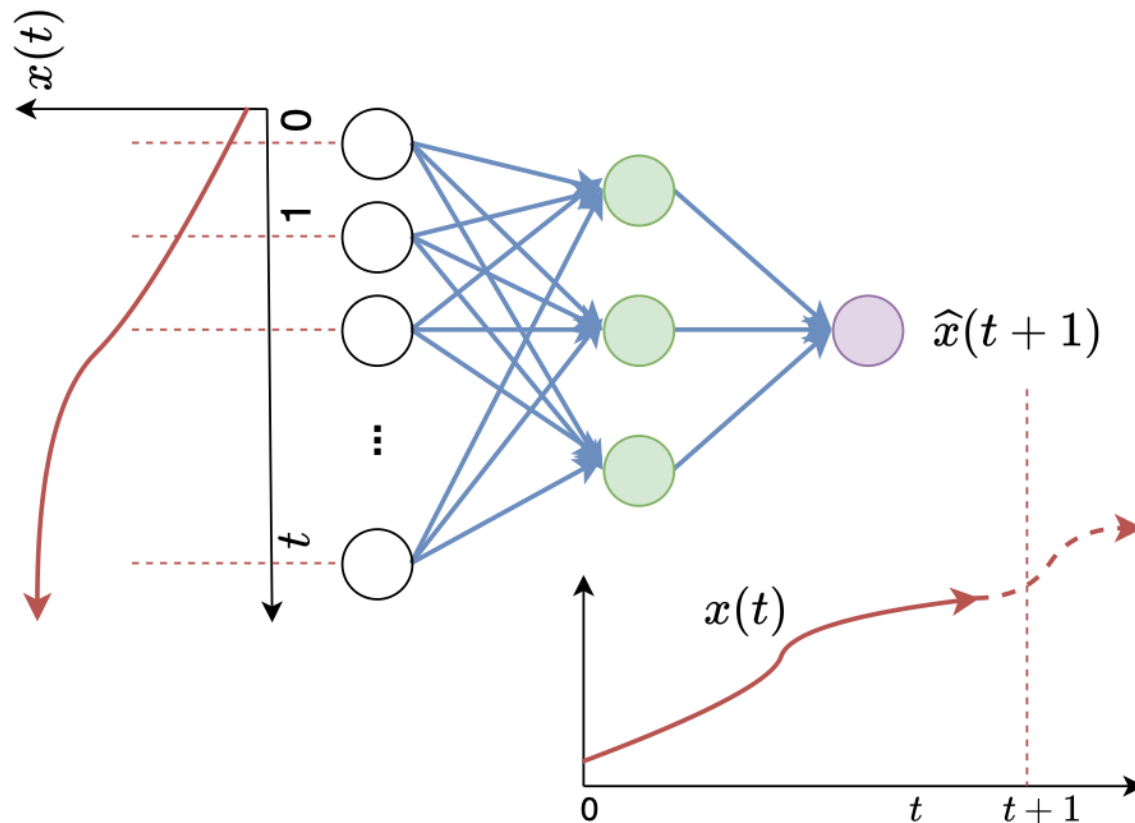
Verte  
Rouge  
Rouge

- Une fois appris, le réseau est figé
- Beaucoup plus rapide

# Vers des modèles génératifs : réseaux de neurones récurrents

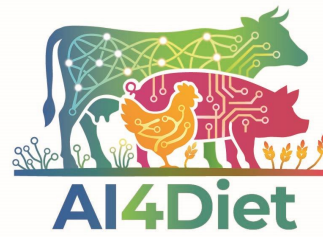


- Prédire une valeur / prédire l'évolution d'un processus (= séries temporelles)

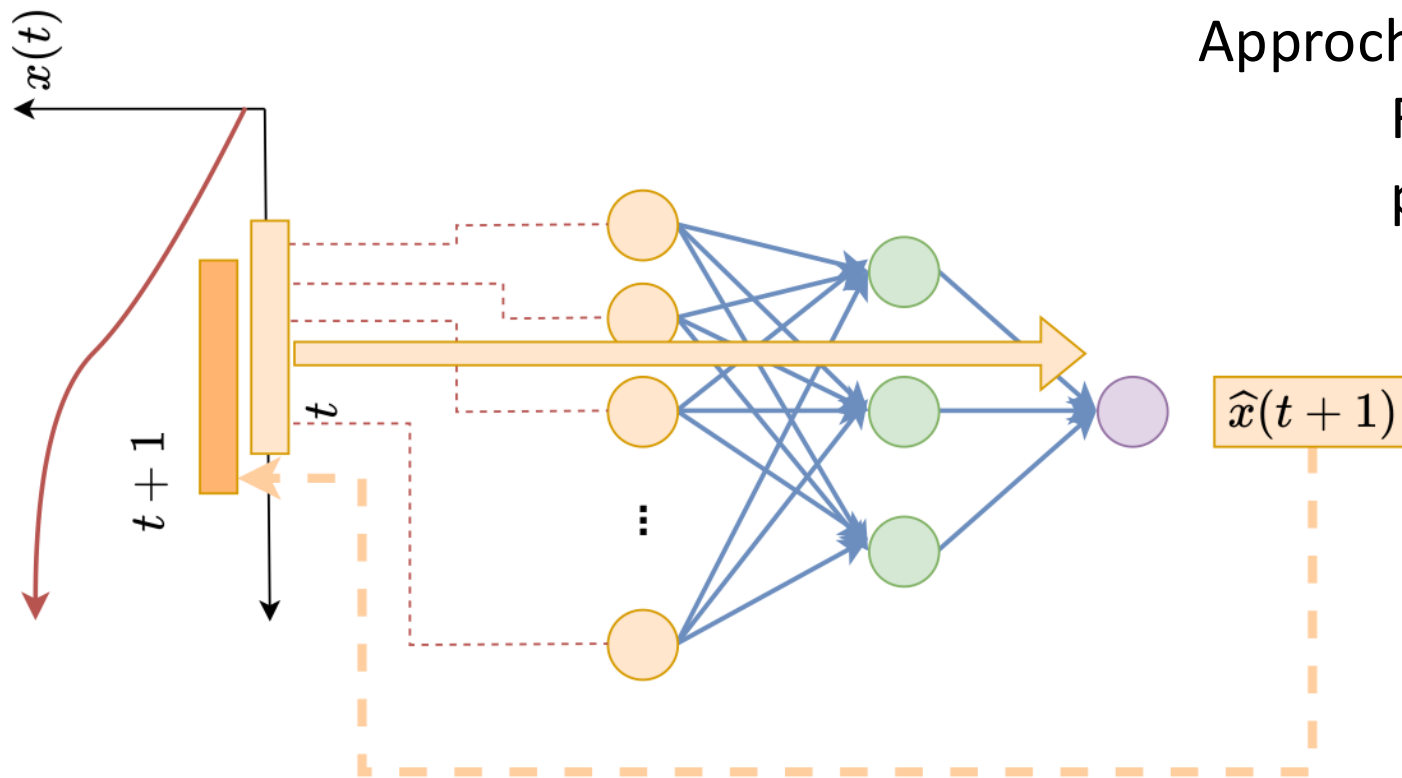


- Cours de bourse
- Consommation électrique
- Trafic routier
- Courbe de croissance animale/végétale
- ...

