

Journée formation IA en biologie médicale

INTRODUCTION

Dr Alexandre Godmer
Dr Guillaume Bachelot

Contexte (1) : l'intelligence artificielle pour tous ?

Le Monde

Quand la fiction dessine l'avenir de l'intelligence artificielle, de « Frankenstein » à « Terminator »

Par Elisa Thévenet

Publié le 02 septembre 2022 à 20h00 - Mis à jour le 03 septembre 2022 à 06h05

[Quand la fiction dessine l'avenir de l'intelligence artificielle, de « Frankenstein » à « Terminator » \(lemonde.fr\)](#)

Radiologie : pourquoi l'IA n'a (toujours) pas remplacé le médecin

Publié: 28 juin 2022, 19:04 CEST

[Intelligence artificielle et médecins : qui va gagner ? | EY - France](#)

Intelligence artificielle et médecins : qui va gagner ?

[Radiologie : pourquoi l'IA n'a \(toujours\) pas remplacé le médecin \(theconversation.com\)](#)

Santé numérique : des radiologues bientôt au chômage ?
<https://www.inria.fr/fr/intelligence-artificielle-radiologie-benefices>



[Faut-il avoir peur de l'Intelligence Artificielle ? - {Sciences2} \(lemonde.fr\)](#)

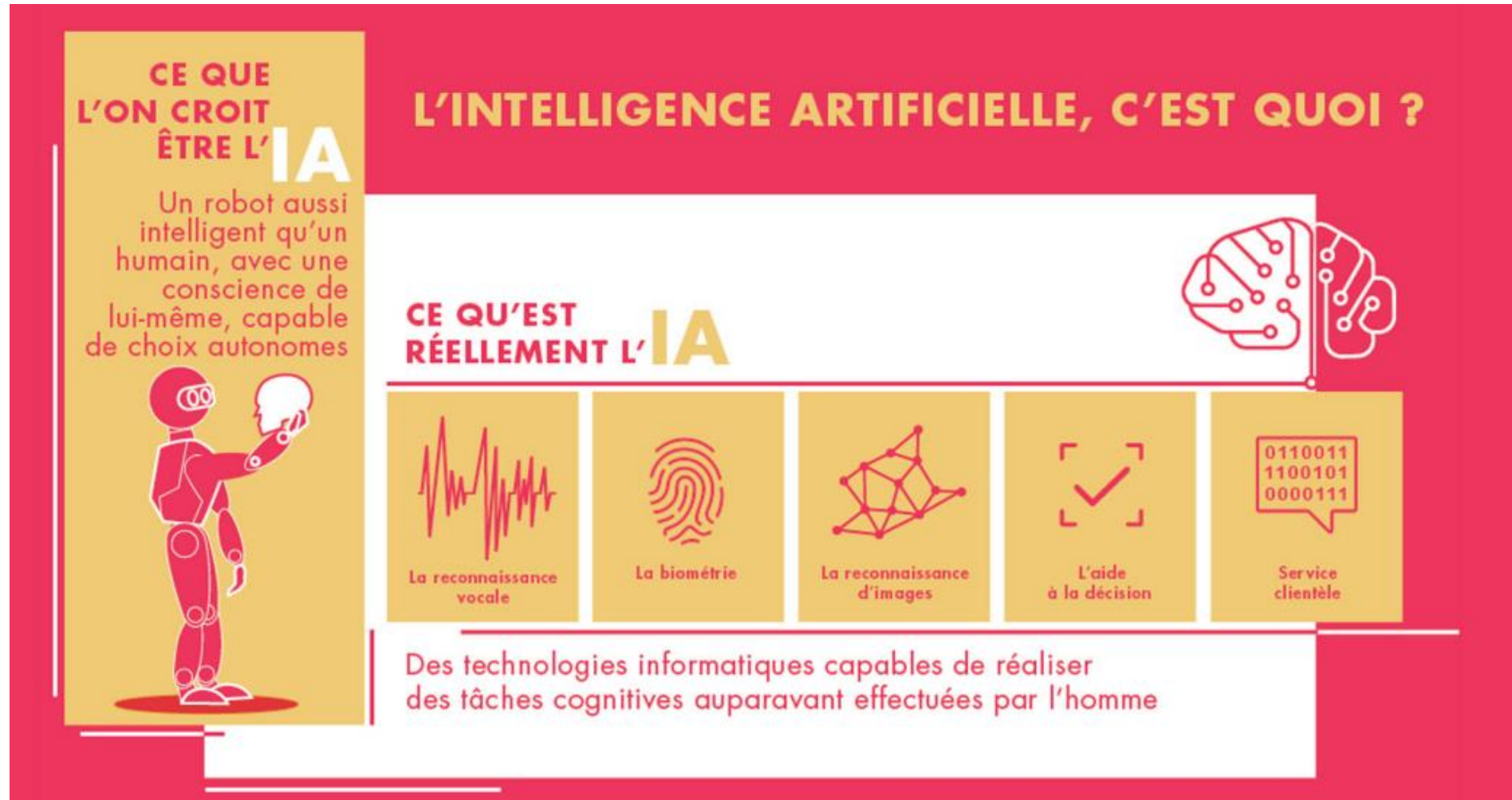
CNRS
LE JOURNAL

Une IA, juge de beauté des poissons tropicaux

[Une IA, juge de beauté des poissons tropicaux | CNRS Le journal](#)

Alexandre Godmer

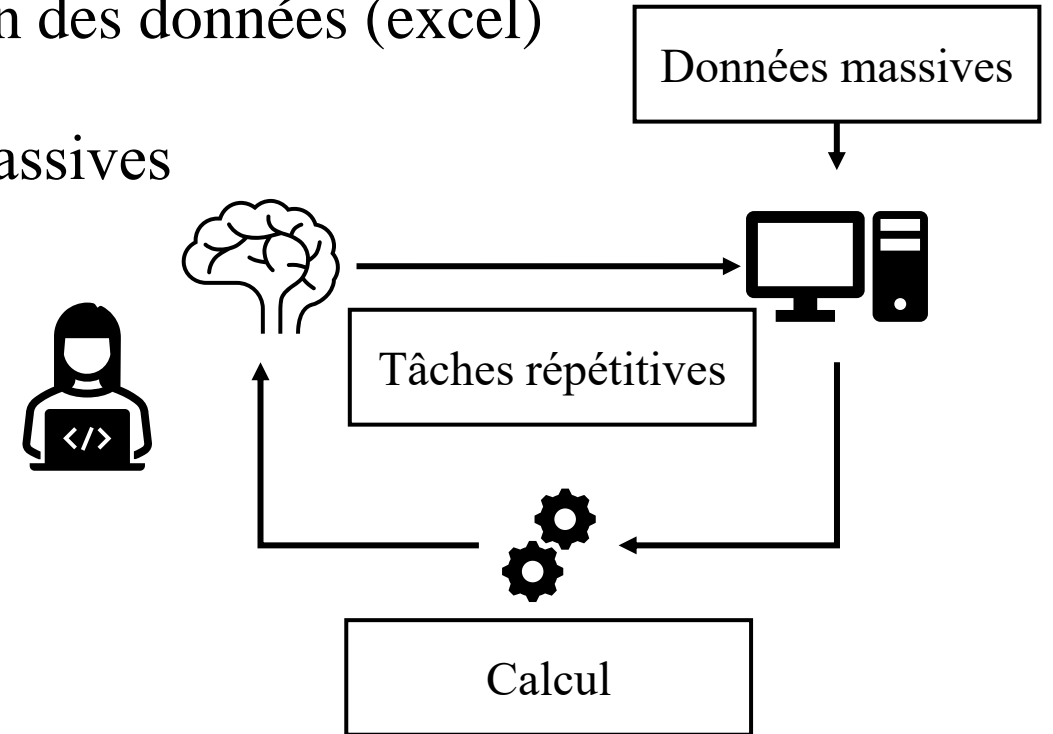
Contexte (2) : mais en fait...



[revue_personnel_s_benhamou_intelligence_artificielle.pdf \(strategie.gouv.fr\)](#)

Contexte (3) : pourquoi a-t-on besoin de l'intelligence artificielle ?

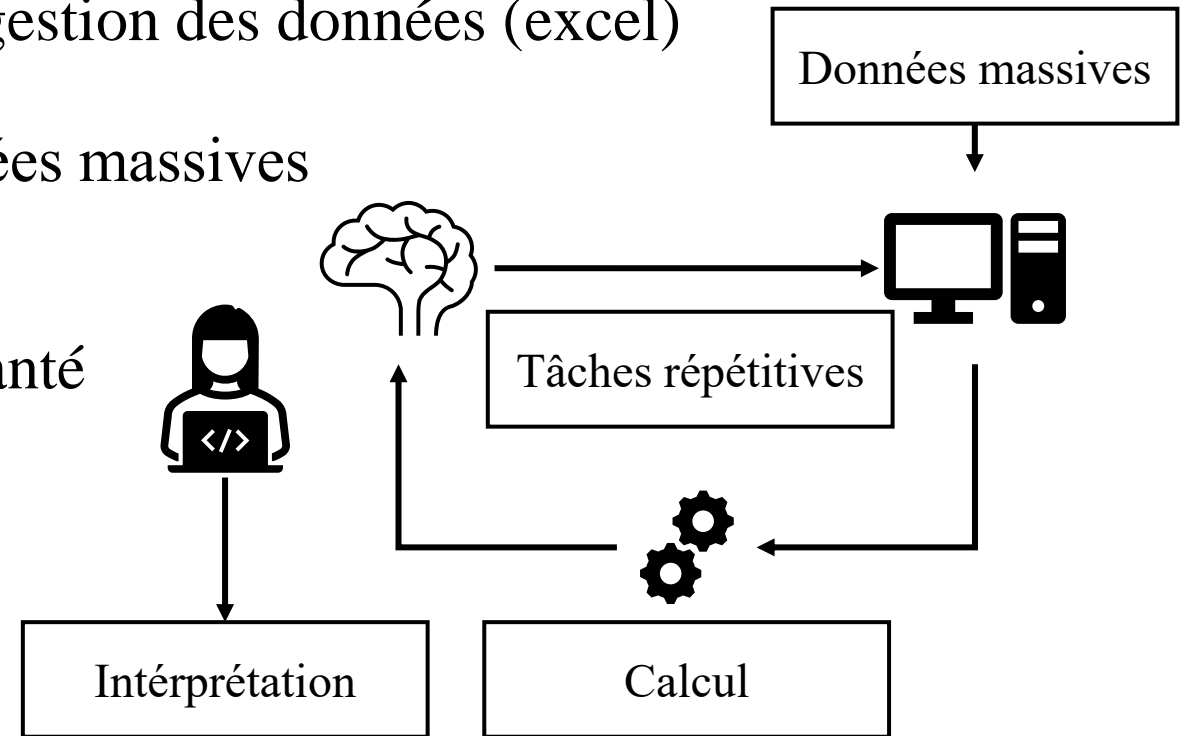
- **Augmentation de la production de données massives hétérogènes = big data**
 - limites de certains logiciels pour la gestion des données (excel)
- **Intelligence artificielle :**
 - pour l'analyse, tri des données de données massives
 - effectue des tâches répétitives
 - travaille en continu
 - création de systèmes de prédiction en santé
 - diminution du taux d'erreur
 - aide au diagnostic



cerveau et intelligence artificielle = collaboration utile !

