



inetum.

Introduction

Intelligence Artificiel

PLAN

- Première introduction :
 - Tour d'horizon de l'intelligence Artificiel
 - Introduction à la Data
- Data et Classification :
 - Approfondissement des problématiques lié à la Data
 - Machine Learning : Problématique de classification
- Machine Learning et Modèle statistique :
 - Machine Learning : Problématique de régression
 - Série temporelle : du modèle Statistique au Deep Learning
- Problématique non-structurée et IA générative:
 - Machine Learning : Problématique de Clustering
 - Détour vers le Reinforcement Learning
 - IA générative : Bonne pratique et usage.

EVALUATION

- Test de placement (Non noté)
- Contrôle continu répartition:
 - Projet machine: 80%
 - Participation: 10%
 - Veille technologique et fonctionnel: 10%
- Examen final :
 - Evaluation écrite des connaissances 2h

01



QUELQUES FONDAMENTAUX

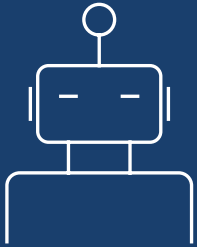
L'IA, C'EST QUOI ?

**L'INTELLIGENCE
ARTIFICIELLE** c'est le fait
d'utiliser les données comme
base de l'apprentissage d'une
action vers une machine, en
vue de la lui faire reproduire
seule par la suite
(et sans utiliser de règles de gestions)

PLUSIEURS TERMES SOUS JACENTS

ARTIFICIAL INTELLIGENCE

Machine "intelligente"



MACHINE LEARNING

Algorithmes apprenants statistiques



DEEP LEARNING

Réseaux de neurone, la nouvelle révolution de l'IA



GENAI

Une « prédiction » générative



SNN

Plus près du cerveau



1950's

1980's

2010's

2020's

?

L'IA N'EST PAS UNE **REVOLUTION** !

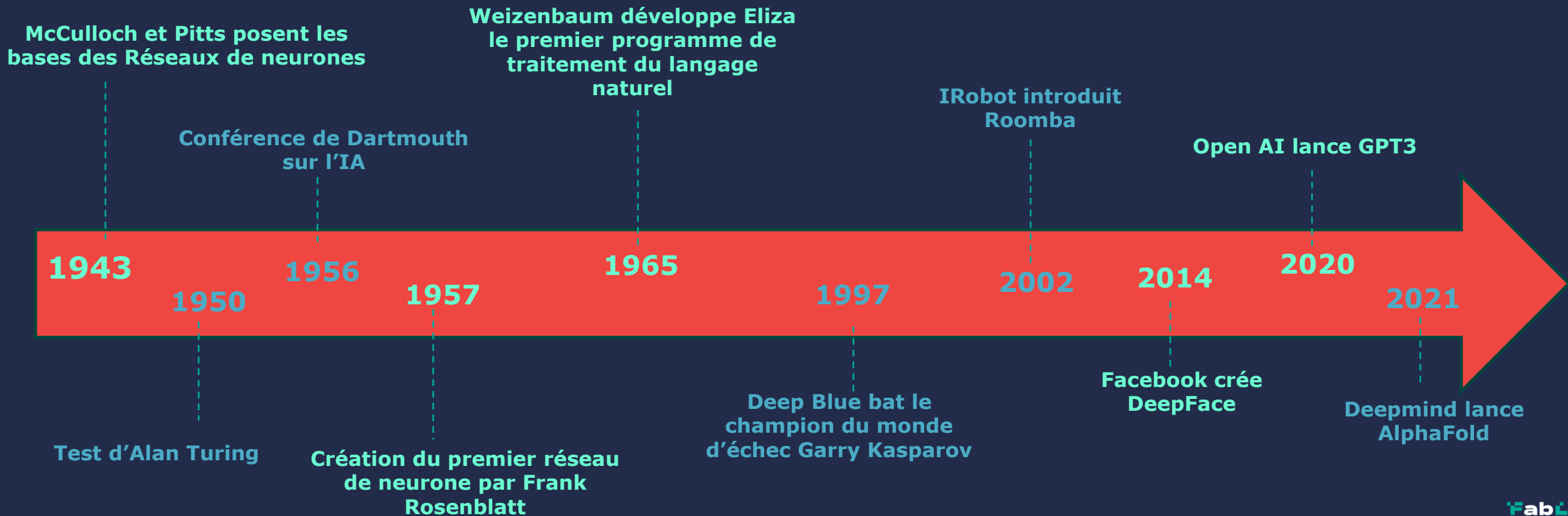


L'IA NE DATE PAS D'HIER

1957, discipline
universitaire aux
Etats Unis



Quelques dates Clées



Une acceleration visible

66,834 model

Grok

72,300 heures A100-GPU
pour un modèle de 12B
de paramètre

GPT 4 : 1,7T de
paramètre

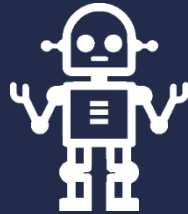
02



L'IA à notre service

ELLE EST DÉJÀ **PARTOUT** !

On a tendance à ne
focaliser que sur les
sujets « **polémiques** » ...



(...)



L'IA POUR QUOI ?

inetum.