# Vers des modèles génératifs : réseaux de neurones récurrents

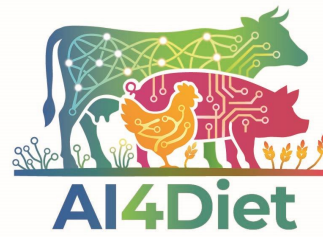


- Prédire une valeur / prédire l'évolution d'un processus

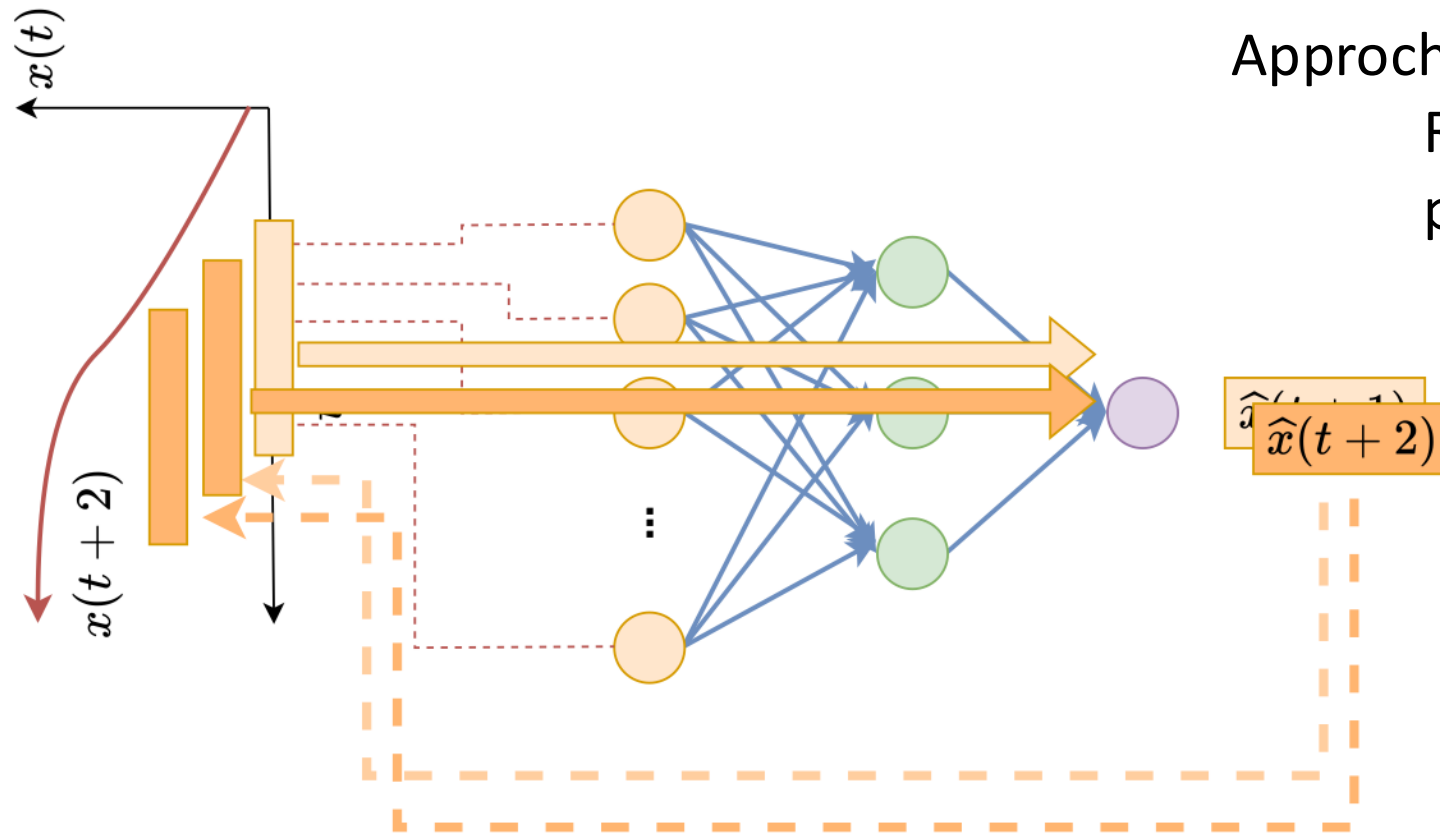


Approche auto-régressive:  
Ré-injection des valeurs prédites =  
prédictions à long terme

# Vers des modèles génératifs : réseaux de neurones récurrents

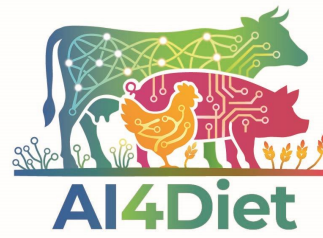


- Prédire une valeur / prédire l'évolution d'un processus



Approche auto-régressive:  
Ré-injection des valeurs prédites =  
prédictions à long terme

# Données nombreuses & hétérogènes



- Production de lait d'une vache
  - Alimentation
    - Quantité, composition, dynamique
  - Données de contexte
    - Température, humidité, ...
    - Saison
    - Altitude
    - Type de bâtiment
  - Caractéristique de la vache
    - Génétique / lignée
    - Age
    - Stade de lactation
    - Morphologie
    - Date de vêlage, succès reproductif
  - Santé
    - Mamite, ...
  - Pratique d'exploitation

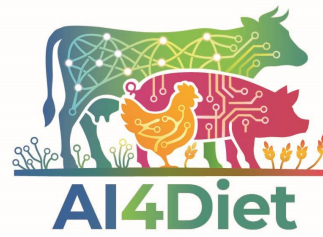
- Taille troupeau,
- Traite robot/salle
- Temps de paturage