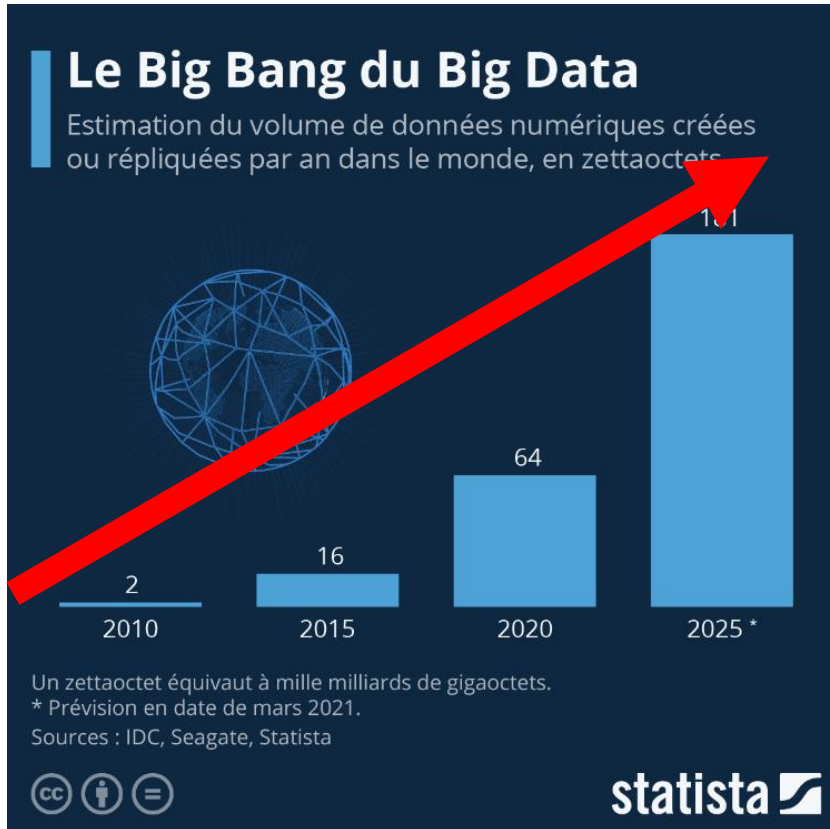
Contexte (3) : pourquoi a-t-on besoin de l'intelligence artificielle ?

- **Augmentation de la production de données massives hétérogènes = big data**
 - limites de certains logiciels pour la gestion des données (excel)
- **Intelligence artificielle :**
 - pour l'analyse, tri des données de données massives
 - effectue des tâches répétitives
 - travaille en continu
 - création de systèmes de prédiction en santé
 - diminution du taux d'erreur
 - aide au diagnostic



cerveau et intelligence artificielle = collaboration utile !

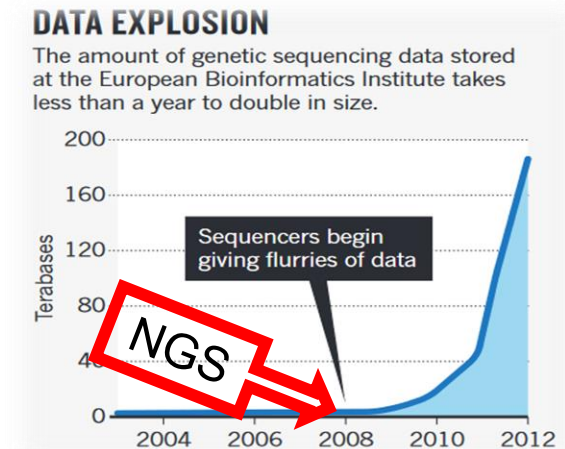
Contexte (4) : l'ère du Big Data



Graphique: Le Big Bang du Big Data | Statista

En santé → nouvelle compréhension des mécanismes des maladies

- recherche biomédicale axée sur les données massives
- computationnelle
- intégrative (de façon globalisée)



Marx, V. The big challenges of big data. *Nature* **498**, 255–260 (2013). <https://doi.org/10.1038/498255a>

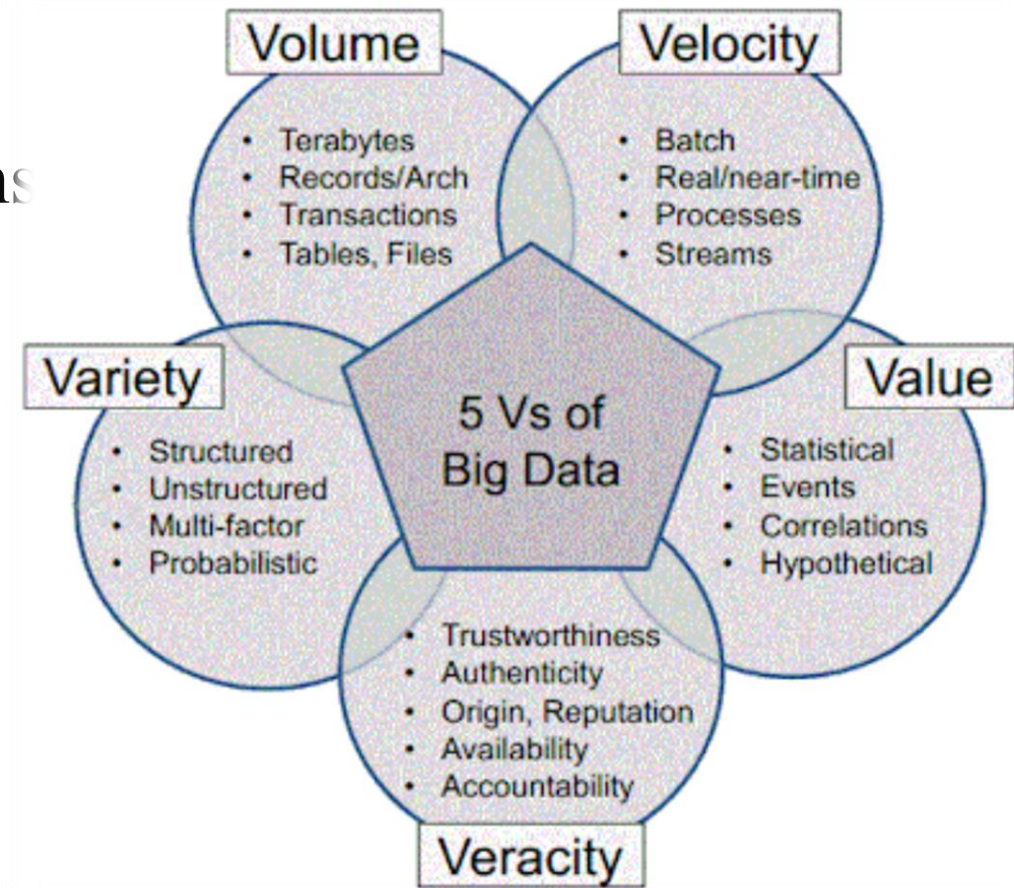
Big Data

« stockage, agencement et traitement de données massives hétérogènes »

Contexte (5) : l'ère du Big Data

Les caractéristiques des Big Data :

- **volume** :
→ considérable et en augmentation constante
- **vitesse** :
→ données collectées en temps réel
- **variété** :
→ données hétérogènes
- **valeur** :
→ quelle valeur ajoutée ?
→ pertinence
- **véracité** :
→ fiabilité des données




Marx, V. The big challenges of big data. *Nature* **498**, 255–260 (2013). <https://doi.org/10.1038/498255a>

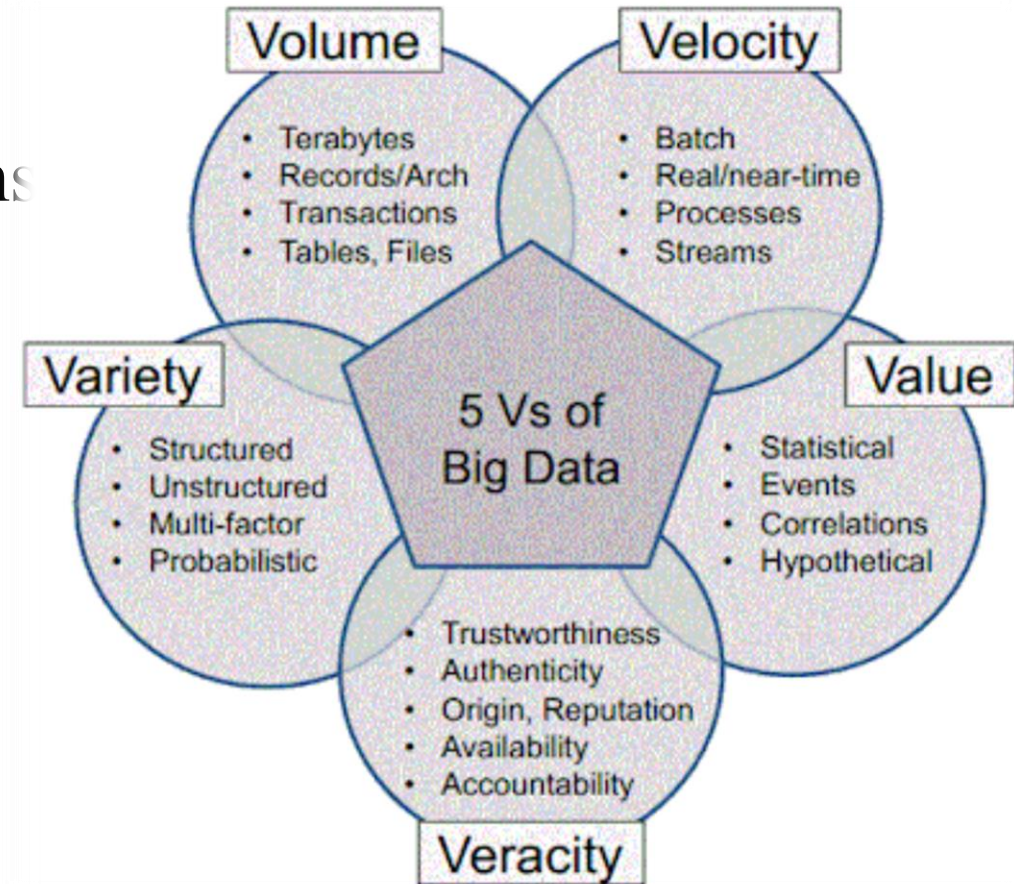
Domaine de la santé : importance de la véracité des données +++

Contexte (5) : l'ère du Big Data

Les caractéristiques des Big Data :

- **volume** :
→ considérable et en augmentation constante
- **vitesse** :
→ données collectées en temps réel
- **variété** :
→ données hétérogènes
- **valeur** :
→ quelle valeur ajoutée ? 
- pertinence
- **véracité** :
→ fiabilité des données