Smart Cities

Analyse des flux, gestion des déchets, gestion de la pollution



Loisirs

Recommandations personnalisées



Santé

Diagnostic assisté



Education

Apprentissage personnalisé, Professeur augmenté



Collectivités

Maintenance prédictive, de flux



Télécom

Optimisation du réseau



Véhicules autonomes



Banques et Finances

Détection de fraude, Conseils personnalisés



Marketing

Publicité, Analyse comportement



Distribution

Optimisation des stocks, Maintenance



Industrie

Maintenance, Contrôle qualité



Environnement

Surveillance et réduction de la pollution, Collecte de l'eau de pluie



Energie

Analyse et suivi, Optimisation



Agritech

Agriculture de précision,



Légal / conformité

Analyse de contrats, Surveillance

“ L'utilité perçue de l'intelligence artificielle par types de tâches

Q. Considérez-vous que l'intelligence artificielle pourrait vous être utile ou non pour réaliser les tâches suivantes... ?



Travailleurs du savoir
501 personnes

Oui

Trouver les informations que vous cherchez dans des documents, messages...



76%

Les tâches répétitives



73%

Les tâches administratives basiques



69%

Résumer et récapituler des échanges, des réunions...



66%

Les tâches à faible valeur ajoutée



65%

Rédiger pour vous des messages, des réponses à des mails...



61%



Travailleurs de première ligne
584 personnes

Oui



63%



60%



60%



56%



48%



55%

● Oui, très utile

● Oui, assez utile

● Non, assez peu utile

● Non, pas du tout utile

● NSP

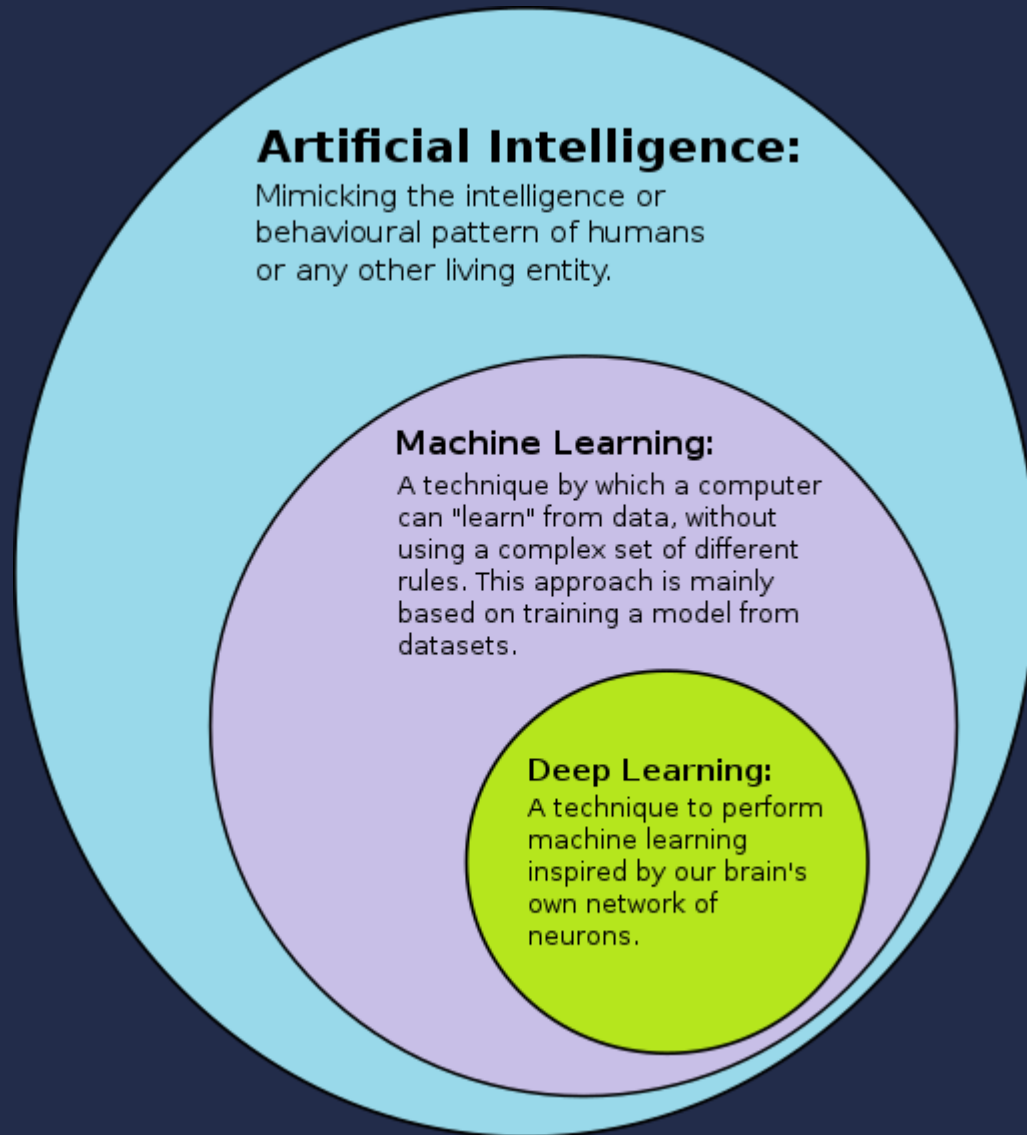
03



L'IA par domaine

IA ? MACHINE LEARNING ? DEEP LEARNING ?