- Variable d'intérêt
  - Quantité
  - Taux protéique

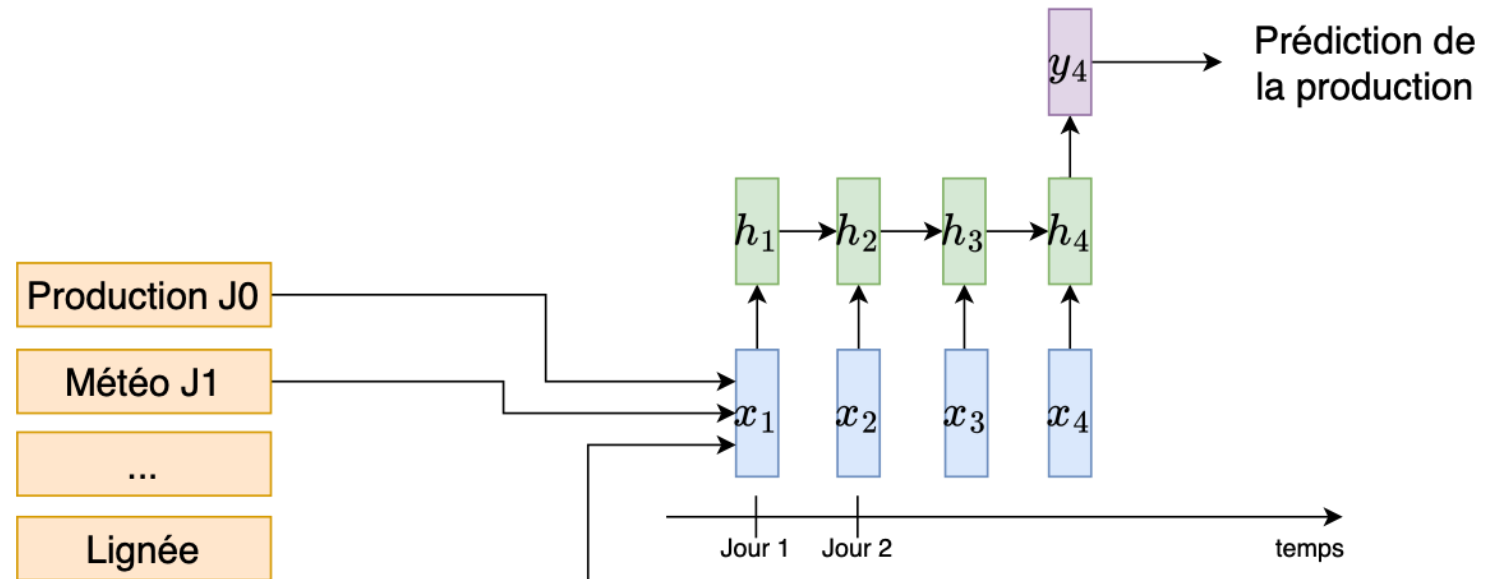


Il faut aussi prendre en compte l'inertie de la vache:  
=> la référence la plus classique est de faire une  
moyenne du passé

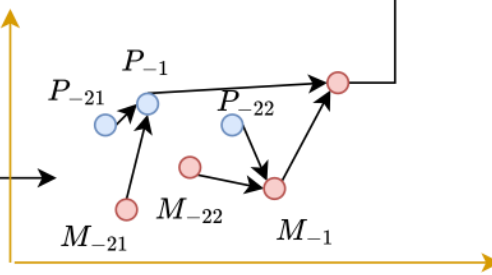
# Construire une architecture sur mesure



- Variables continues
- Variables discrètes
- Dynamique temporelle
- ...

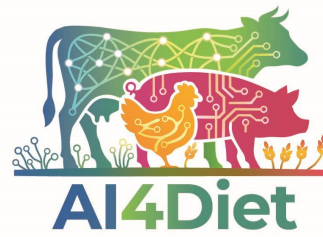


Il est possible d'inclure des données textuelles...  
Si elles sont intéressantes + si on les a observées dans notre jeu de données