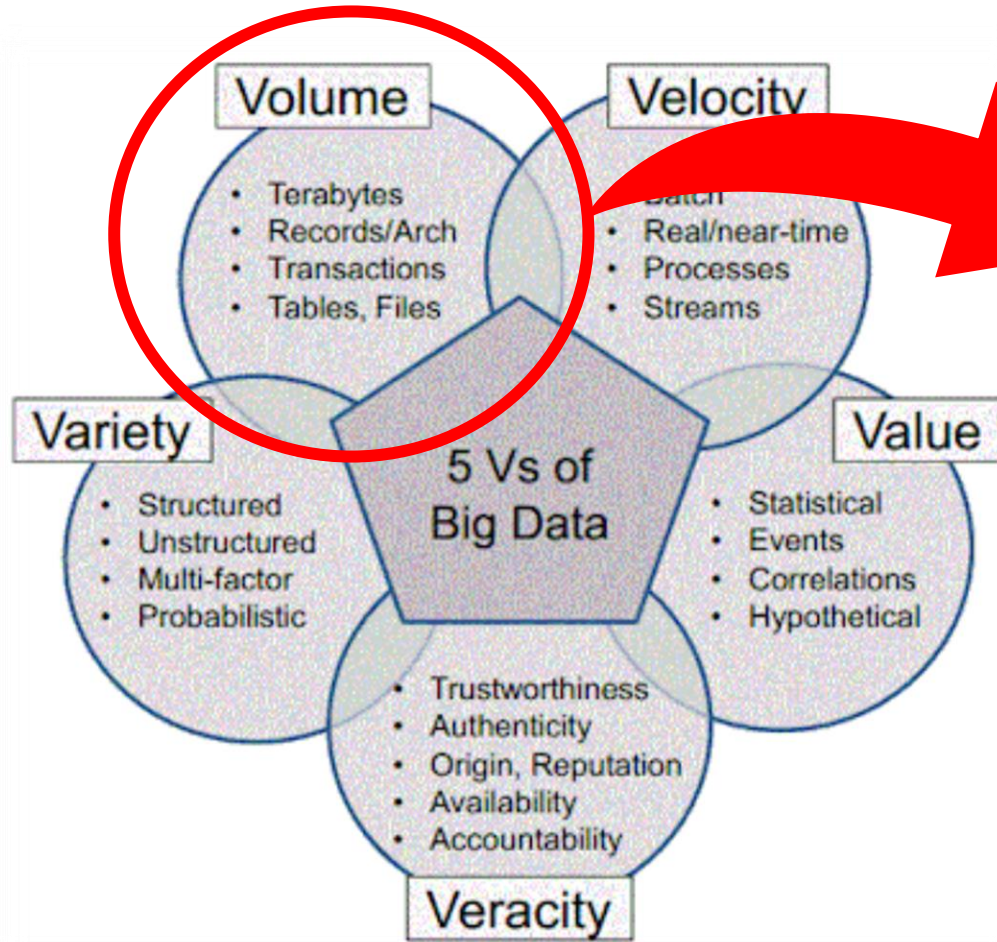
Nettoyage et vérification
des données



Marx, V. The big challenges of big data. *Nature* **498**, 255–260 (2013). <https://doi.org/10.1038/498255a>

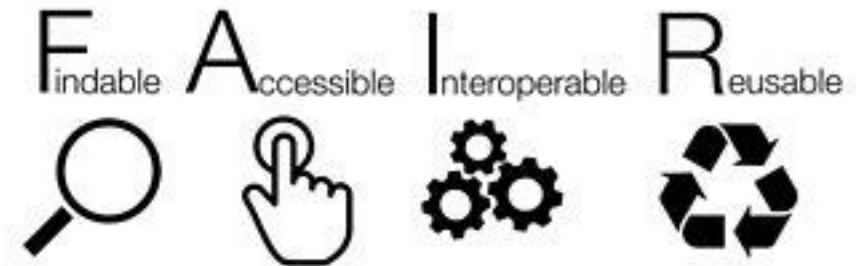
Domaine de la santé : importance de la véracité des données +++

Contexte (6) : Big data et outils de gestions



L'Entrepôt de Données de Santé de l'AP-HP, le plus gros entrepôt de données de santé hospitalier en Europe, contient aujourd'hui les données de plus de 11 millions de patients. Il appuie plus de 60 projets de recherche et d'innovation.

<https://eds.aphp.fr/nos-services/recherche-innovation>



https://fr.wikipedia.org/wiki/Fair_data#/media/Fichier:FAIR_data_principles.jpg

Marx, V. The big challenges of big data. *Nature* **498**, 255–260 (2013). <https://doi.org/10.1038/498255a>

Domaine de la santé :
Principe FAIR et sécurité des informations médicales +++

Alexandre Godmer

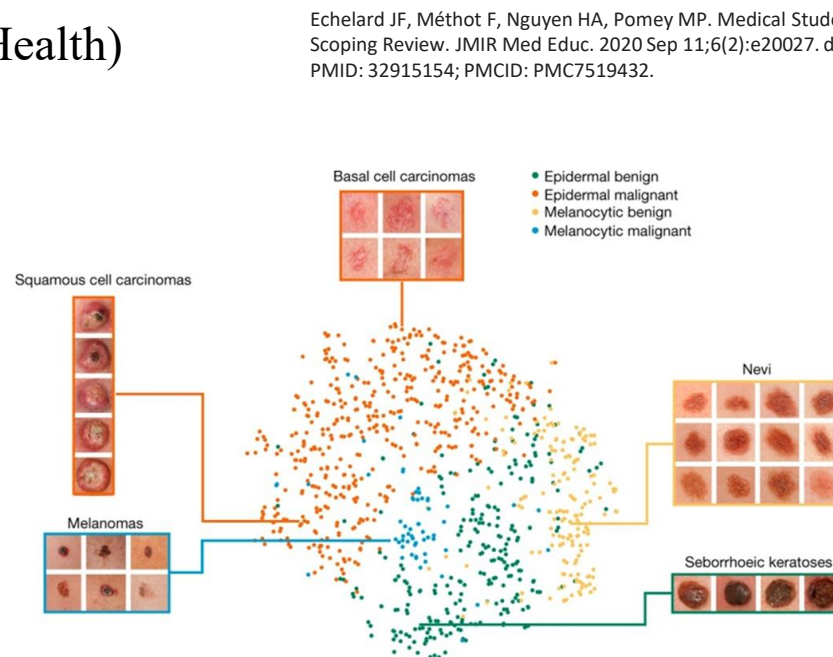
Contexte (7) : La santé numérique (e-santé, eHealth) :

- **Définition (OMS) (<https://www.who.int/observatories/global-observatory-for-ehealth>) :**
« *eHealth is the use of information and communication technologies (ICT) for health* »
- **Englobe de nombreuses technologies faisant appel à la programmation informatique (codage) :**
 - **Intelligence Artificielle (IA)**
 - Appareils connectés
 - Applications mobiles (mHealth)
 - ...

Year	Total Number of FDA Approved Algorithms
2014	1
2015	0
2016	4
2017	8
2018	25
2019 (as of June)	8
Total (as of June 2019)	46

Source: The Medical Futurist (6 June 2019) FDA Approvals For Smart Algorithms In Medicine In One Giant Infographic.

The rise of the data-driven physician. Stanford Medicine 2020 Health Trends Report. 2020. URL: <https://med.stanford.edu/dean/healthtrends.html>



Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017 Feb 2;542(7639):115-118. doi: 10.1038/nature21056. Epub 2017 Jan 25. Erratum in: Nature. 2017 Jun 28;546(7660):686. PMID: 28117445; PMCID: PMC8382232.

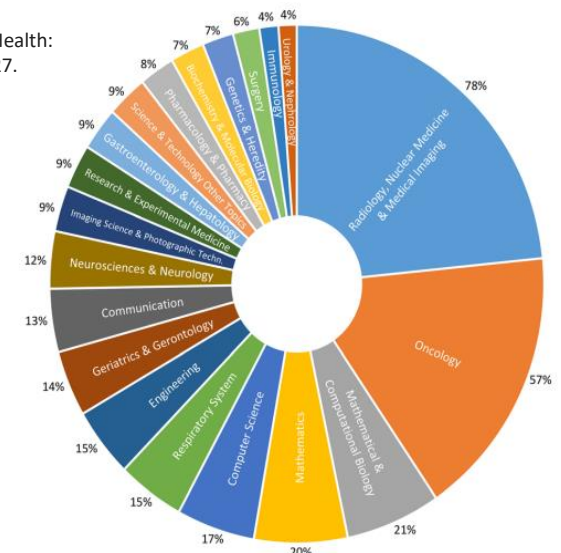


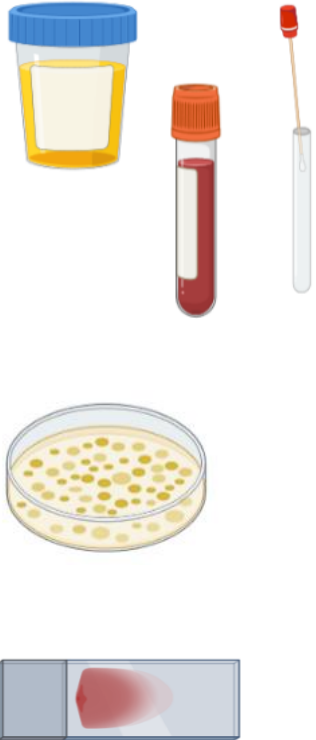
Fig. 3 Share of 3009 radiomics publications in the clinical and technical research areas assigned by the Web of Science for 2011 to 2019. Multiple assignments of research areas per publications are possible (see Supplement for further details on the methods)

Bukowski, M., Farkas, R., Beyan, O. *et al.* Implementation of eHealth and AI integrated diagnostics with multidisciplinary digitized data: are we ready from an international perspective?. *Eur Radiol* 30, 5510–5524 (2020). <https://doi.org/10.1007/s00330-020-06874-x>

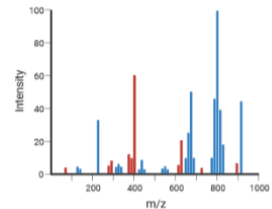
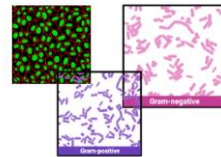
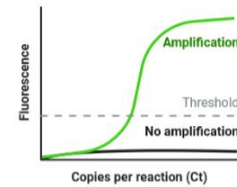
Intelligence artificielle = domaine en pleine expansion dans la santé

Intégration de l'IA au laboratoire

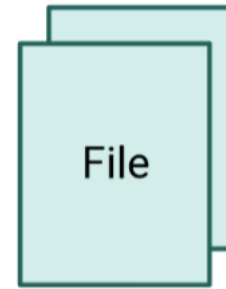
Echantillons



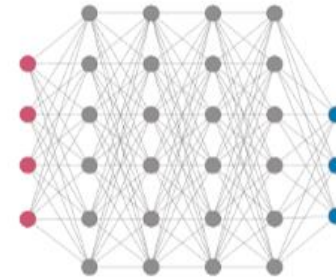
Production des données



Numérisation des données (digitalisation)



Algorithmes d'intelligence artificielle

















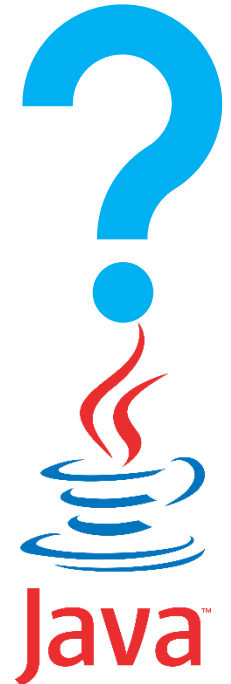
Système d'aide au diagnostic



Quel(s) langage(s) de programmation ? (1)



