

■

Webinaire Les données Du Cauchemar à l'or Noir

16 Mai
2023

Comment Structurer & Valoriser vos données ?



Alia Santé

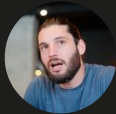
Qui sommes nous ?



Segolène Combettes - R&D Directrice
segolene@alia-sante.com



Yann Peter - Responsable Innovation et Développement des Entreprises
Yann.PETER@castres-mazamet.com



Sylvain Françoise Dal-Mas - President
Sylvain.fd@alia-sante.com



Issu d'un regroupement de chercheurs du tissu académique Toulousain



Docteurs en Mathématique
Nous fédérons des docteurs en mathématique pour comprendre l'intelligence artificielle



Expert en Intelligence artificielle
L'expertise de Alia Santé est focalisé sur l'étude et la création de modèles d'intelligence artificielle



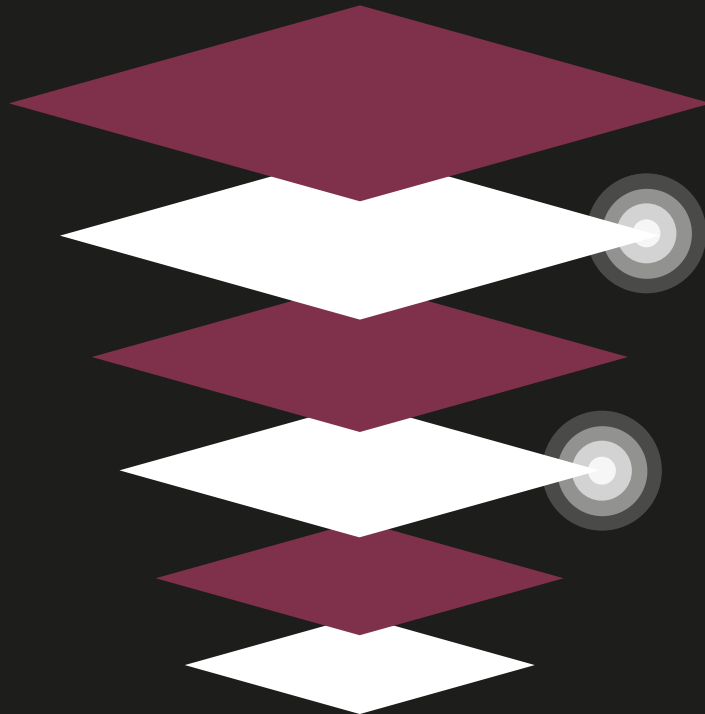
Déroulement du Webinaire



■ Parcourons ensemble la route de la data science appliquée à vos projets.

Les étapes projet en IA

Les étapes clefs de la création d'un modèle d'intelligence artificielle



■ LA PROBLÉMATIQUE

■ STRATÉGIE DE
COLLECTE & STOCKAGE

■ QUALIFICATION

■ TRAITEMENT

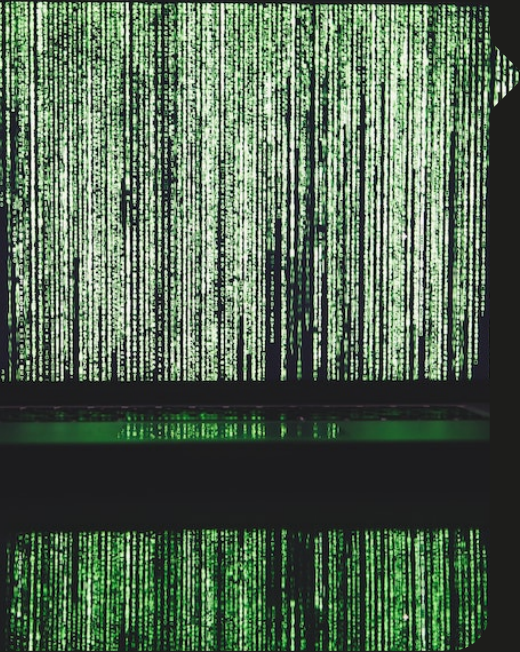
■ ENTRAÎNEMENT

■ DÉPLOIEMENT

Les données sont la pièce maîtresse des algorithmes d'intelligence artificielle. Elles permettent à l'IA de comprendre et d'apprendre la manière de traiter des informations. La performance de l'IA dépend donc de la disponibilité, de la quantité, et de la qualité des données.