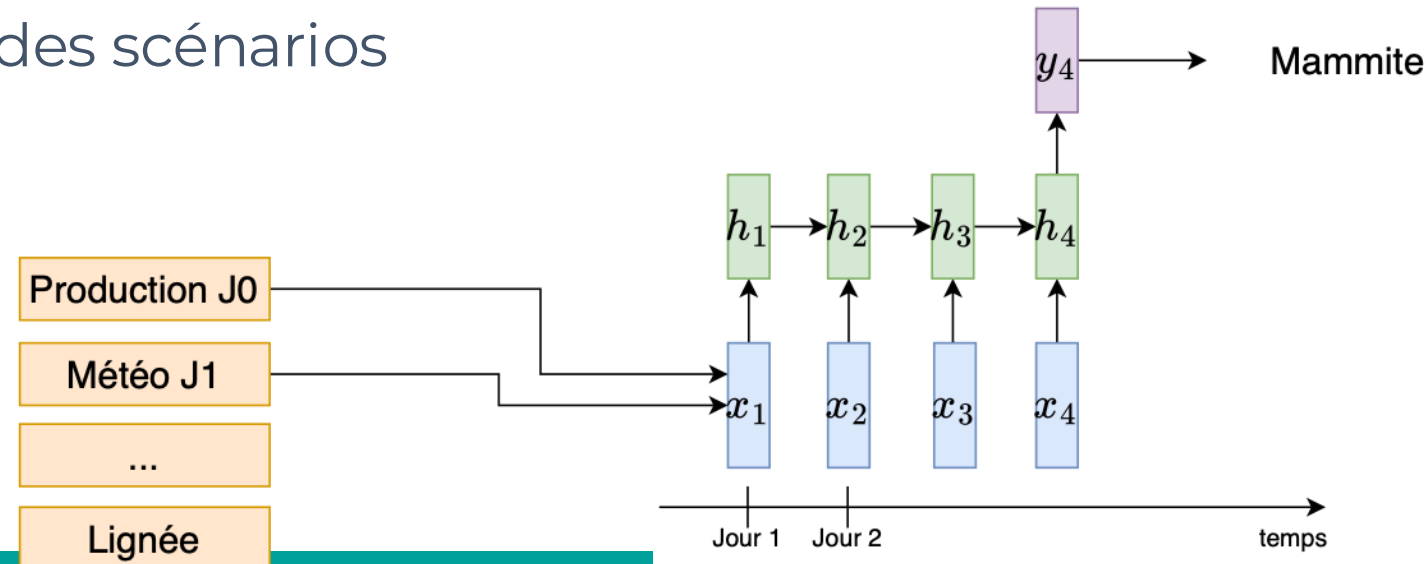


Représentation des variables discrètes  
Modélisation des dépendances

# Une modélisation, différents problèmes

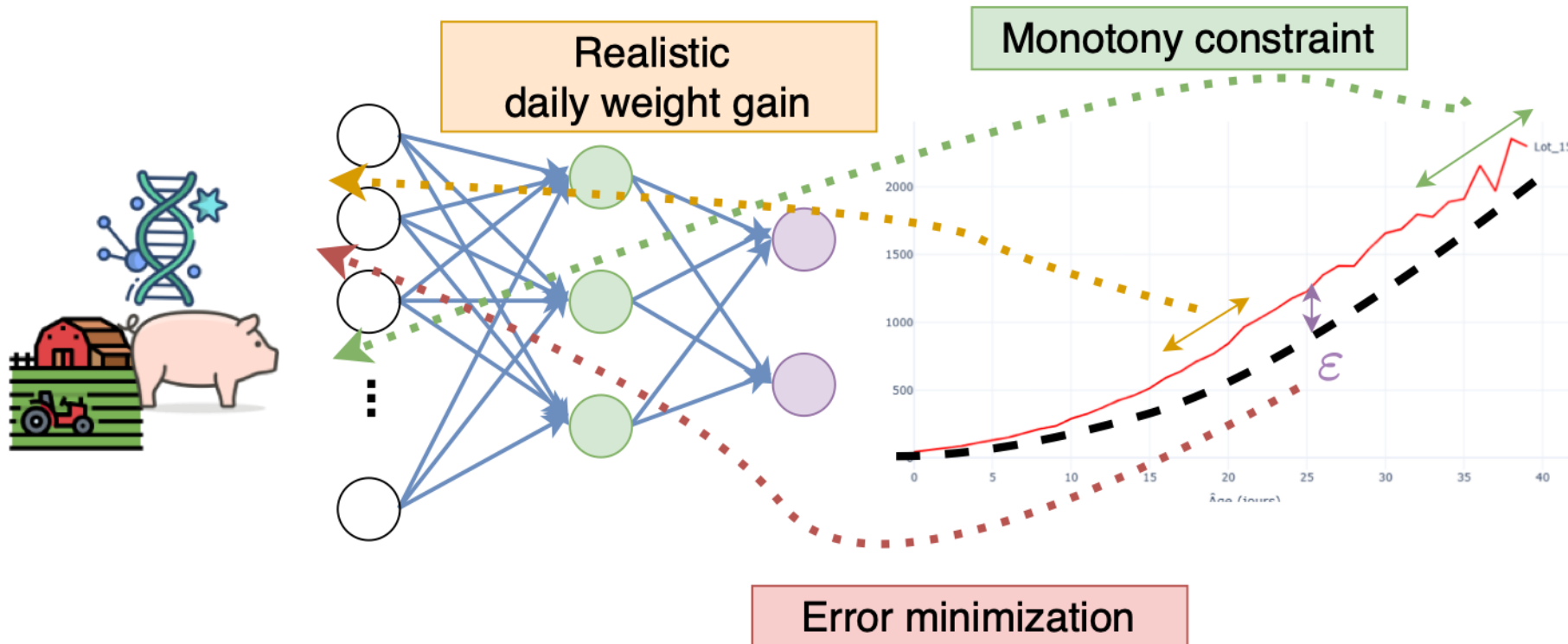


- Prédiction de la production de lait
- Bien-être animal
- Prédiction des maladies / impact des facteurs observés
  - Y a-t-il des signaux précurseurs détectables dans les données collectées?
  - L'activité de la vache dans la journée impacte-t-elle la production de lait?
- Utiliser le prédicteur pour projeter des scénarios
  - Tester virtuellement des aliments
  - Optimiser l'exploitation
  - = base des jumeaux numériques



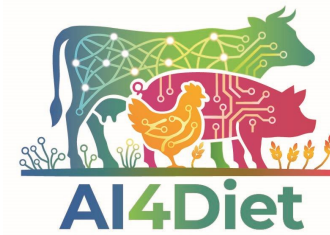
# Réseaux de neurones : architectures malléables

- Multiplier les contraintes = multiplier les objectifs (fonctions de coût)
  - Une autre manière d'introduire de l'expertise
- Design de l'architecture = garantie de certaines propriétés (e.g. monotonie)

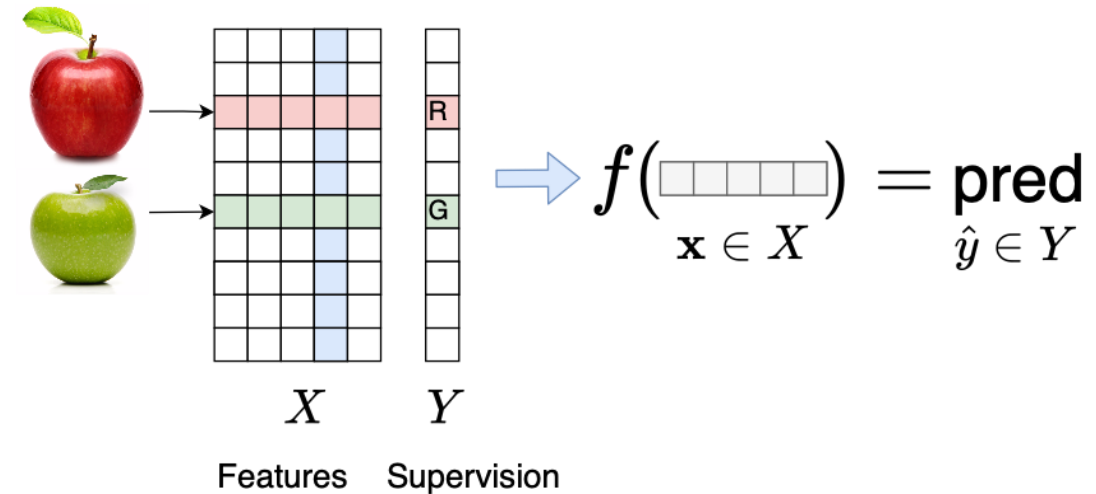


- La vache n'a pas produit de lait
  - Maladie?
  - Données non-collectées?
  - => Recoupement de bases de données + hypothèses
- Les pratiques agricoles ne sont pas disponibles
  - On peut modéliser l'impact de la variable *Exploitation* sur la production
  - + Cartographier les exploitations les uns par rapport aux autres
  - ... Mais dans un **espace abstrait** + à condition d'avoir des données de production comparable sur **plusieurs exploitations**
- J'ai des données en faible quantité...
  - ... Mais je dispose d'un modèle de production de lait
    - Mélanger les données réelles et synthétiques?

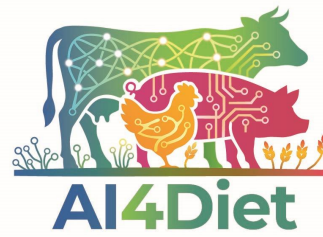
# Qu'est ce qu'un jeu de données exploitable?