Oct 2022	Oct 2021	Change	Programming Language	Ratings	Change
1	1		 Python	17.08%	+5.81%
2	2		 C	15.21%	+4.05%
3	3		 Java	12.84%	+2.38%
4	4		 C++	9.92%	+2.42%
5	5		 C#	4.42%	-0.84%
6	6		 Visual Basic	3.95%	-1.29%
7	7		 JavaScript	2.74%	+0.55%
8	10	^	 Assembly language	2.39%	+0.33%
9	9		 PHP	2.04%	-0.06%
10	8	v	 SQL	1.78%	-0.39%
11	12	^	 Go	1.27%	-0.01%
12	14	^	 R	1.22%	+0.03%
13	29	^^	 Objective-C	1.21%	+0.76%
14	13	v	 MATLAB	1.18%	-0.02%



<https://www.tiobe.com/tiobe-index/>

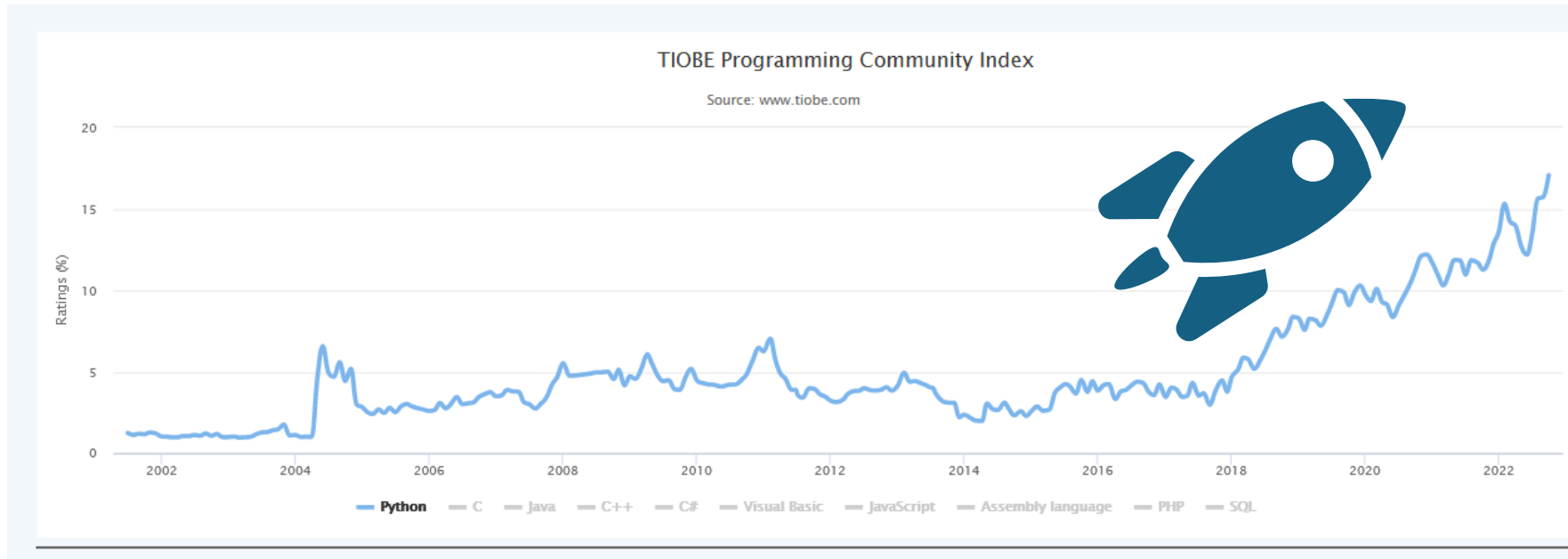
Quel(s) langage(s) de programmation ? (2)

Python

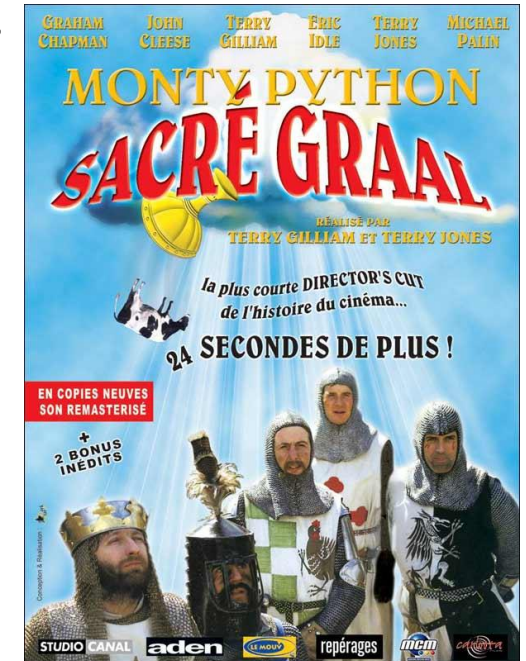
- Créé en 1989 par Guido van Rossum
- Langage multiplateformes
- Langage orienté objet
- Gratuit, simple à prendre en main
- Syntaxe claire, simple
- Utilisation de bibliothèques extérieures (modules) = multiples fonctionnalités



<https://commons.wikimedia.org/wiki/File:Guido-portrait-2014-drc.jpg?uselang=fr>



<https://www.tiobe.com/tiobe-index/>

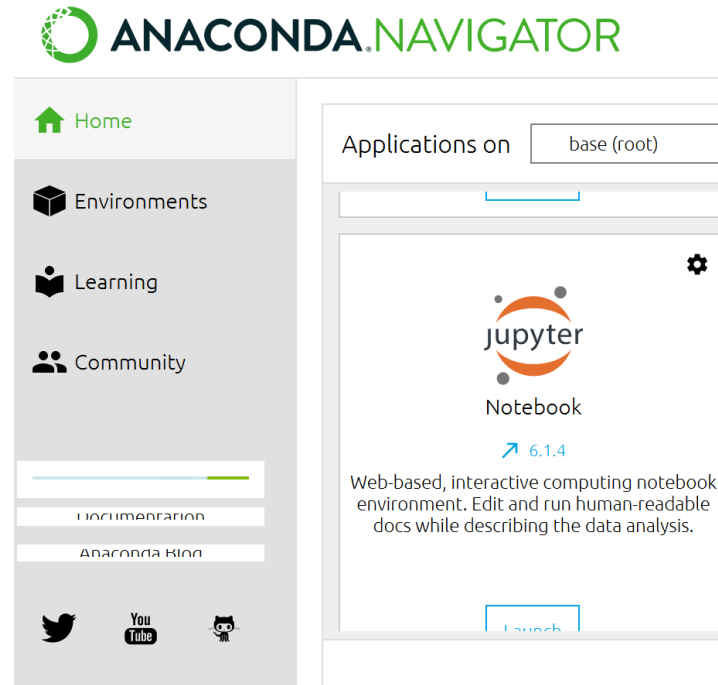


<https://www.allocine.fr/film/fichefilm-248/dvd-blu-ray/?cproduct=9802>

Quel(s) langage(s) de programmation ? (3)

Quelques suggestions pour python :

- Pour l'installation : <https://www.anaconda.com/products/distribution>
- Utilisation de Jupyter notebook pour la création de notebook
 - contient les résultats de l'exécution
 - facilite le partage, la lisibilité, la documentation du code



Quel(s) langage(s) de programmation ? (4)

Quelques suggestions pour python :

jupyter WE ARE 2022 Rennes Dernière Sauvegarde : 26/03/2021 (modifié)

File Edit View Insert Cell Kernel Widgets Help

Exécuter

Exemple de notebook

Entrée [1]: `# Ajouter un commentaire
a = 2+3
print(a)`

5

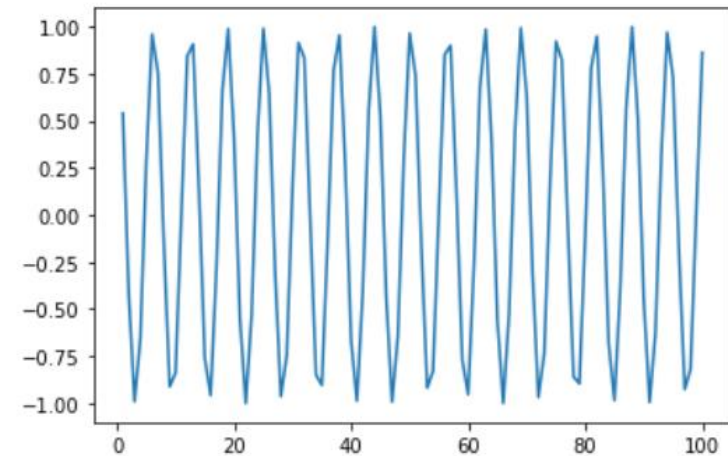
Entrée []: |

Entrée [2]: `# Importer fonctions provenant des modules
import matplotlib.pyplot as plt
import numpy as np`

Entrée [6]: `x = np.linspace(1, 100, 100)
y = np.cos(x)`

Entrée [7]: `plt.plot(x, y)`

Out[7]: [`<matplotlib.lines.Line2D at 0x21840f08520>`]



**Notebook ou cahier électronique
formidable outil d'échange pour l'analyse des données**

L'intelligence artificielle (IA) en 3 définitions

Intelligence artificielle

“The use of computers to perform human-like tasks such as learning, perception and problem solving”

Machine Learning

“Field of study that gives computers the ability to learn without being explicitly programmed”

Deep Learning

“Extension of machine learning which use artificial neuronal network”



Arthur L. Samuel, 1959

<https://cs.stanford.edu/memoriam/professor-arthur-samuel>

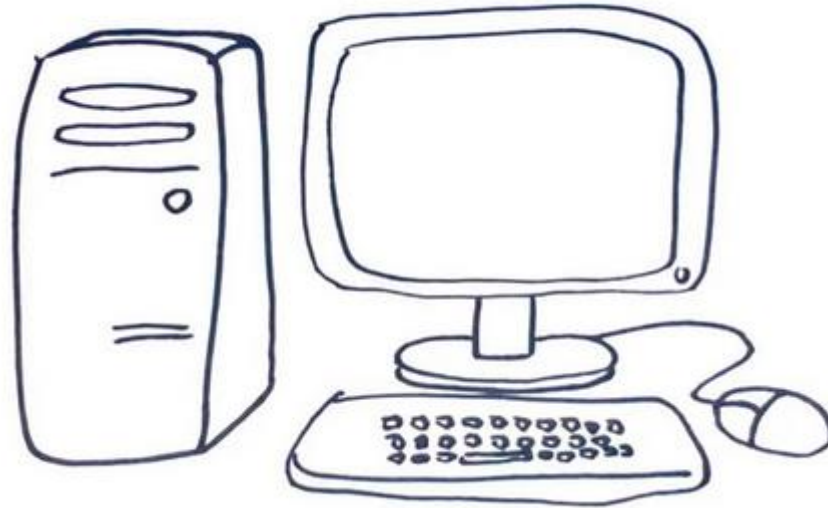
IA et ordinateur

?????