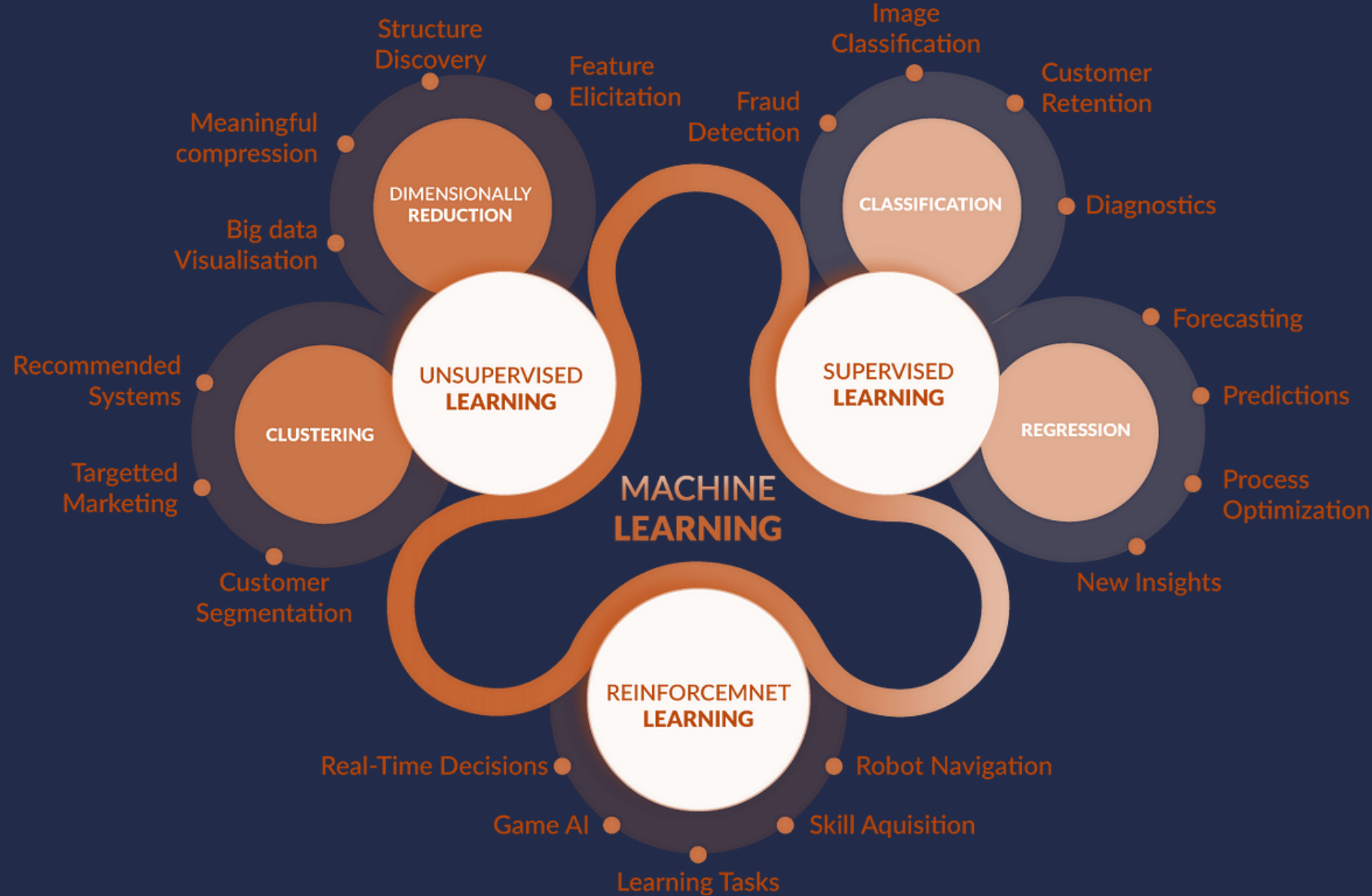
inetum.



Disclaimer :

« Vous utiliser pas de modèle type GPT vous ne faites pas d'IA alors ?! »

MACHINE LEARNING « CLASSIQUE »



Pourquoi utiliser le Deep Learning ?

