

10 **Science des données et technologies émergentes**

Table des Matières du Module

10	Science des données et technologies émergentes	1
10 - 1	Objectifs humanitaires grâce à la science des données	7
10 - 2	Rôle d'une équipe bénévole en science des données	11
10 - 3	À quoi pourrait ressembler une description du poste de Data Scientist	12
10 - 4	Processus de mise en œuvre durable et amélioration de la maturité	13
10 - 5	Données, technologies émergentes et objectifs de développement durable	15
10 - 6	Diversité sociale, équité et inclusion	21
10 - 7	Cartographie des aspects du travail en science des données	26
10 - 8	Pourquoi avons-nous besoin de la science des données ?	29
10 - 9	Aider à la prise de décision dans des contextes spécifiques	33
10 - 10	Vue d'ensemble de la plateforme et des processus typiques de la science des données	38
10 - 11	Comment mettre en œuvre un modèle de science des données	49
10 - 12	Analyse du taux de désabonnement de la Croix-Rouge - Jupyter Notebook	52

Science des données et technologies émergentes

Ce module présente aux équipes les concepts de base de la science des données et leur lien avec les objectifs humanitaires. Il vise à guider les équipes qui souhaitent explorer les possibilités de la science des données et des technologies émergentes dans leur travail quotidien vers l'innovation et la transformation.

Questions que ce module explore

- ▶ Comment pouvons-nous adopter et utiliser efficacement la science des données et les technologies émergentes dans notre travail ?
- ▶ Comment pouvons-nous comprendre et soutenir l'utilisation de la science des données et des technologies émergentes dans notre travail ?

Objectifs d'apprentissage

- ▶ Explorer comment les équipes comprennent et utilisent la science des données et les technologies émergentes dans leur travail.
- ▶ Guider les équipes vers les facteurs critiques de succès qu'elles doivent connaître et les questions clés qu'elles doivent poser lors de l'adoption de projets de science des données.
- ▶ Évaluer le niveau requis de sensibilisation aux technologies émergentes

Sujets du Module

- ▶ Pourquoi utiliser les technologies émergentes dans notre travail ?
- ▶ Comment utiliser les données et les principes fondamentaux pour relever les défis et les transformations humanitaires stratégiques ?
- ▶ Quelles sont les initiatives et les perspectives de la science des données et des technologies émergentes dans notre travail ?

Recettes

Une proposition de processus étape par étape pour atteindre les objectifs d'apprentissage

- 1 Commencez par identifier les **Objectifs humanitaires grâce à la science des données (10 - 1)** pour comprendre pourquoi les technologies émergentes sont introduites dans le travail humanitaire.
- 2 Discutez ensuite des **Processus de mise en œuvre durable et amélioration de la maturité (10 - 4)** qui est illustrée dans le tableau d'apprentissage des technologies émergentes.
- 3 Examiner de plus près **Données, technologies émergentes et objectifs de développement durable (10 - 5)** et la façon dont ils sont représentés dans le cadre des objectifs de développement durable.
- 4 Permettez aux groupes de faire la **Cartographie des aspects du travail en science des données (10 - 7)** qu'ils rencontrent dans leur travail et discutez des raisons pour lesquelles ils ont besoin du machine learning.
- 5 Guidez les participants sur **Comment mettre en œuvre un modèle de science des données (10 - 11)**.
- 6 Présenter aux groupes **Aider à la prise de décision dans des contextes spécifiques (10 - 9)**.
- 7 Lors d'une discussion de groupe, demandez aux participants d'indiquer s'ils ont l'intention d'intégrer des projets de science des données pour les aider dans leur travail humanitaire.

Ingrédients

Choisissez les ingrédients pour créer votre propre recette. Avez-vous un ingrédient qui nous manque ? Envoyer un mail à data.literacy@ifrc.org

Exercices

Présentations à utiliser et/ou à adapter :

- ▶ Possibilités d'aborder les objectifs humanitaires grâce à la science des données
- ▶ Cadre pour les objectifs de durabilité : un regard plus approfondi sur la durabilité dans l'IA
- ▶ Principes : diversité sociale, équité, inclusion dans la science des données
- ▶ Cartographie des différents aspects du travail en science des données
- ▶ Pourquoi avons-nous besoin du machine learning ?
- ▶ Questions pour aider à guider les décisions sur les plateformes, les outils et la sélection des compétences
- ▶ Comment mettre en œuvre un modèle de science des données ?

Diaporamas

Présentations à utiliser et/ou à adapter :

- ▶ Processus de mise en œuvre durable pour améliorer la maturité. Fournit un contexte sur le tableau d'apprentissage des technologies émergentes.