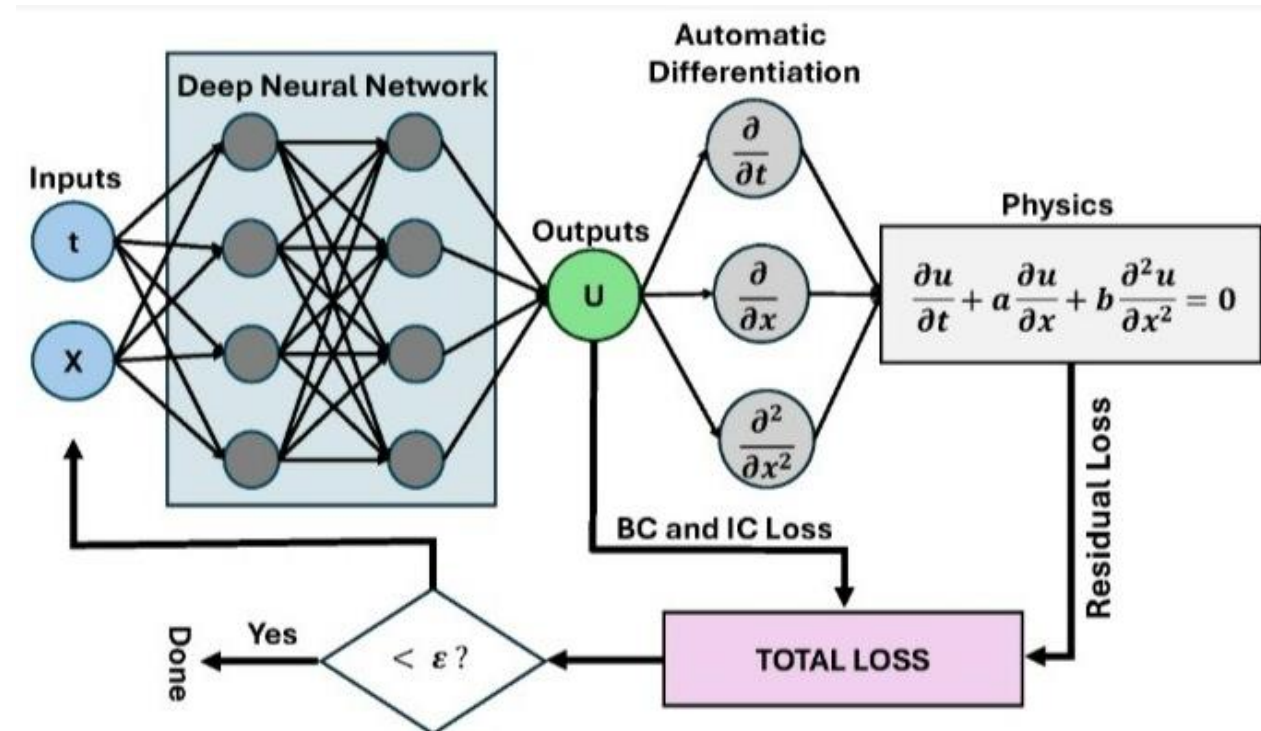
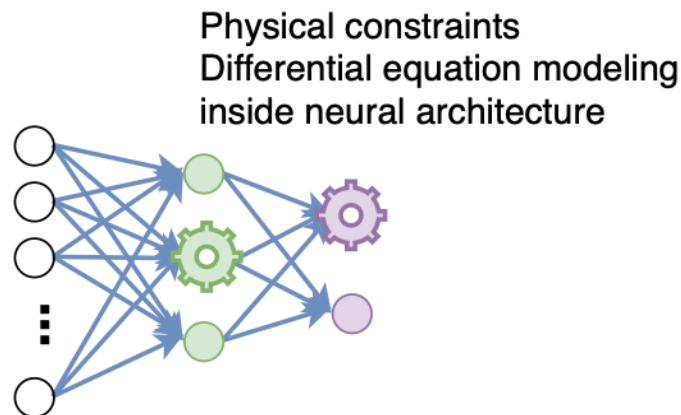
- Quantité des données nécessaire
  - Vérifier l'intérêt de l'application
- Type des données
  - Excel, capteurs numériques
  - Données catégorielles
  - Séries temporelles
  - Evènements (pratiques agricoles, ...)
- Supervision
  - Est-ce qu'il est possible/facile/difficile pour un humain de faire cette prédiction?
  - Est-ce qu'on a un historique complet/suffisant de ce qu'on cherche à prédire?
    - Coté observation + coté supervision
- Variation
  - Quels sont les facteurs de variations impactant la cible?
  - Combien de type de variations sont observées?



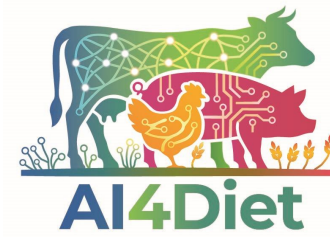
# Hybridation des approches : data + mécaniste



- Pourquoi hybrider?
  - Gagner de la vitesse
  - Universalité
  - Exploitation des masses de données disponibles
- Physics Informed Neural Networks
  - Combiner EDP / réseaux de neurones
  - Beaucoup de théorie



# Hybridation des approches : data + mécaniste (2)



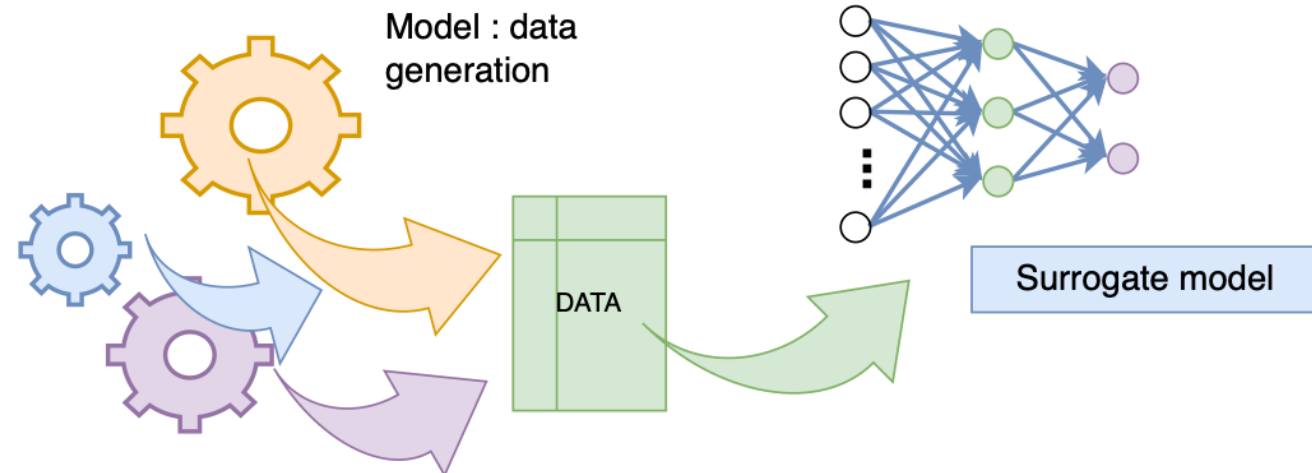
- Pourquoi hybrider?
  - Gagner de la vitesse
  - Universalité
  - Exploitation des masses de données disponibles
  - Limiter les entrées nécessaires
- Génération de données
  - + Combinaison avec observations
  - Calibration
- Unification de modèles