Donnée.

Entité numérique sur une information quelconque.



Elle peut prendre de **multiple formes**.

On distingue principalement deux modèles de données ou variables :

Les données **quantitatives** et les données **qualitatives**.

Un jeu de données est un ensemble de données portant sur des **entités de même nature**.



Donnée.

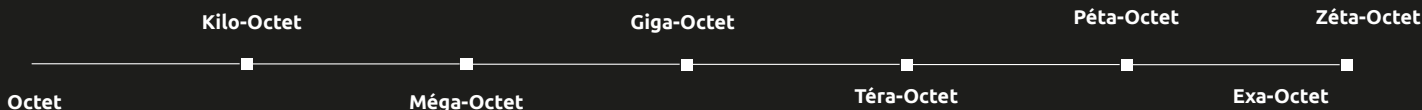
Une croissance exponentielle.

Selon les dernières estimations, le volume de données numériques créées ou répliquées à l'échelle mondiale a été multiplié par plus de trente au cours de la dernière décennie, passant de 2 zettaoctets en 2010 à 64 zettaoctets en 2020.

56

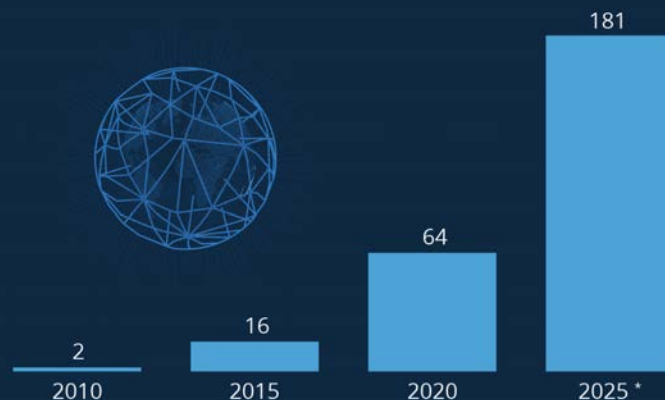
milliards d'euros

Chiffre d'affaires en 2020 du marché du numérique.



Le Big Bang du Big Data

Estimation du volume de données numériques créées ou répliquées par an dans le monde, en zettaoctets



Un zettaoctet équivaut à mille milliards de gigaoctets.

* Prévvision en date de mars 2021.

Sources : IDC, Seagate, Statista



statista

Les acteurs de régulation

Régissent les règles et cadre d'utilisation de la donnée.



Autorité administrative indépendante française. La CNIL est chargée de veiller à ce que l'informatique soit au service du citoyen et qu'elle ne porte atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

Règlement Général sur la Protection des Données.

Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne.

- Continuité de la Loi française Informatique et Libertés de 1978
- Renforce le contrôle par les citoyens de l'utilisation de leurs données
- Harmonise les règles en Europe



Qu'est ce qu'une donnée personnelle ?

Une donnée se rapportant à une personne physique identifiée ou identifiable.

Une personne physique peut être identifiée :



Directement

Nom et prénom



Indirectement

Numéro de téléphone, le numéro de sécurité sociale, une adresse mais aussi la voix ou l'image.

07

L'identification d'une personne physique peut être réalisée :



A partir du **croisement d'un ensemble de données** (exemple : une femme vivant à telle adresse, née tel jour et membre dans telle association).