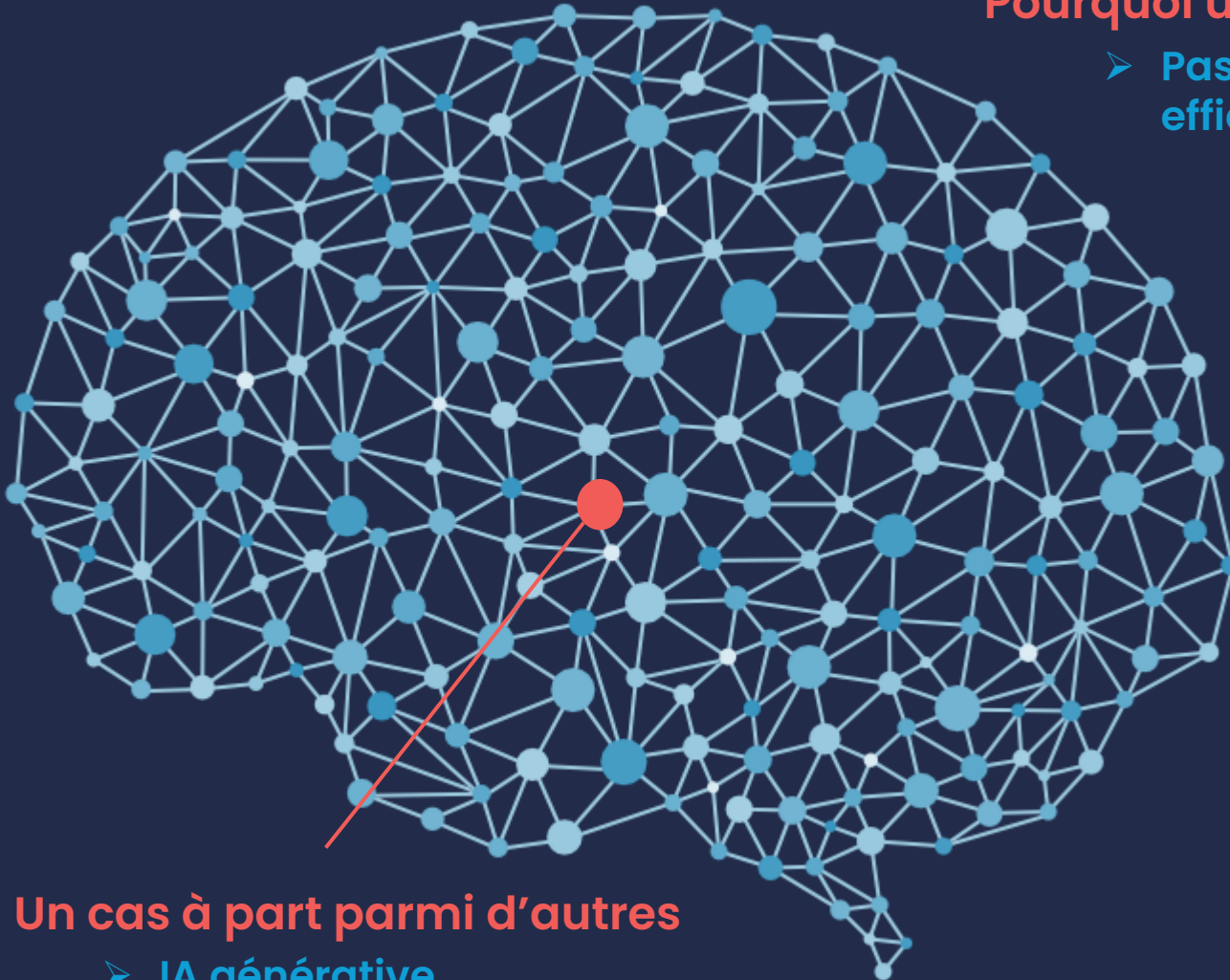
- Pas de solution Machine Learning efficace

Quel cas d'usage ?

- Classification d'image et de vidéo
- Analyse du Langage naturelle
- Problématique de série temporelle
- IA générative

Des limites ?

- Explicabilité limité
- Contrainte de donnée
- Entraînement plus long et coûteux



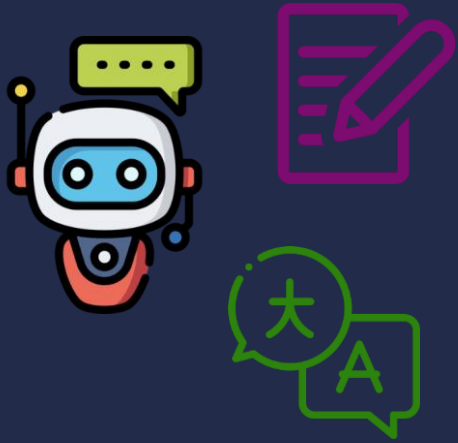
Un cas à part parmi d'autres

- IA générative
- Une architecture à part mais pas une généralité

DEEP LEARNING : QUELQUES CAS D'USAGES

inetum.

Traitement du Langage



Traitement de l'image



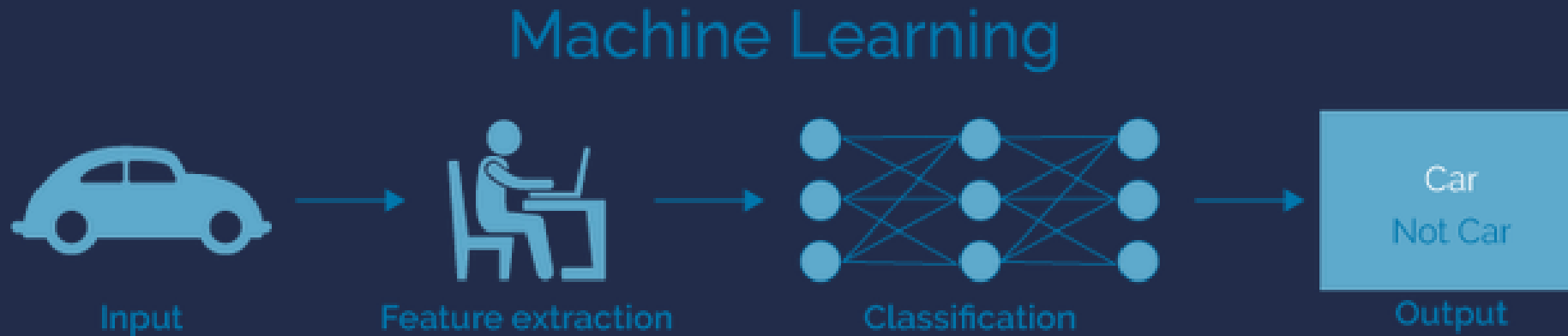
Problématique de série temporelle



IA « Hybride »