Inputs :

$x = 1$

$y = 2$



Output :

1

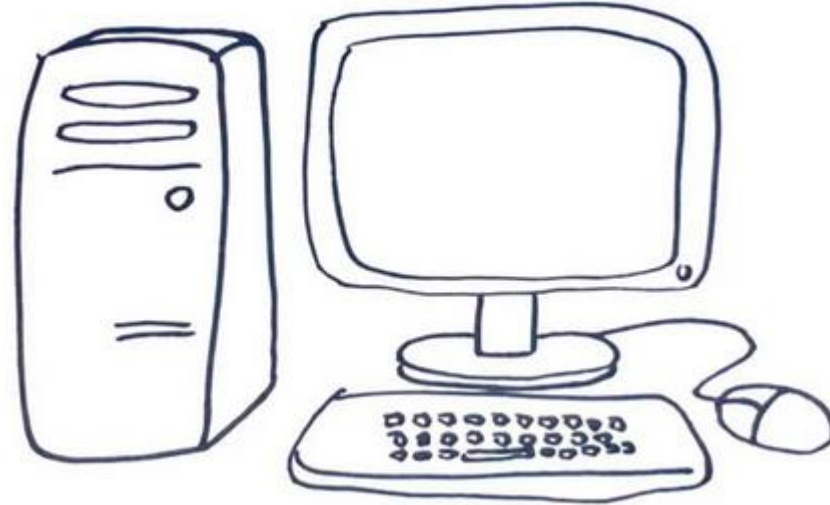
Algorithme

Algorithme

Inputs :

$x = 1$

$y = 2$



Output :
1

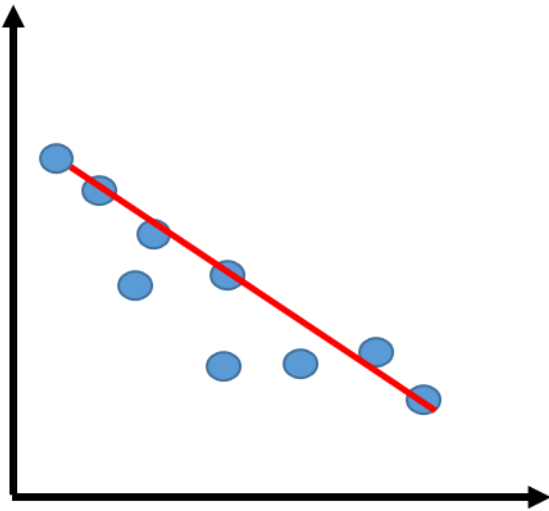
« Ensemble **de règles opératoires** dont l'application permet de **résoudre un problème** énoncé au moyen **d'un nombre fini d'opérations**. Un algorithme peut être traduit, grâce à un langage de programmation, en un programme exécutable par un ordinateur. »

Définition « algorithme », Larousse, 2020

ML, quelques concepts

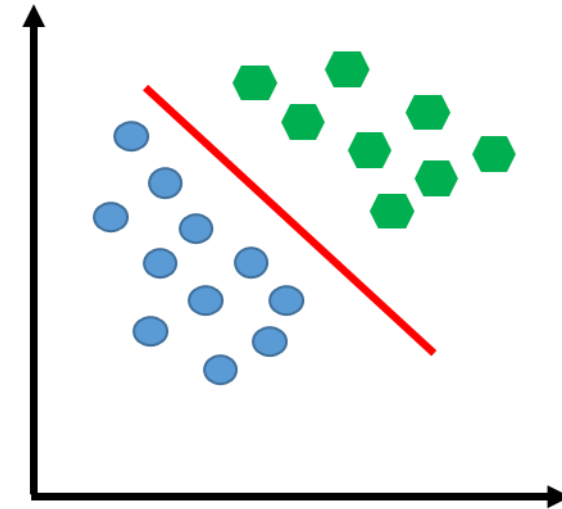
Régression

→ Variable continue



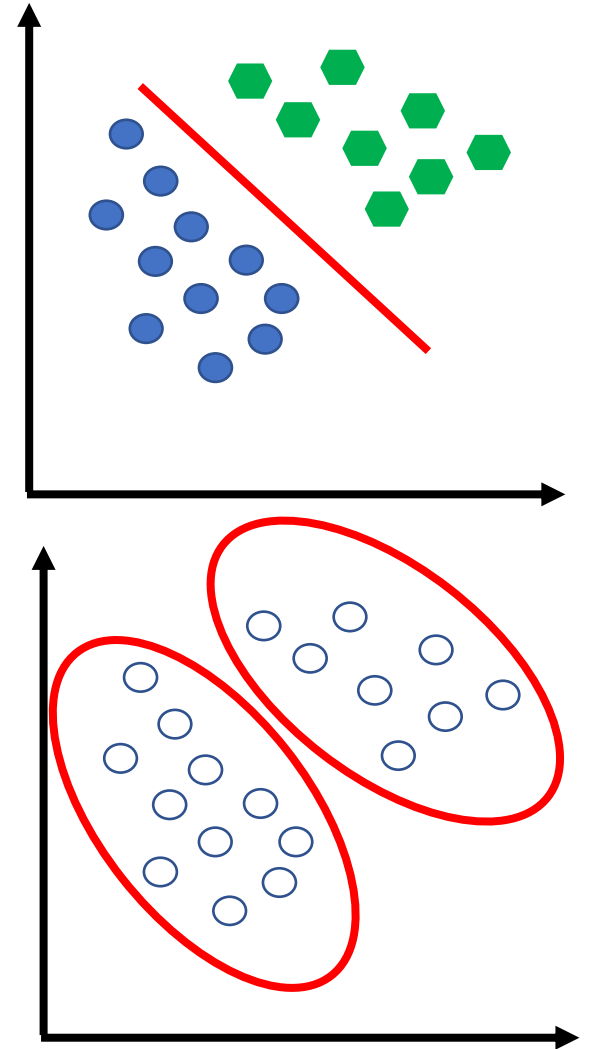
Classification

→ Variable discrète

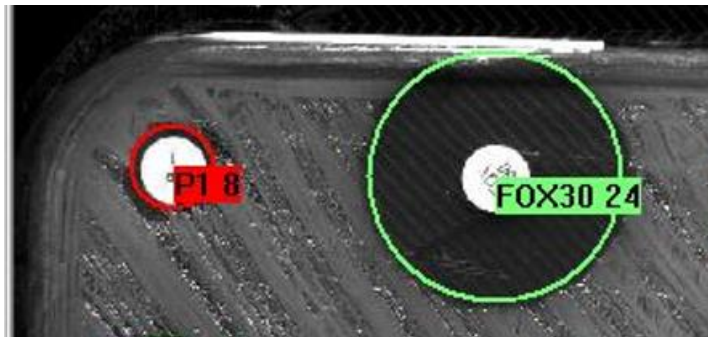
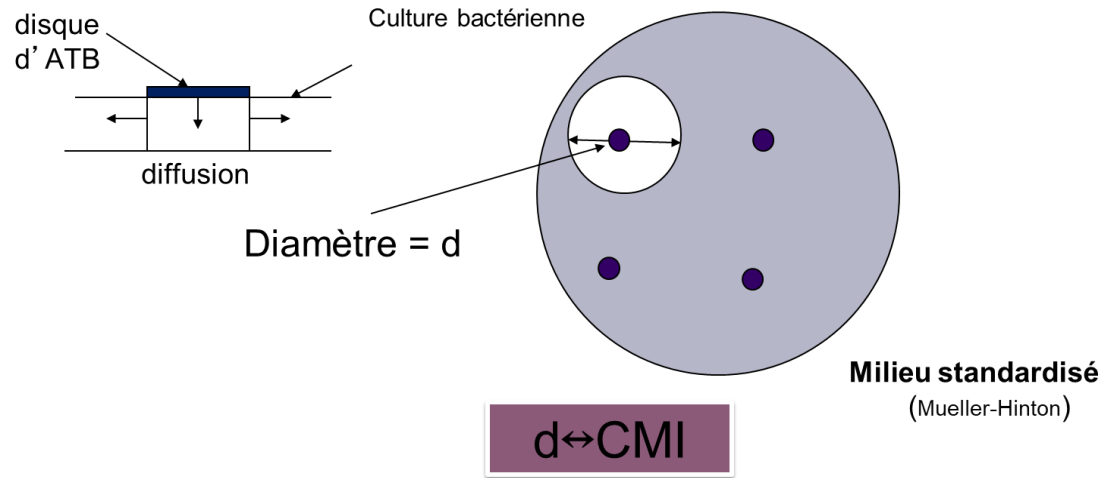


ML, quelques concepts

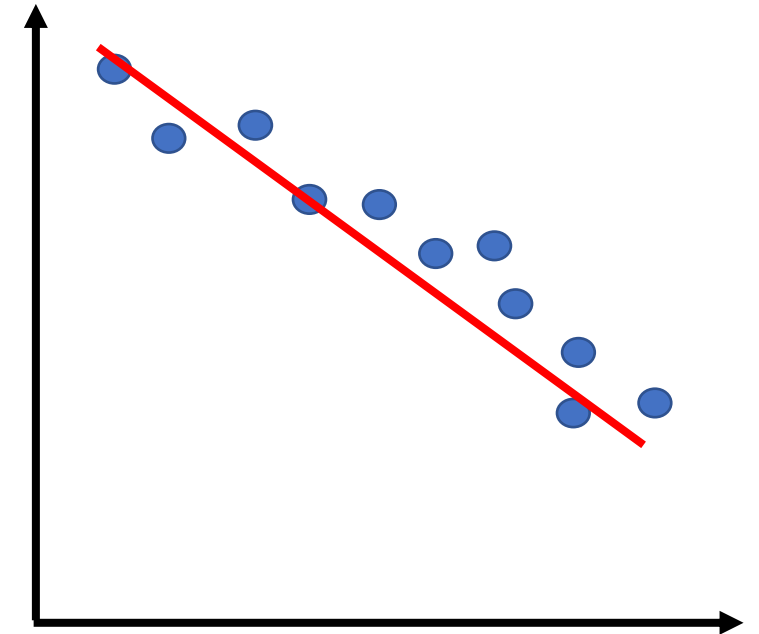
- **Apprentissage supervisé :**
 - apprendre à partir d'exemples étiquetés
 - exemple : réseau de neurones, support vector machine (SVM)
- **Apprentissage non supervisé :**
 - données non étiquetées
 - reconnaissance de structures propres aux données
 - exemple: analyse composante principale (ACP), clustering



Exemple simple : ML et microbiologie (1)