A partir d'une seule donnée (exemple : nom)

Les données à caractère personnel concernant la santé sont les données relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique qui révèlent des informations sur l'état de santé de cette personne.

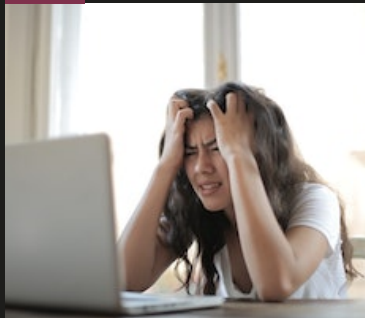
Comment établir une stratégie de collecte de données ?

Savoir collecter de la donnée permettant la mise en place du modèle IA.

Définir la problématique pour identifier la solution

01

Identifier, comprendre et synthétiser la problématique

**02**

Analyser la stratégie et les méthodes de collectes

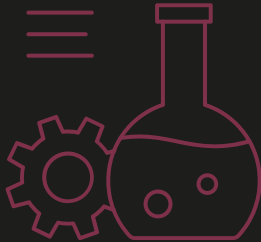
**03**

Vérifier la conformité réglementaire



Comment établir une stratégie de collecte de données ?

Information à déterminer au préalable



Une collecte de données vraiment nécessaire

- Quels sont les objectifs de la collecte ?
- A-t-on le droit de collecter ces données ?
- Les données sont-elles pertinentes à l'utilisation recherchée ?
- Faut-il un consentement ?
- D'où proviennent les données personnelles collectées ?



Respect des droits des personnes

Les droits d'accès, modification, suppression, portabilité doivent être respectés selon les principes du RGPD

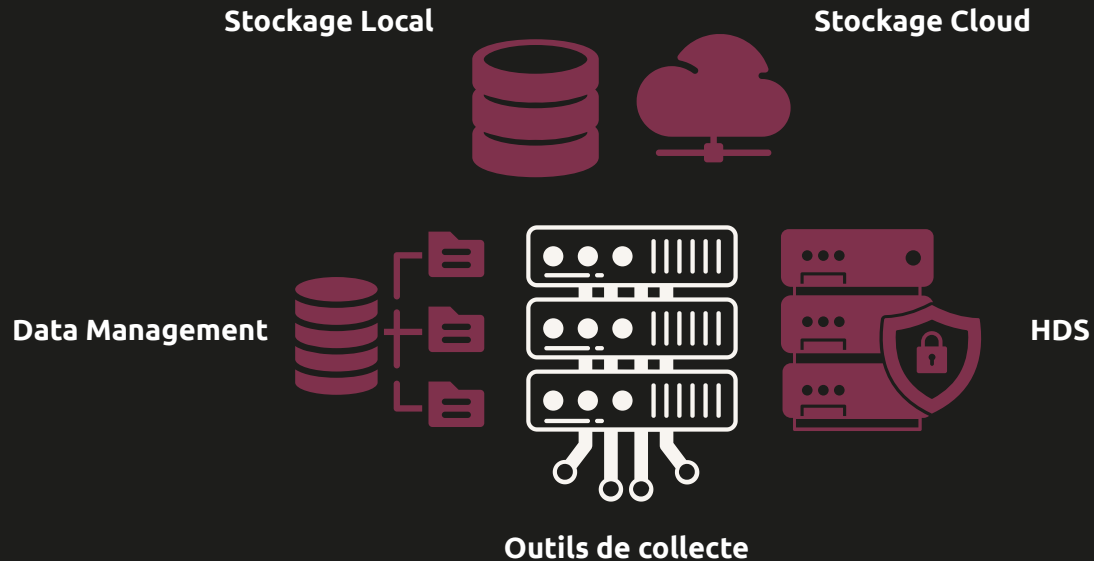


La transparence Une bonne information des personnes = confiance

Le respect de ces réflexes permet d'être conformité avec les exigences du RGPD mais également d'utiliser des systèmes d'IA éthiques, respectueux des droits des personnes et des principes européens de protection des données personnelles.