Mechanistic model / simulation

Slow / costly  
Accurate  
Interpretable

Data driven

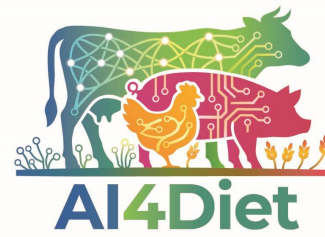
Fast  
Approximation



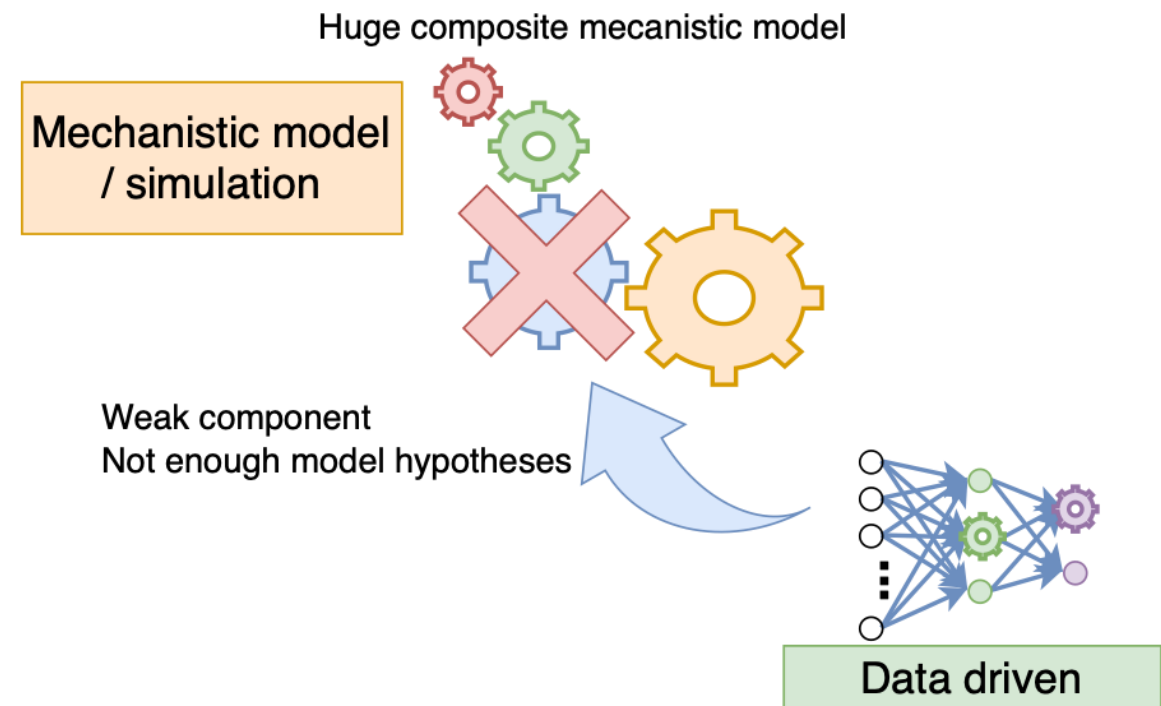
Ex1: Wofost

Ex2: Projet AFZ

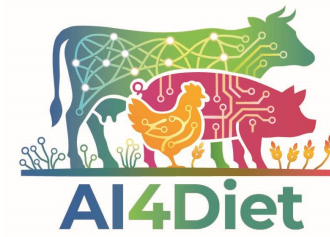
# Hybridation des approches : data + mécaniste (3)



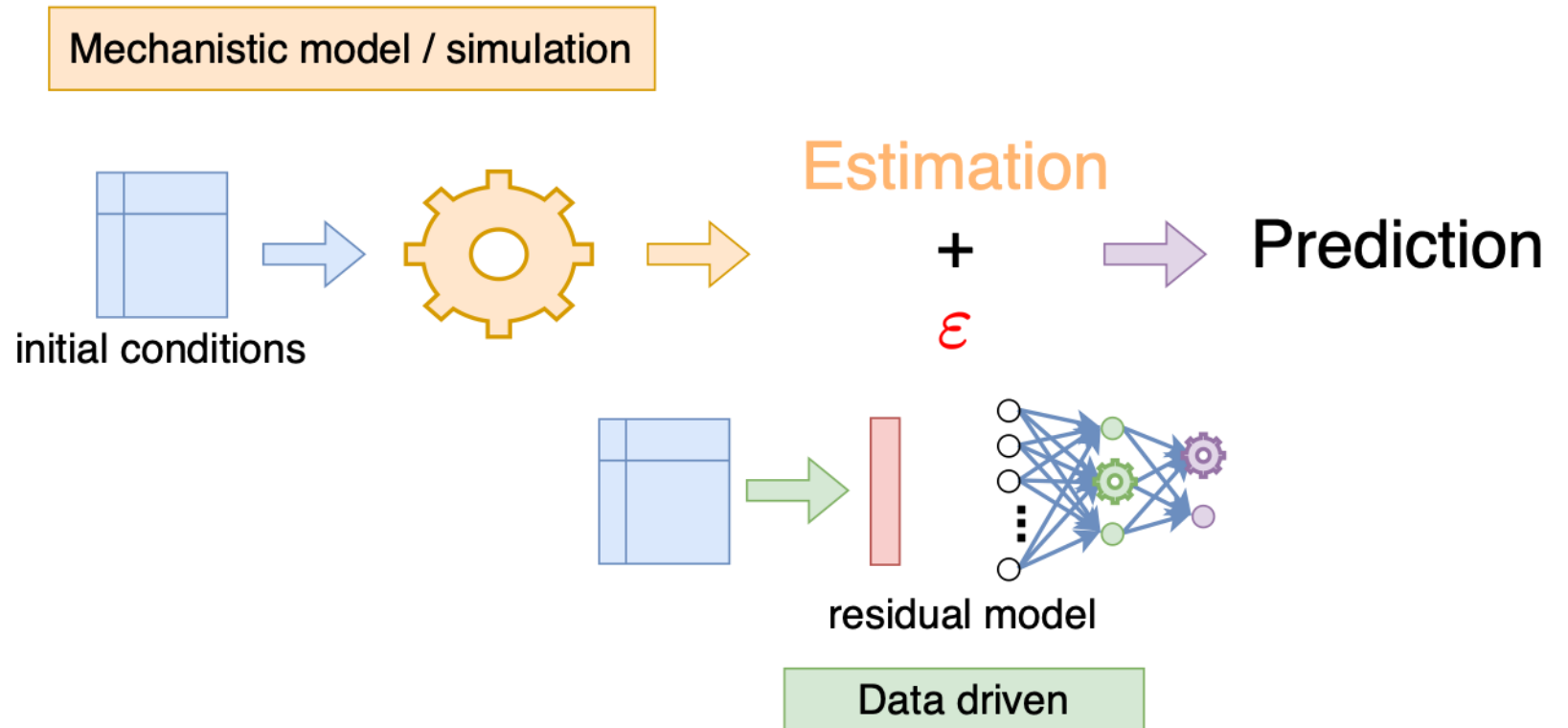
- Pourquoi hybrider?
  - Gagner de la vitesse
  - Universalité
  - Exploitation des masses de données disponibles
- Système complexe
  - Remplacer le(s) maillon(s) faible(s)