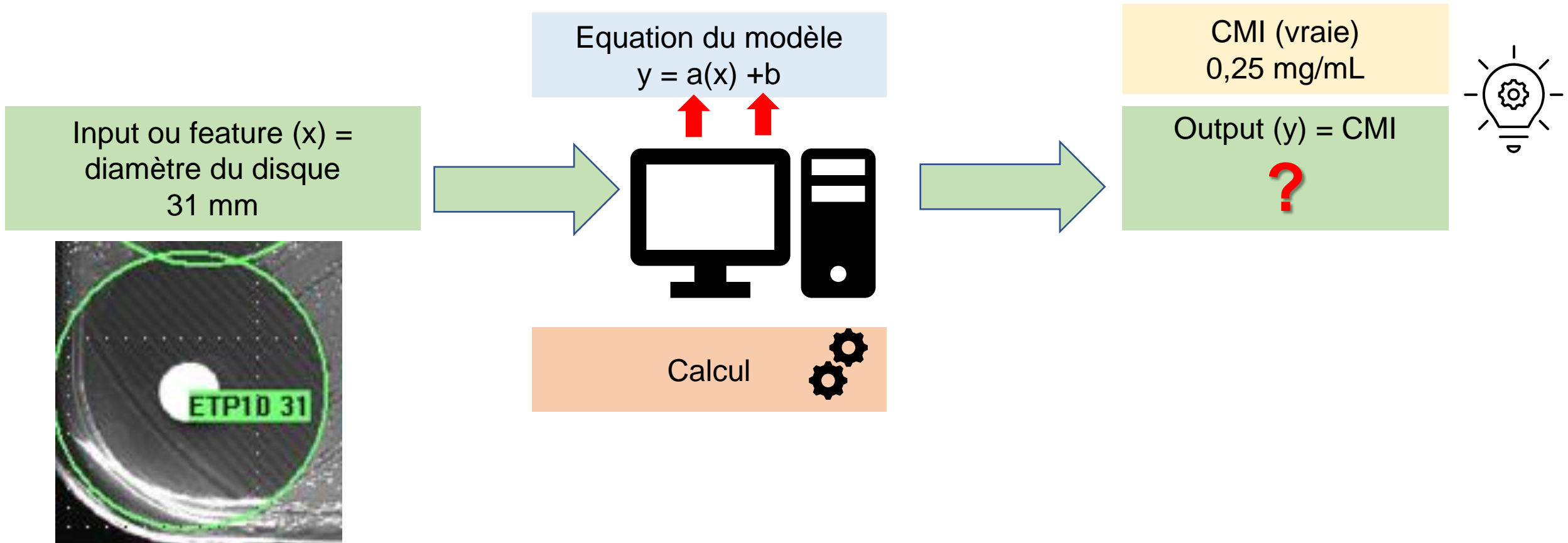
Diamètre



CMI : concentration minimale inhibitrice à partir de laquelle on n'observe plus de pousse visible
Objectif : concentration en antibiotique au site de l'infection $>$ CMI \rightarrow succès thérapeutique

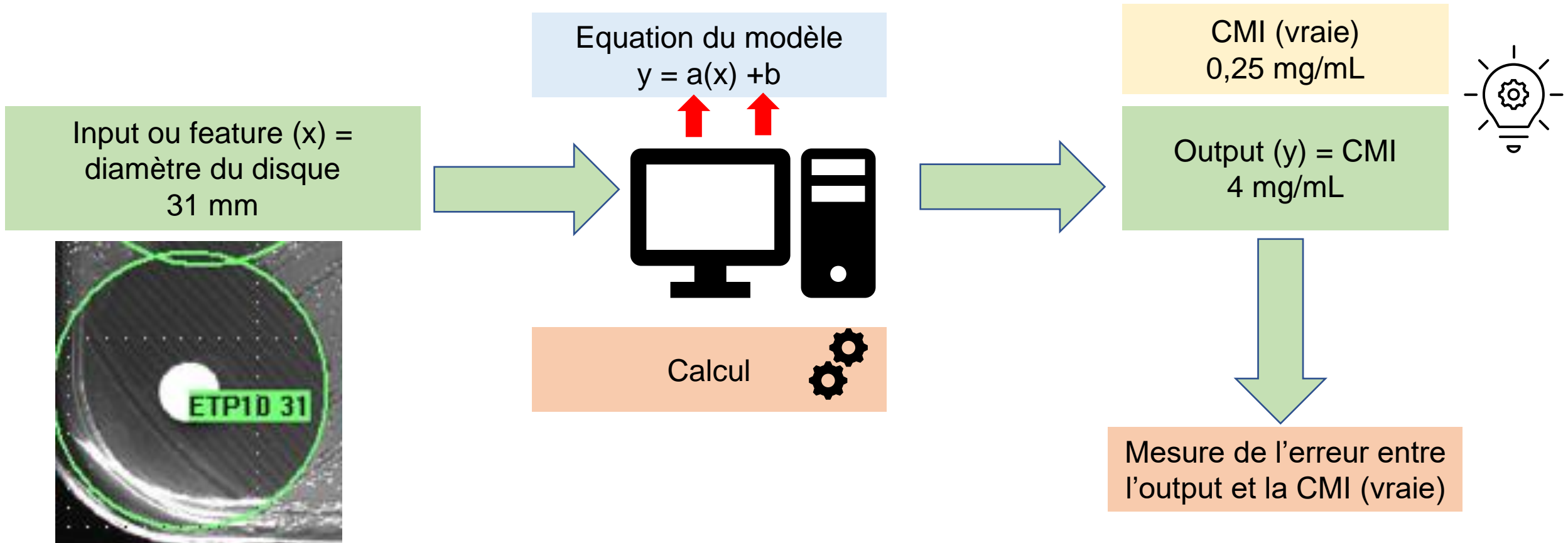
Peut-on estimer la CMI (mg/mL) à partir du diamètre mesuré (mm) ?
Exemple avec IA = apprentissage supervisé et régression

Exemple simple : ML et bactériologie (2)

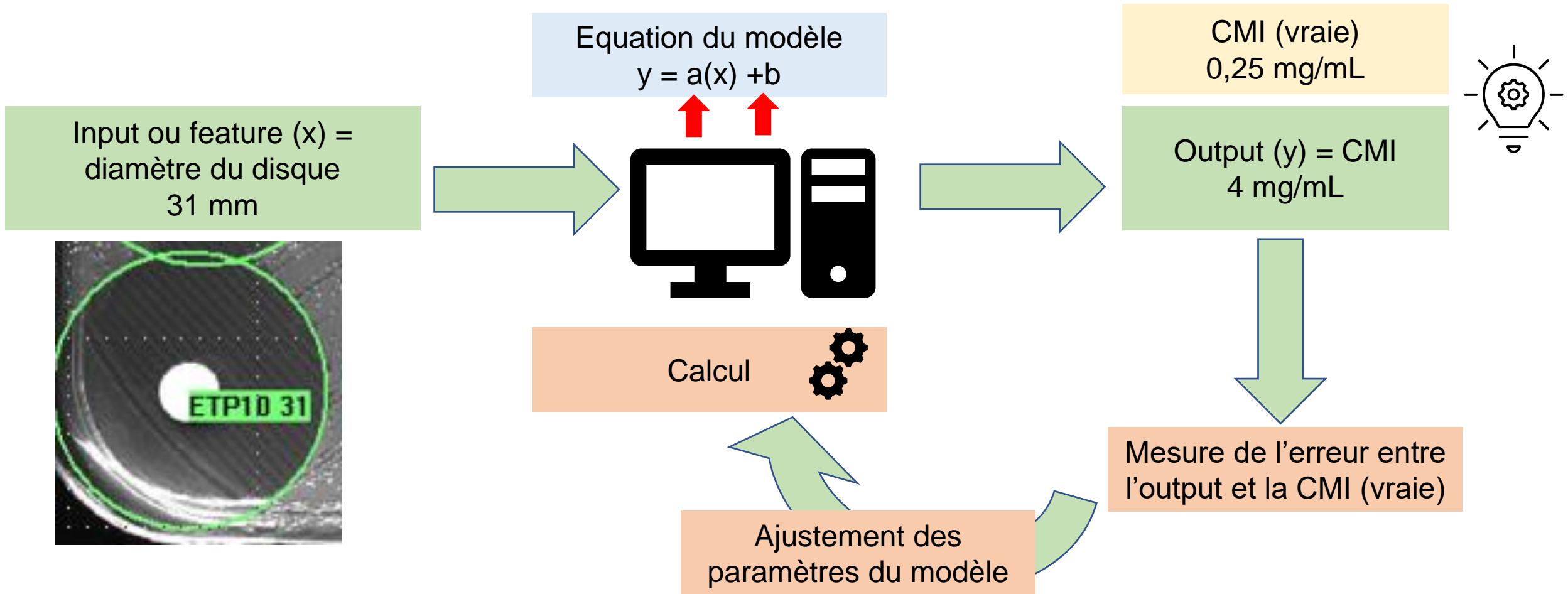


**Il faut déterminer les meilleurs paramètres (a) et (b) de l'équation du modèle
→ entraînement du modèle**

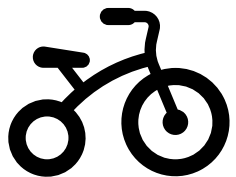
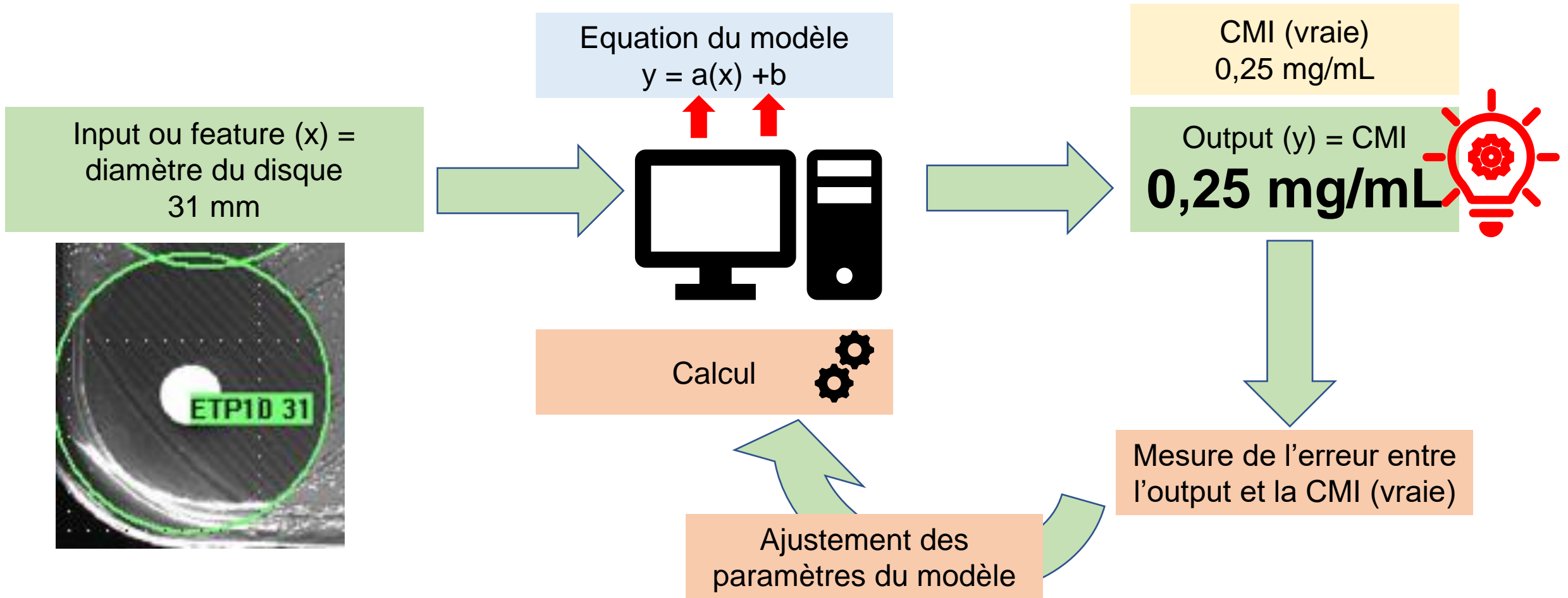
Exemple simple : ML et bactériologie (2)