Comment stocker de la donnée ?

La collecte de la donnée nécessite la sécurisation et l'organisation du système de stockage.

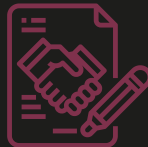


Comment stocker de la donnée ?

Le cadre réglementaire du stockage de la donnée



LICÉITÉ
LOYAUTÉ
TRANSPARENCE



DES FINALITÉS
DÉTERMINÉES
EXPLICITES
LÉGITIMES



DONNÉES
ADÉQUATES
PERTINENTES
LIMITÉES A
L'USAGE

EXACTITUDE
MISE A JOUR DES
DONNÉES



LIMITATION DE
LA DURÉE DE
CONSERVATION
DES DONNÉES



INTÉGRITÉ ET
CONFIDENTIALITÉ
DES DONNÉES



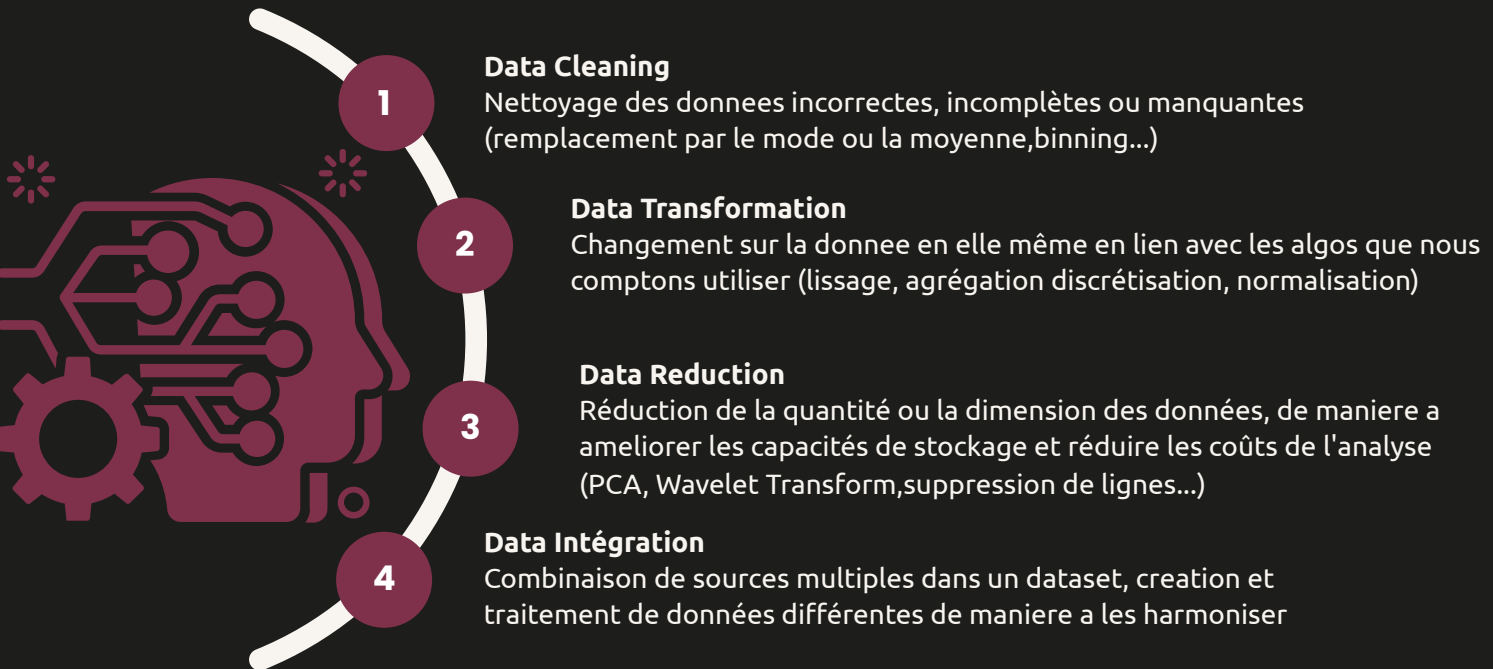
Les données à caractère personnel doivent être conservées « *pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées* ».