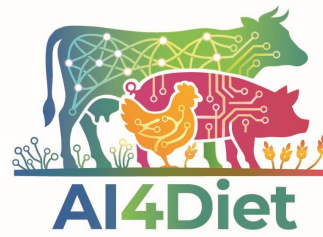
# Hybridation des approches : data + mécaniste (4)



- Pourquoi hybrider?
  - Gagner de la vitesse
  - Universalité
  - Exploitation des masses de données disponibles
- Combinaison résiduelle



# Limites des modèles: le domaine de définition



## Approches data-driven

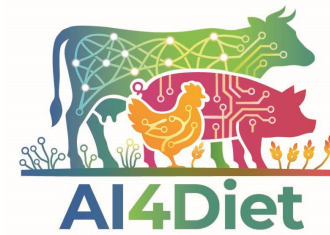
- Les données observées décrivent un domaine de validité

## Approches modélisation expert

- Les hypothèses de modéliser forment un domaine de validité

Cas difficile = OOD : Out Of Distribution

# Limites des modèles: le domaine de définition



## Approches data-driven

- Les données observées décrivent un domaine de validité

## Approches modélisation expert

- Les hypothèses de modéliser forment un domaine de validité

- **Chaque échantillon est caractérisé par :**

- Son Nom : 73 différents
- Ses valeurs physico-chimiques
- Ses valeurs énergétique sur différents animaux → **Valeurs cibles**

- **6422** échantillons uniques (une mesure faite sur un aliment)

Nom	MS% brut	PB % brut	CB % brut	MGR % brut	MM % brut	NDF % brut	ADF % brut	Lignine % brut	Amidon % brut	Sucres % brut
Avoine	87	9,5	12,6	3,2	3,8	33,4	15,4	2,4	46,2	1,4
Avoine	87	11,9	11,8	2,6	2,5	31,7	14,5	2,3	44,4	1,7

EB (kcal) kcal/kg brut	ED porc croissance (kcal) kcal/kg	EM porc croissance (kcal) kcal/kg	EN porc croissance (kcal) kcal/kg	EMAn coq (kcal) kcal/kg brut	EMAn poulet (kcal) kcal/kg brut	UFL 2018 par kg brut	UFV 2018 par kg brut	PDIA 2018 g/kg brut	PDI 2018 g/kg brut	BalPro Ru 2018 g/kg brut
3920	2460	2380	1850	2610	2480	0,81	0,76	17	62	-10
3970	2580	2480	1880	2590	2460	0,88	0,85	22	69	5