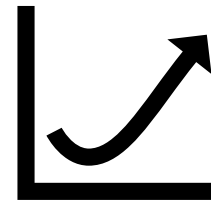
Exemple simple : ML et bactériologie (2)



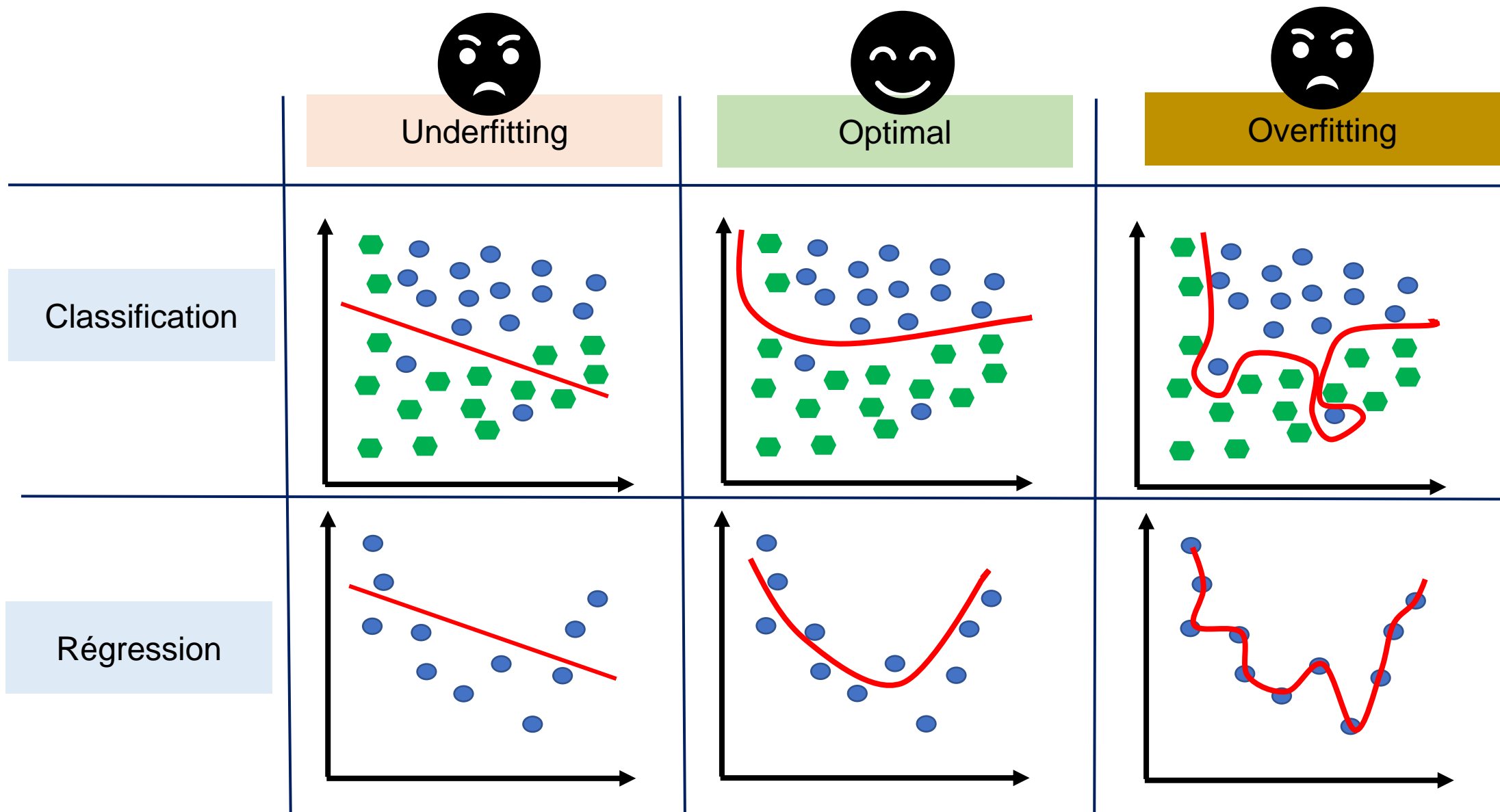
Exemple simple : ML et bactériologie (2)



Performance = tâche × expérience

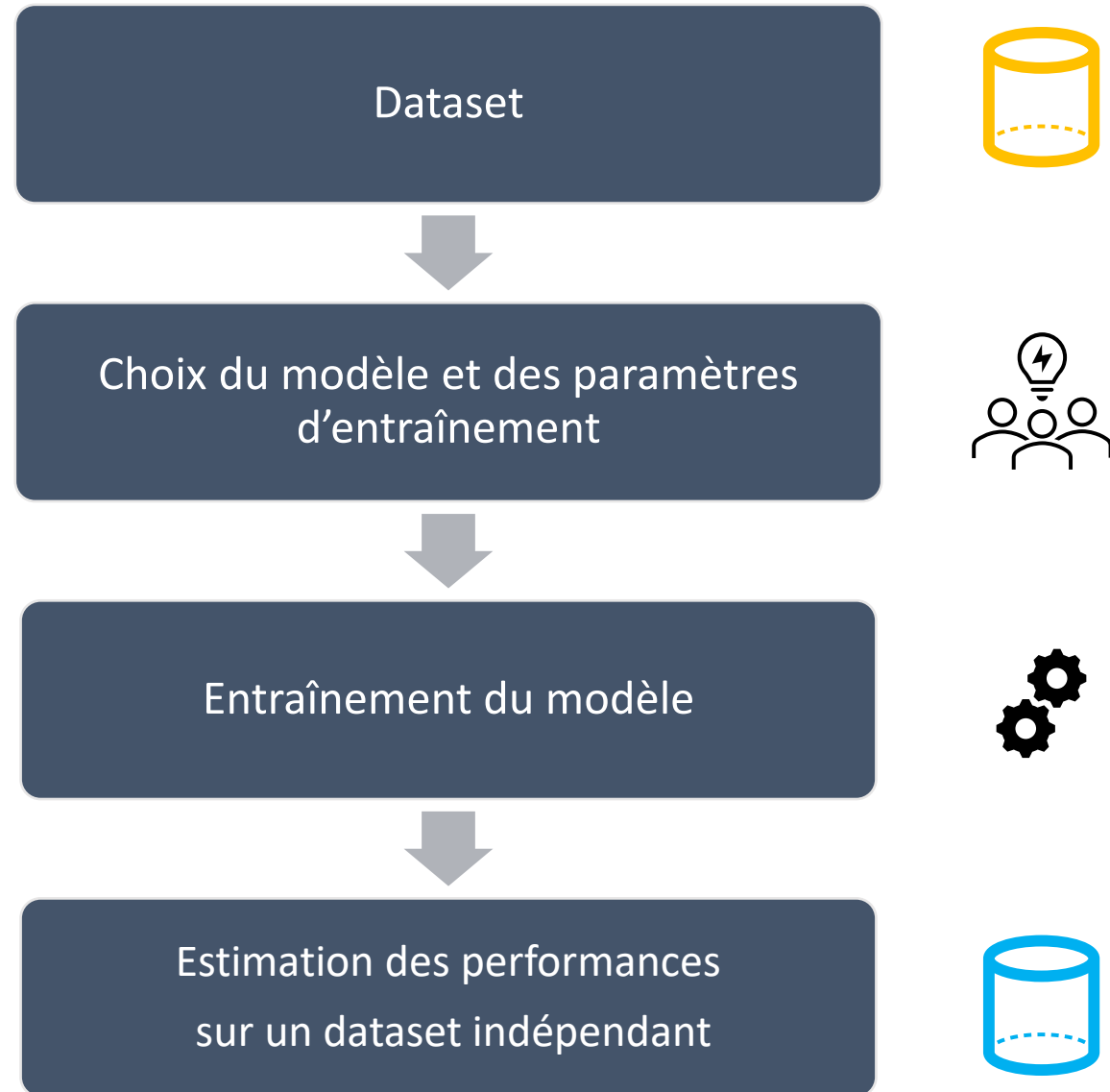


Bon ou mauvais modèle ?



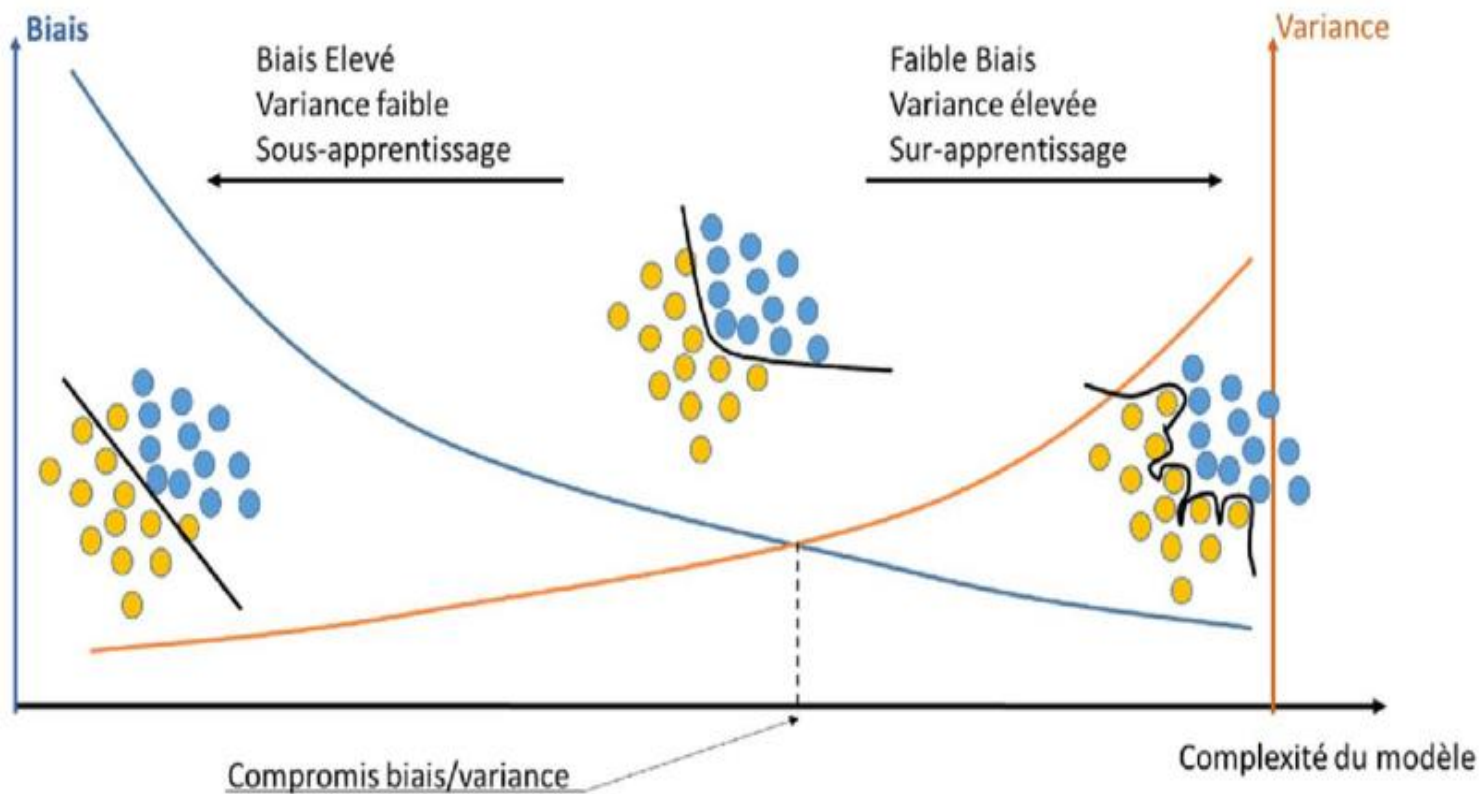
Le bon modèle ? Toujours à vérifier sur un jeu de données externe !