DEEP LEARNING X MACHINE LEARNING



04

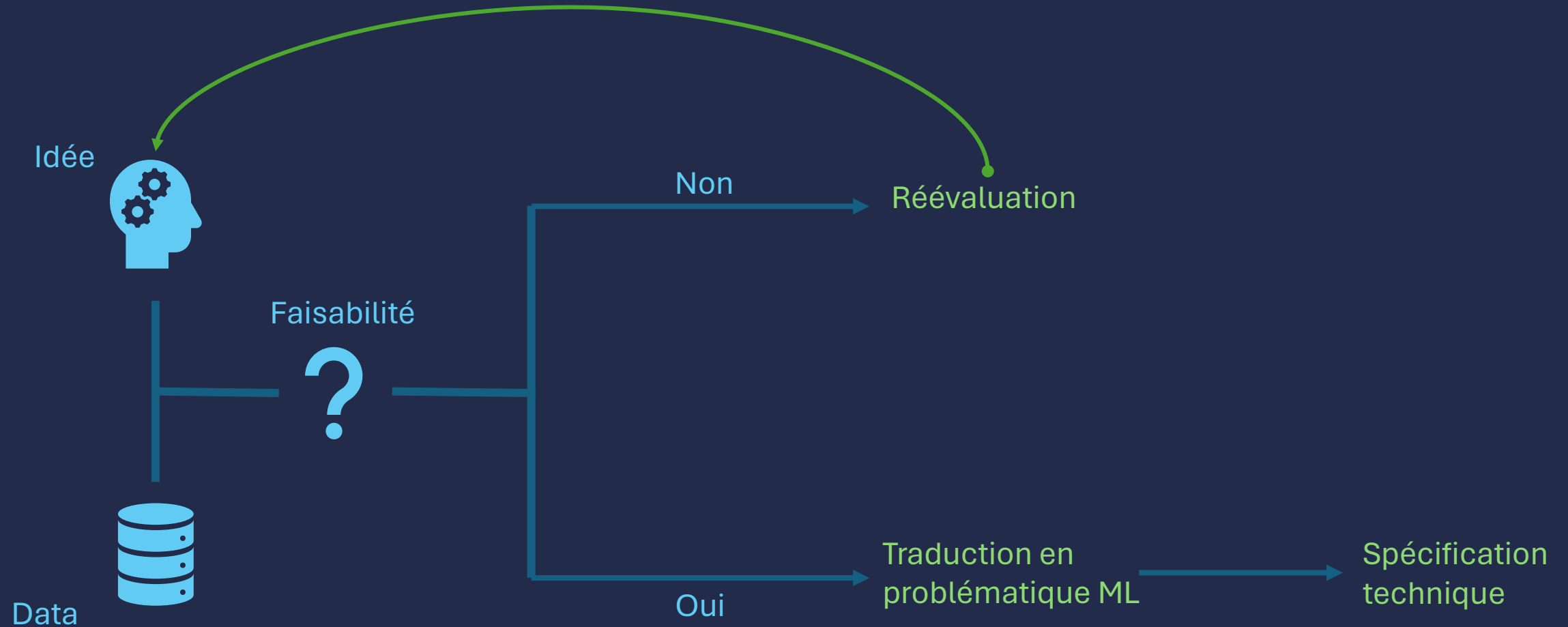


Machine Learning : quelques bases

OVERVIEW : PROCESSUS EXPÉRIMENTAL



PREMIÈRE ÉTAPE : LES BESOINS CLIENTS



SECONDE ÉTAPE: ANALYSE DE LA DONNÉE



Accéder aux sources
de données



Exploration
de la donnée



Premier rapport

TROISIÈME ÉTAPE: MODÉLISATION

Nettoyage de la donnée



On enlève les défauts identifiés dans notre data.

Feature engineering



On ajoute de la valeur à notre data.

Pre-processing



On « traduit » notre data dans un format que le modèle appréhende le mieux.

Choix des modèles



On choisit les modèles les plus adaptés à notre data.