Déterminer les durées de conservation strictement nécessaires à la réalisation des finalités des traitements

Être en mesure d'expliquer en quoi cette durée est justifiée, eu égard à la finalité du traitement

Comment préparer la valorisation des données ?



Comment valoriser ses données ?

La Data Visualisation

Déterminer les informations de pilotage importantes

Méthodes de représentation

Déterminer le But

Méthodes statistiques d'analyse

Déterminer la cible

Méthode de mise à disposition

Déterminer les acteurs de cette analyse

Fréquence de mise à jour des informations

Déterminer les utilisateurs de ces informations



Comment valoriser ses données ?

Générer des données synthétiques

Vérifier la représentativité et la quantité de données



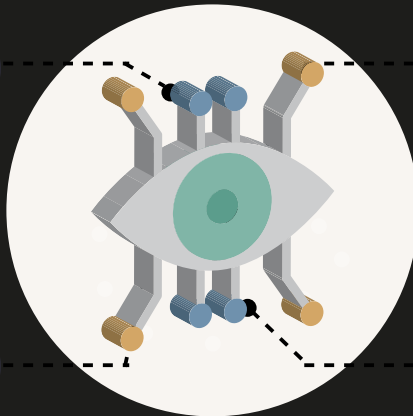
Mesurer la qualité des données synthétiques



Procéder à l'augmentation des données



Faire vérifier par des experts ces données synthétiques



Comment valoriser ses données ?

Les problèmes liés aux données

Selon Gartner, 85% des projets d'apprentissage automatique échouent en raison de problèmes de données.



Données
difficilement
accessibles



Données
Manquantes



Données
Insuffisantes



Données
Sensibles

15

Avec les données synthétiques

Selon Gartner, d'ici 2025 les données synthétiques réduiront la collecte des données personnelles des personnes, évitant 70% des sanctions pour violation de la vie privée.



Augmentation des
données



Partage et utilisation
des données
simplifiés



Diversité des
données



Fiabilité des données

Comment valoriser ses données ?

L'anonymisation

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière irréversible.

Méthodes



Permet de conserver la donnée au-delà de la durée de conservation



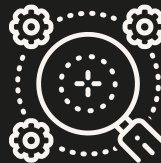
Ouvre la possibilité d'une réutilisation des données sans porter atteinte à la vie privée des personnes donnée au-delà de la durée de conservation

16

Comment ?



Randomisation



Généralisation



Données synthétiques

Segmentation, mesure et calculs automatique en radiologie ostéo-articulaire

Par Alia Santé

Processus de création de modèles IA :

- Définition de la problématique & solution
- Définition des objectifs des modules
- Consortium des acteurs experts
- Identification des sources de données
- Mise en place du protocole de collecte
- Méthodes d'annotations
- Protocole d'anonymisation
- Outils de stockage
- Augmentation des données
- Entraînement des modèles



07

Déploiement

Processus qualité

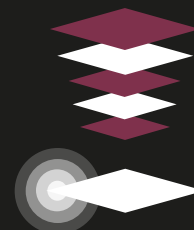
Certification

Méthodes de validation



Interface d'usage

Amélioration continue



Merci
pour
votre
écoute

Castres·Mazamet
Technopole



Alia Santé



Segolène Combettes - R&D Directrice
segolene@alia-sante.com



Yann Peter - Responsable Innovation et Développement des Entreprises
Yann.PETER@castres-mazamet.com



Sylvain Françoise Dal-Mas - President
Sylvain.fd@alia-sante.com

 Alia Sante

 Castres Mazamet
technopole

DES QUESTIONS ?