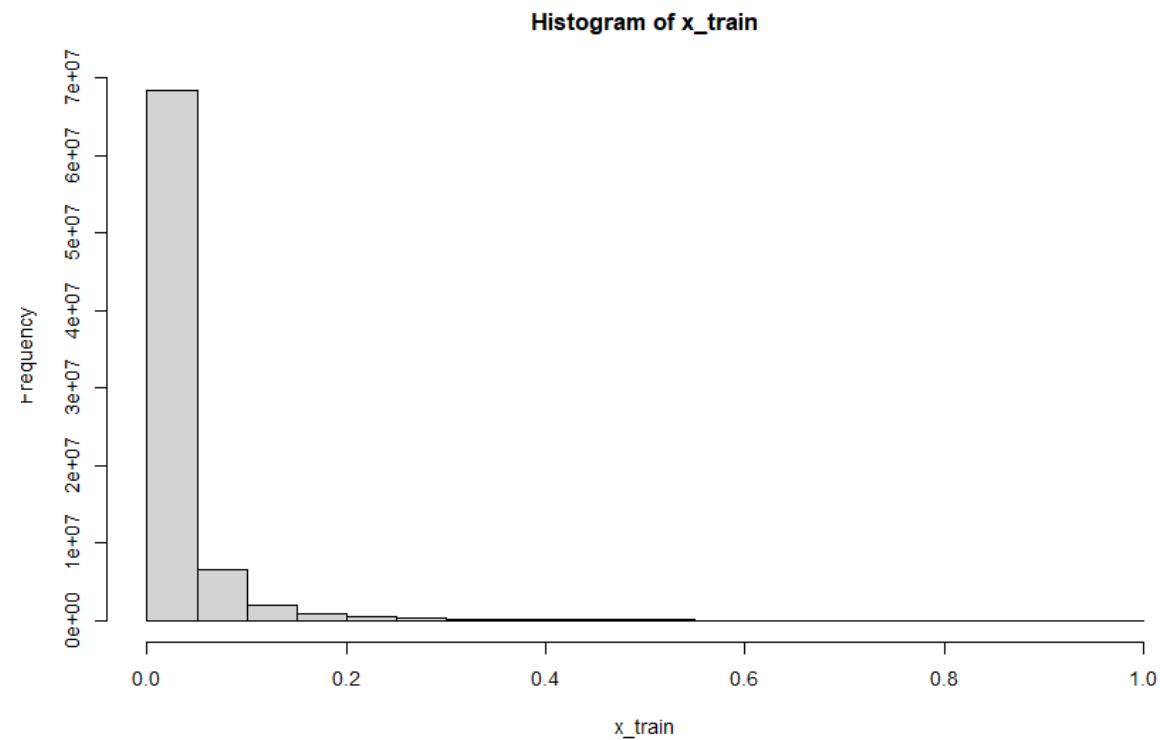
Workflow pour créer un modèle de ML



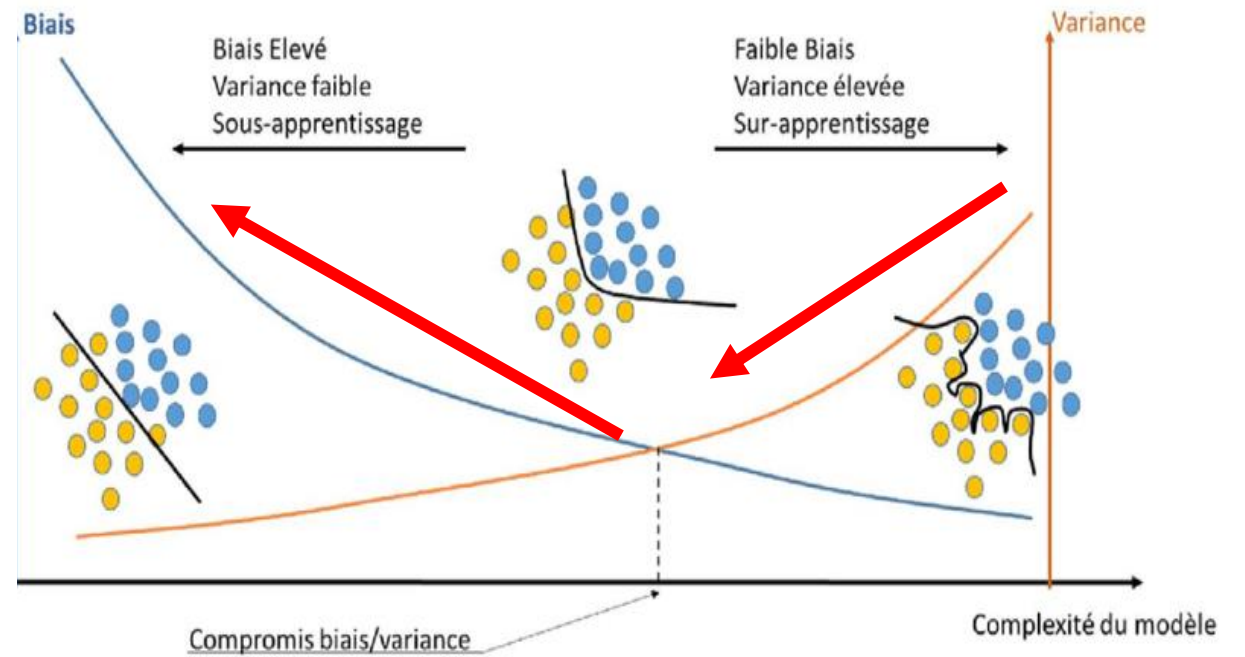
Un biais élevé peut être lié à un algorithme qui manque de relations pertinentes entre les données en entrée et les sorties prévues

Modélisation le bruit aléatoire des données

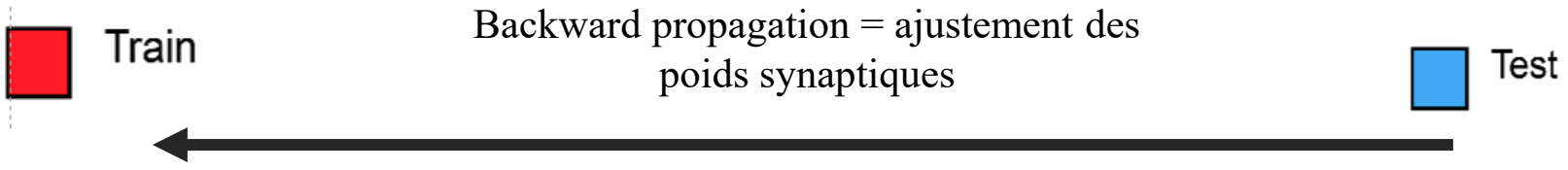




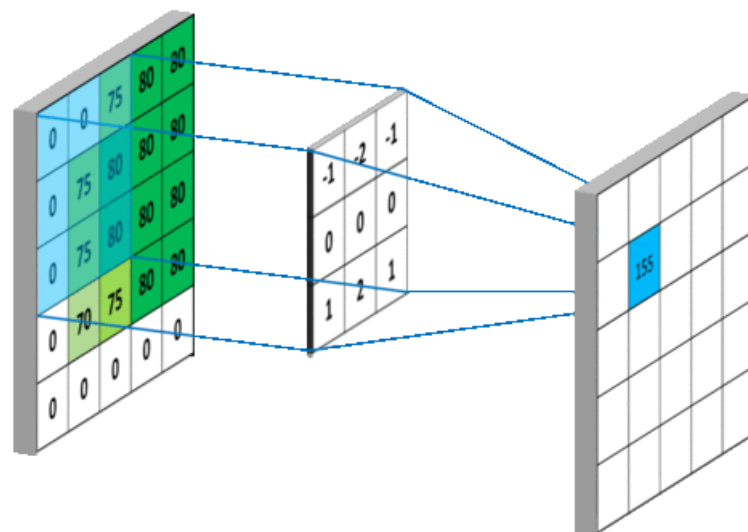
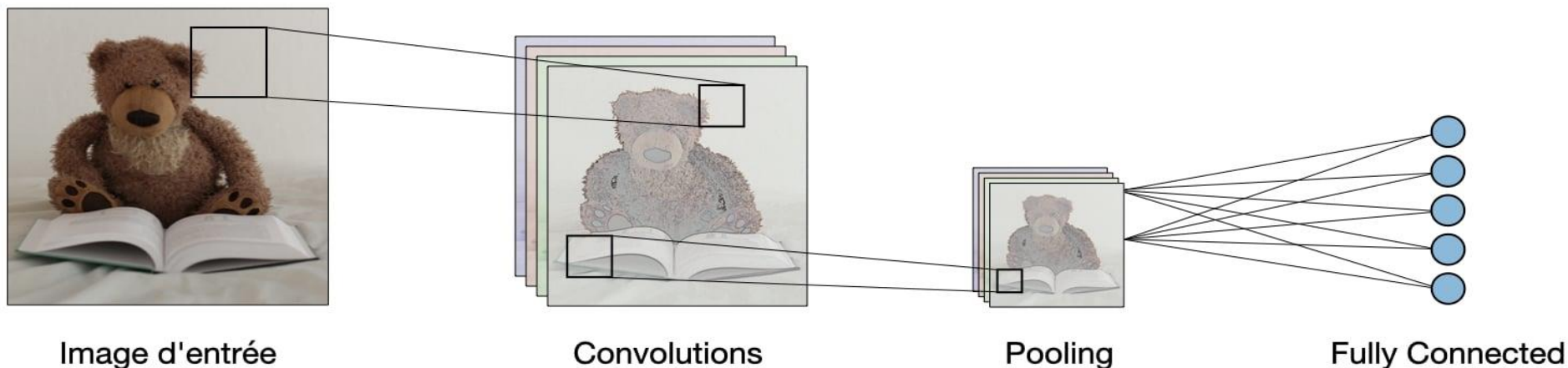
Données bruitées +++



Diminution de la variance
= augmentation du biais



Rétrogradation neuronale = ajustement des poids synaptiques (phase d'entraînement)



Outil de filtrage qui parcourt la totalité de l'image :
« carte de caractéristiques »

<https://stanford.edu/~shervine/tf/teaching/cs-230/pense-bonne-reseaux-neurones-convolutionnels>
extracteur de features : filtration de l'image

Réseaux de neurones à convolution = inspiré du cortex visuel des animaux