Entraînement



Notre IA apprend

Evaluation

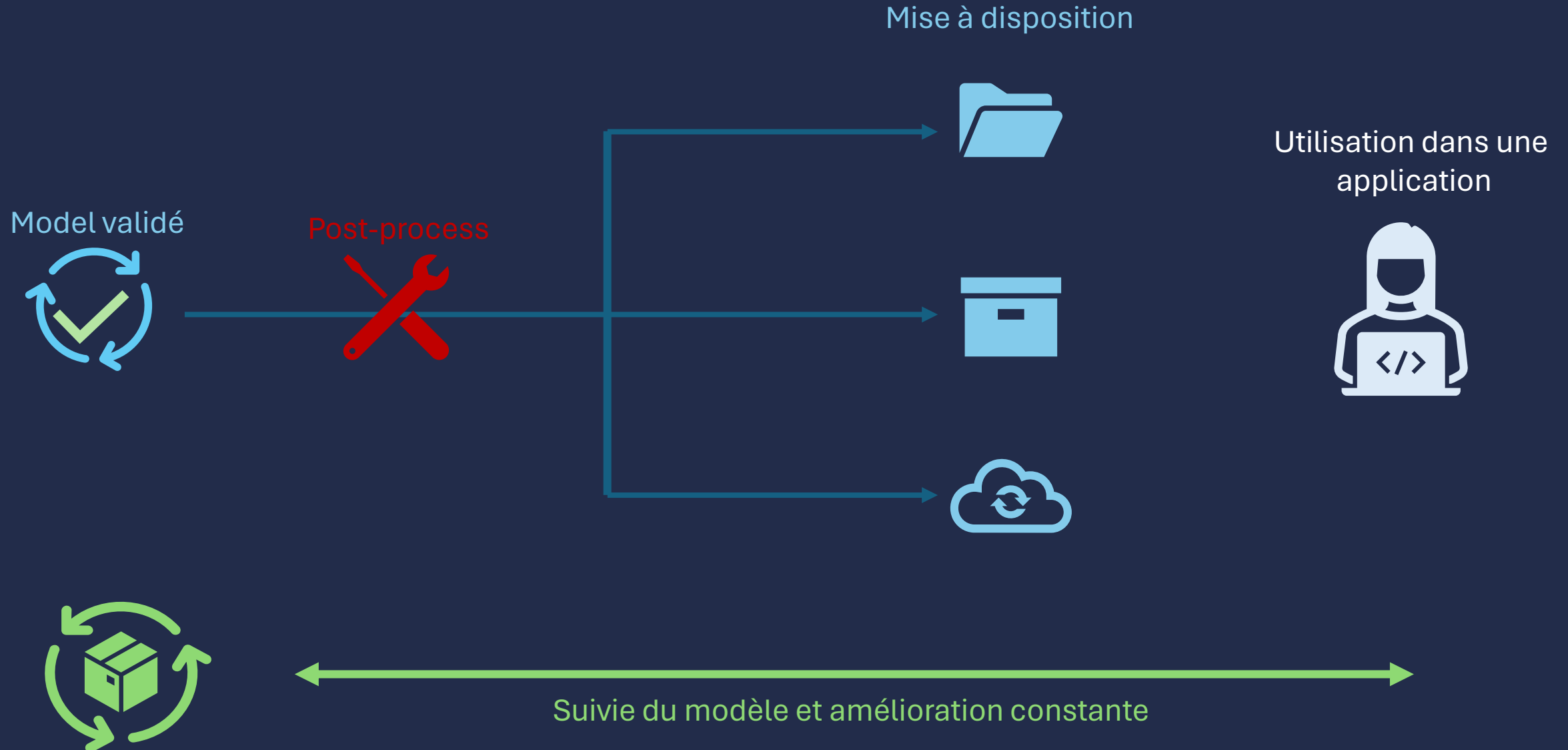


Les résultats sont-ils correct ?

Tuning / optimisation



QUATRIÈME ÉTAPE: INDUSTRIALISATION

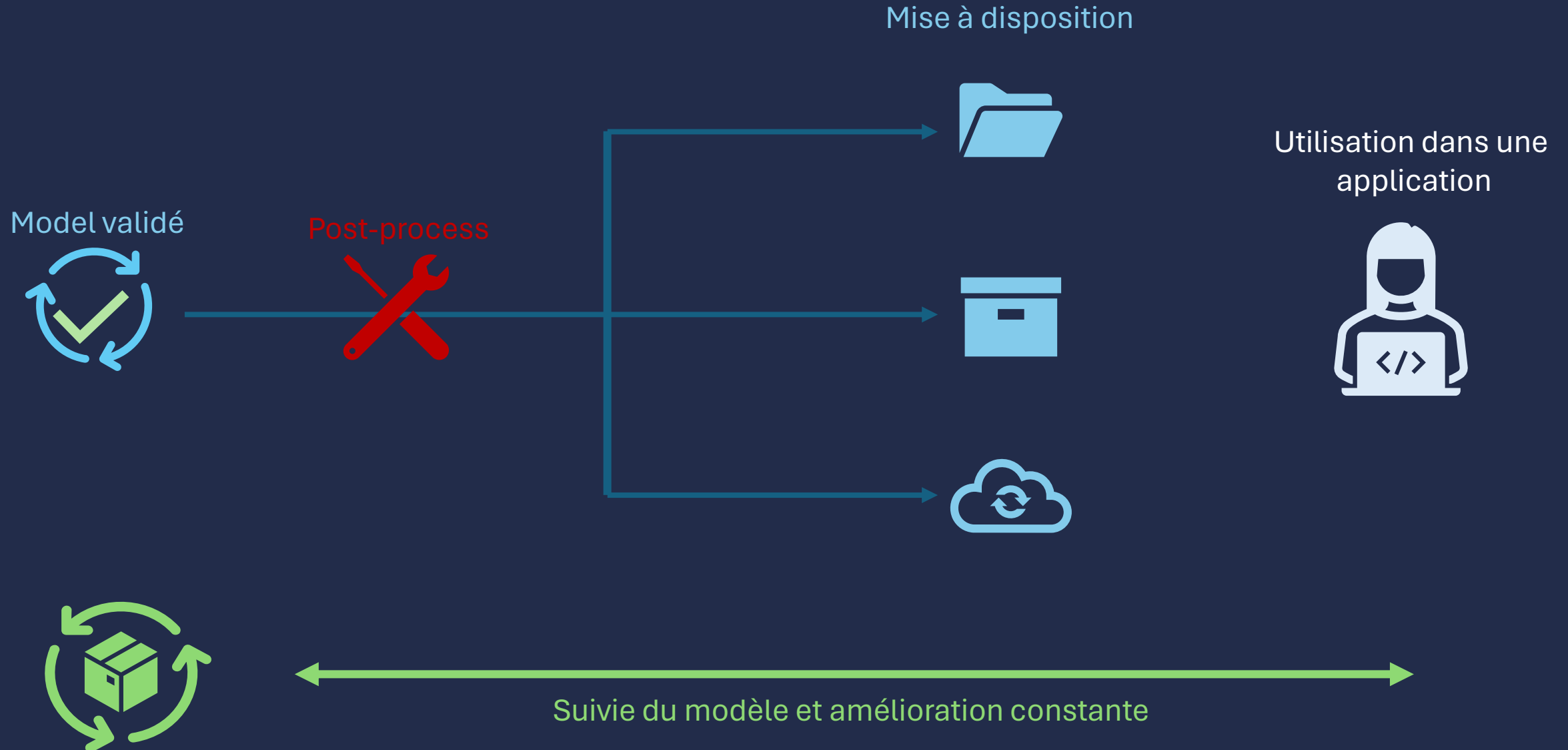




ML Ops

« ML Ops est un ensemble de pratiques qui vise à déployer et maintenir des modèles d'apprentissage automatique en production de manière fiable et efficace. »

QUATRIÈME ÉTAPE: INDUSTRIALISATION



04



Data

LE MOMENTUM DE L'IA

Nouvelles architectures de modèles

Monde académique en ébullition
Sortie de nouvelles approches tous les ans

Accessibilité technique

Nombreux framework disponibles
Nombreuses implémentations
packagées

Puissance de calcul

GPU, TPU, NPU
Neural computer

L'algorithme seul ne peut pas tout ...