**10 valeurs chimiques et 11 valeurs énergétiques cibles**

# Limites des modèles: le domaine de définition

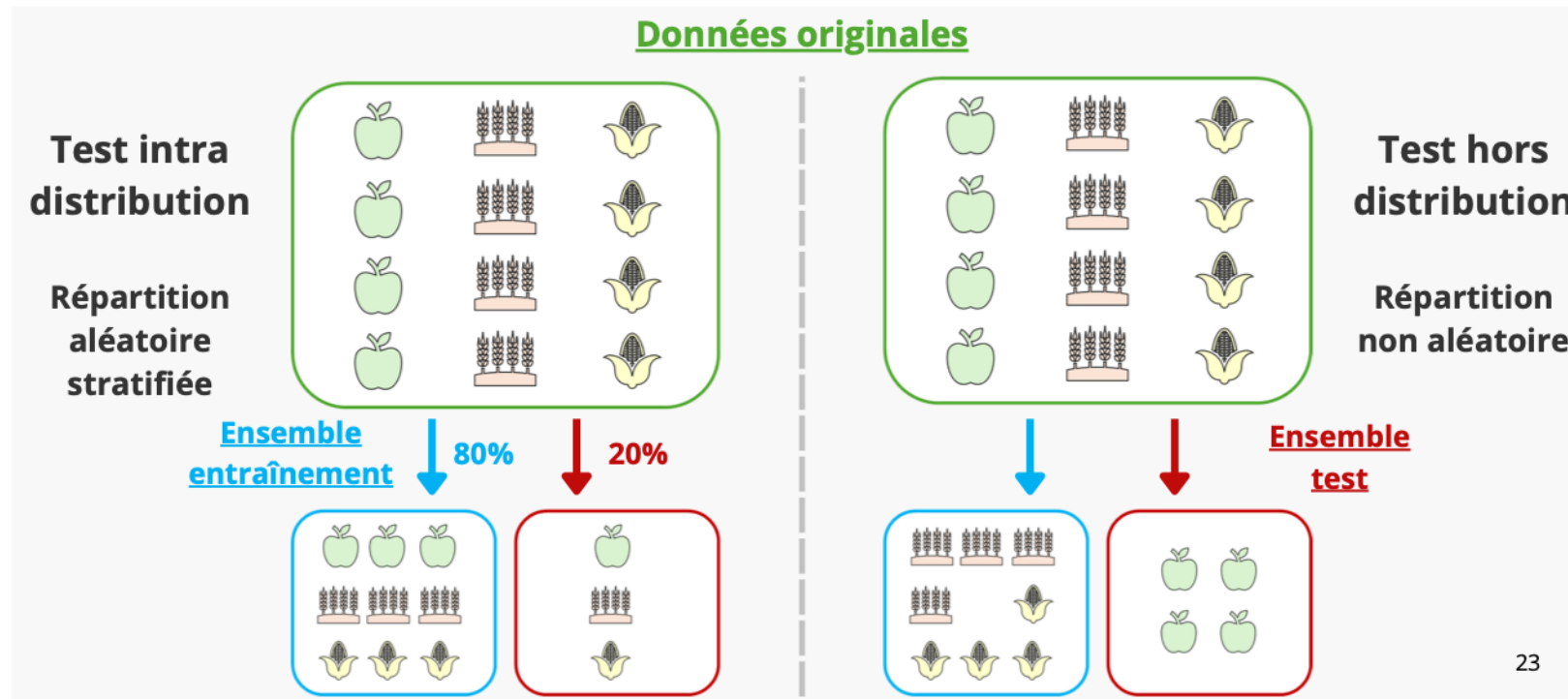


## Approches data-driven

- Les données observées décrivent un domaine de validité

## Approches modélisation expert

- Les hypothèses de modéliser forment un domaine de validité



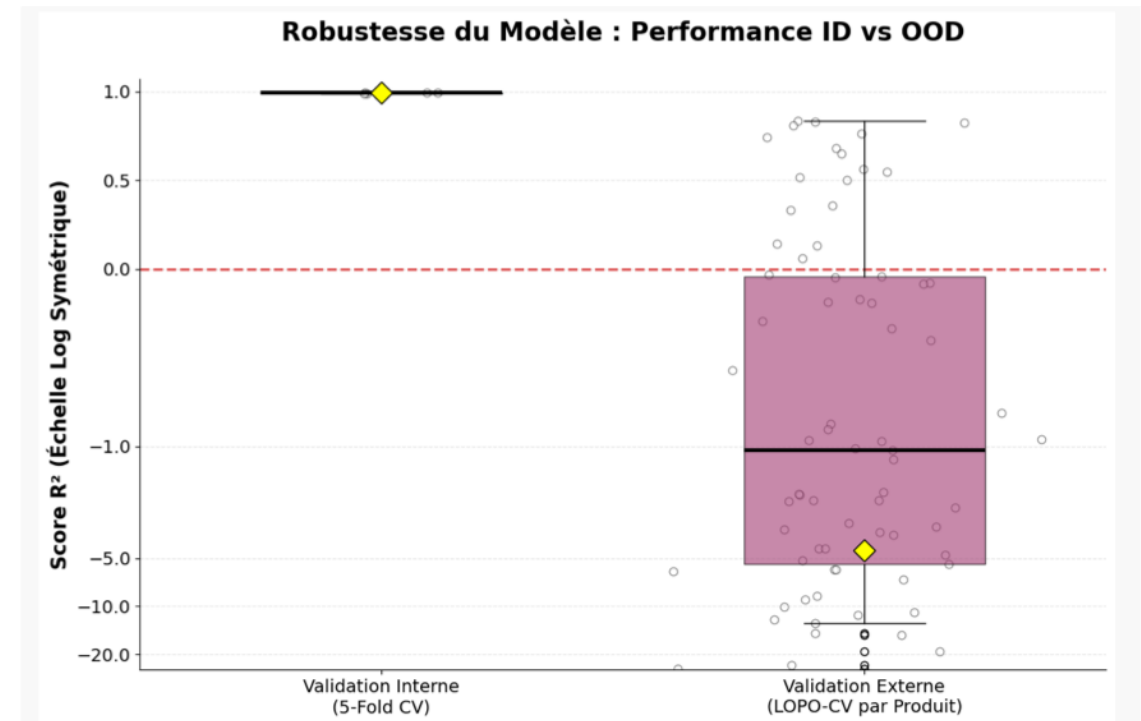
## Approches data-driven

- Les données observées décrivent un domaine de validité

## Approches modélisation expert

- Les hypothèses de modéliser forment un domaine de validité

- Parfait dans le scénario 1...
- Catastrophique dans le scénario 2



- Data: apprentissage sur des observations = un domaine
- Mécaniste : définition d'équations... Variables sur un domaine
- Que se passe-t-il lorsque l'on s'éloigne?
  - Nouvelle génétique
  - Réchauffement climatique
  - Aliments locaux mal caractérisés
  - Aliments exotiques

Photo ( train )



Clipart ( test )



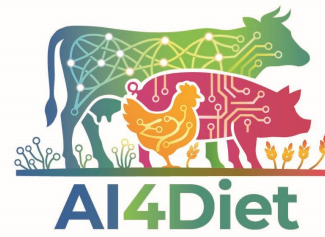
Sketch ( test )



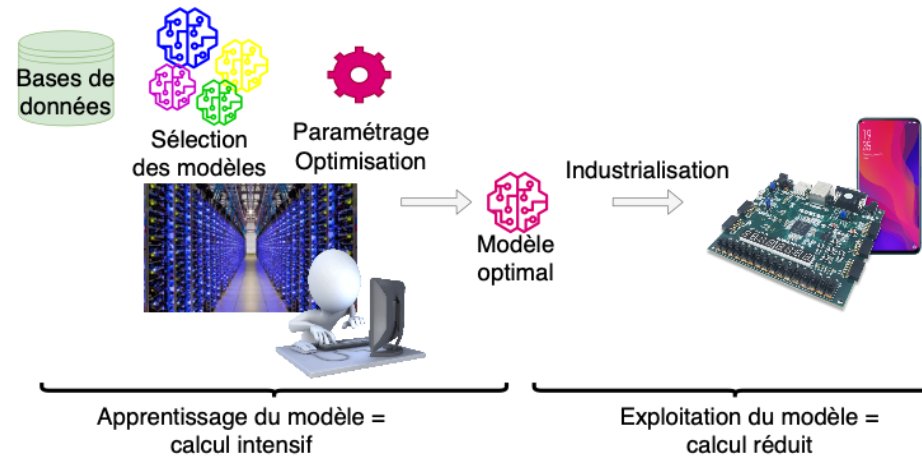
## => Augmenter le domaine de validité des modèles

- Combinaison de modèles
- Hypothèses sur les autres domaines
- Régularisation / Intégration de nouvelles données (e.g. textuelles)

# Conclusion

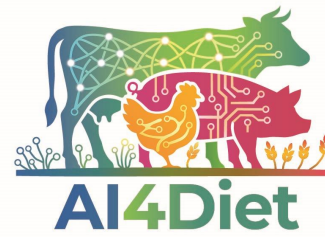


- Ne pas confondre entraînement et exploitation du modèle
- Le machine-learning est souvent plus léger que les approches mécanistes



- On ne prédit que ce que l'on a modélisé et/ou observé
  - Même si on a quelques astuces pour approximer des données manquantes
- Les modèles ont un domaine de validité limité

# Prédire ou Simuler?



- Prédire = ne pas connaître le futur (contexte dynamique)
- Simuler = suivre un scénario pour le contexte futur