 PAS de DATA

PAS d'IA 

 « MAUVAISE » DATA

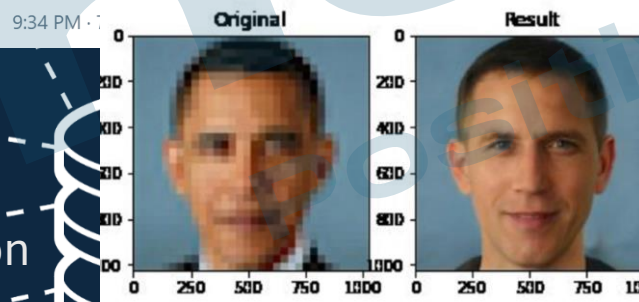
MAUVAISE IA 

6 critères pour un « bon » jeu de données

(et donc une « bonne » IA)

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

The @wife a community-property state, and have been married for a long time. Yet Apple's black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.



Le VOLUME

Est-ce que j'en ai "assez" ?



La VERITE

Est-ce que mon jeu de données représente bien la « vérité » que je souhaite faire porter par mon IA au moment où je l'entraîne ?



La DIVERSITE

Est-ce qu'elles couvrent bien l'ensemble des cas de figure de mon cas d'usages ?



La REPRESENTATIVITE

Est-ce que mon jeu de données est équilibré ?



L'ENVIRONNEMENT de capture

Est-ce que mes jeux de données ont été produits exactement dans les mêmes conditions que l'inférence dont j'ai besoin ?



L'ETHIQUE

Dans quelles conditions mes jeux de données ont été produits ?
D'où viennent les données et qu'est-ce qu'elles concernent ?



LES 4 TYPES DE **DATA** NUMÉRIQUES

**Les
Nombres**

Données
tabulaires



**Les
Images**

Statique /
Dynamique



**Les
Sons**
Vibrations

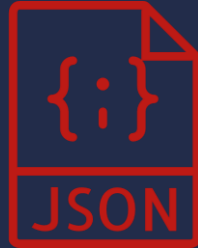


**Les
Mots**
Phrases /
textes

**Multi
Modale**

DONNÉE TABULAIRE (STRUCTURÉE)

SOURCE



Parquet

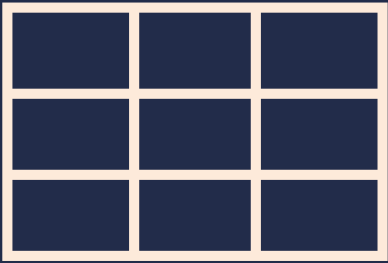


mongo DB

DONNÉE TABULAIRE (STRUCTURÉE)

TYPE DE DONNÉE

Tableau



Numérique



Texte /
catégoriel



Booléen



Date



Autre



Chaque type de donnée va nécessiter une prise en main, des traitements et des visualisations qui leur sont propres

ANALYSE DE LA DONNÉE

GÉNÉRALITÉ

Analyse de forme

- Quelle quantité de donnée ?
- Combien de variable ?
- Combien de ligne ?
- La donnée présente elle des « trous » ? (Nan)
- La donnée a elle la forme attendu ?

Analyse de fond

- A quoi ressemble ma donnée ?
- Réalisation de graphique et visualisation cohérente pour analyser la donnée.
- Analyse univarié
- Analyse multivariée
- Peut-on déceler des valeurs atypique voir aberrante ?

Analyse en corrélation

- Existe-il des corrélations entre nos variables ?
- Ces corrélations sont-elles problématique ?
- Analyse de la target

ANALYSE DE LA DONNÉE

ANALYSE DE FORME

Cette phase nous permet d'avoir une première vision des problèmes structurelles que l'on peut rencontrer dans un projet. Les identifier rapidement permet d'en connaître l'origine et d'éviter de travailler sur une donnée impropre.



Ils manquent certaines variables



Certaines variables n'ont pas la forme désirée



Il manque une partie des données

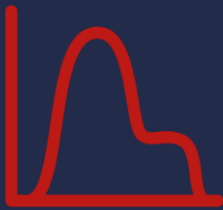
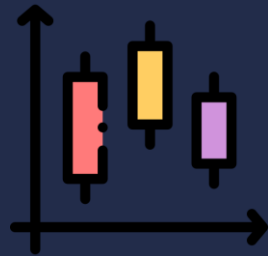
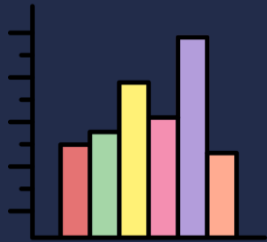


Certaines colonnes sont vides ou présentent une grande part de vide

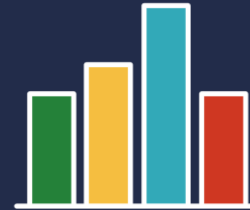
ANALYSE DE LA DONNÉE

ANALYSE DE FOND : UNIVARIÉ

Numérique



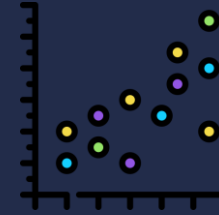
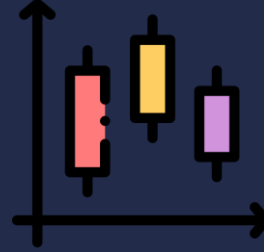
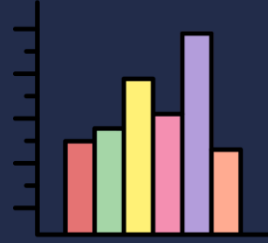
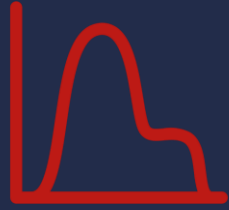
Catégorielle



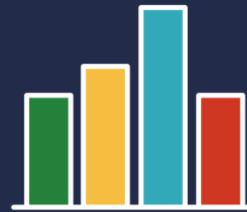
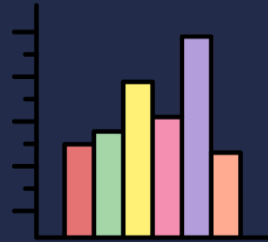
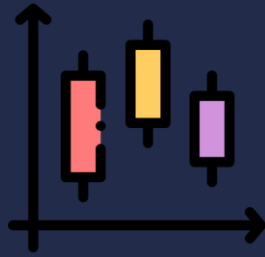
ANALYSE DE LA DONNÉE

ANALYSE DE FOND : MULTIVARIÉ

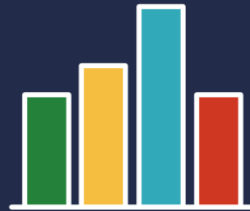
Numérique
X
Numérique



Numérique
X
Catégorielle



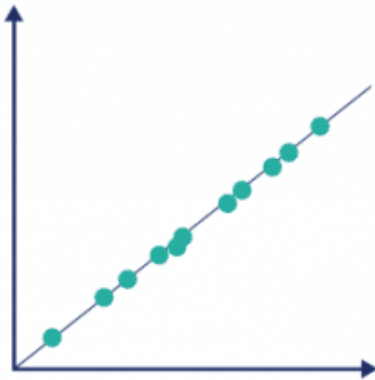
Catégorielle
X
Catégorielle



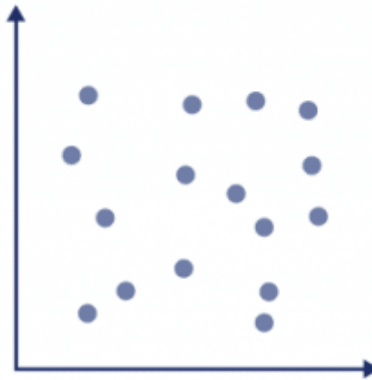
ANALYSE DE LA DONNÉE

ETUDE EN CORRÉLATION

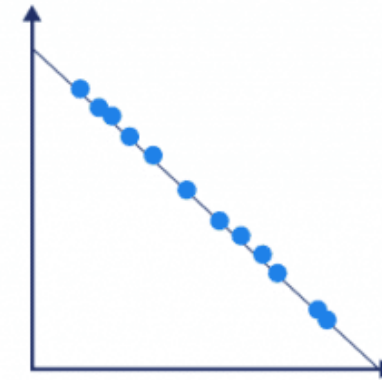
Perfect positive correlation



Zero correlation



Perfect negative correlation



ANALYSE DE LA DONNÉE

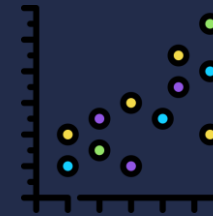
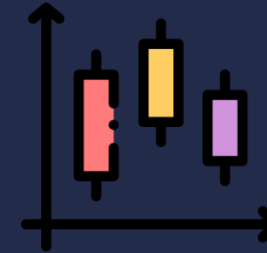
TEST DE CORRÉLATION

En fonction du type de données comparé on va utiliser certains tests statistiques :

- Kendal, Spearman et Pearson pour comparer deux variables numériques entre elles.
- Test d'indépendance du Chi² pour deux variables catégorielles.
- Test d'Anova entre une variable numérique et une variable catégorielle.

Concrètement

-0.24	-0.09	0.04	0.38	1
-0.08	-0.03	0.47	1	0.38
0.28	-0.44	1	0.47	0.04
0.39	1	-0.44	-0.03	-0.09
1	0.38	0.28	-0.08	-0.24



$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$

ANALYSE DE LA DONNÉE

OUTILS

Software



Développement



ANALYSE DE LA DONNÉE

PYTHON : BIBLIOTHÈQUE

Data et statistique



pandas



NumPy



SciPy



GeoPandas

Visualisation



matplotlib



seaborn



plotly



Streamlit



Folium

ANALYSE DE LA DONNÉE

LE DATAFRAME



Attribute 1	Attribute 2	Attribute 3	Attribute 4
...
...	Columns
Rows
...

- Lecture d'une grande variété de source de donnée
- Requêtage « Pseudo-SQL » permettant la plupart des traitements de donnée.
- Traitement des données optimisée sur des vecteurs.
- Visualisation sommaire avec les librairies de visualisation.
- Objet compatible avec la plupart des librairies de data science et de visualisation.

UN PEU DE PRATIQUE