Checklists/Documents/Matériels

Pour la documentation des éléments essentiels de l'expérience d'apprentissage

- ▶ Rôle d'une équipe bénévole en science des données (Canvas). A utiliser avec **Pourquoi avons-nous besoin de la science des données (10 - 8)** (exercice).
- ▶ À quoi pourrait ressembler une description du poste de Data Scientist (Termes de référence). A utiliser avec **Pourquoi avons-nous besoin de la science des données ? (10 - 8)** (exercice).
- ▶ Plateformes de science des données (diapositives). Fournit un contexte sur les composants granulaires qui sont inclus dans les outils de science des données et leur utilisation. Le matériel pourrait être référencé dans **Cartographie des aspects du travail en science des données (10 - 7)** (exercice).
- ▶ Analyse du taux de désabonnement de la Croix-Rouge - Jupyter Notebook. A utiliser avec **Comment mettre en œuvre un modèle de science des données (10 - 11)** (exercice).

Ressources externes

- ▶ [An Abstraction Framework for Reducing Complexity in AI Governance \(Article\)](#)
- ▶ [Knowledge sharing and collaboration: Example of data and digital week](#) (Article de blog)
- ▶ [Examples from Red Cross Red Crescent experiences: Go Github data science projects](#) (Dépôt GitHub)
- ▶ [Artificial Intelligence Suitability Framework](#) (Diaporama)
- ▶ Alsolbi I., et.al (2021). Analyzing Donors Behaviors in Nonprofit Organisations: A Design Science Research Framework.
- ▶ Goffi E. R., Momcilovic A. (dir.) (2022). [Global Trends in AI 2022: Food for thought from GAIEI experts. Global AI Ethics Institute](#), Notes n° 1.

Prochaines étapes

Modules pertinents du Data Playbook

Se référer au module **Pratiques responsables et protection des données (7)** pour mieux comprendre comment atténuer les risques et **Biais cognitifs dans l'analyse humanitaire (6 - 6)** (Polycopié) dans le module **Comprendre et analyser les données (6)**.

Crédit

Mahendra Samarawickrama, Jenny Paola Yela-Bello, et les contributeurs de la V1 Sprint et du Data Playbook de la FICR.

10 - 1 Objectifs humanitaires grâce à la science des données

Lors de l'application des données et des technologies émergentes connexes à l'humanité, il est important que nous ayons identifié les opportunités, l'éthique, les valeurs et la justice sociale. Dans cette perspective, les équipes doivent s'assurer qu'elles comprennent et saisissent pourquoi la science des données et les technologies émergentes sont utilisées dans leur travail. Cet exercice permettra aux équipes d'examiner collectivement certains cadres et d'en discuter. Ces cadres aident les équipes à saisir les opportunités qui répondent à des objectifs humanitaires spécifiques dans le cadre de leur travail.

- ▶ **Personne** : 2 à 8 personnes
- ▶ **Durée** : 40 Minutes
- ▶ **Difficulté** : Moyen
- ▶ **Matériel virtuel** : plate-forme de réunion virtuelle, espace de rédaction/documentation partagé
- ▶ **Matériel en personne** : Tableaux, post-it, marqueurs

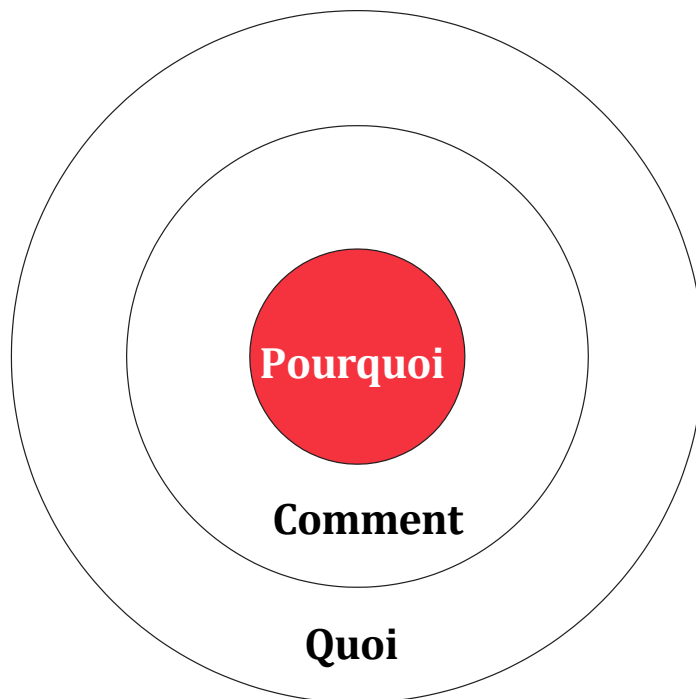
Exercice

Introduction

Dans cet exercice, nous abordons les raisons pour lesquelles les données, l'IA et les technologies émergentes sont importantes à prendre en compte dans les perspectives humanitaires. Le but de l'exercice est de trouver des réponses à la question de savoir pourquoi les données, l'IA et les technologies émergentes devraient être considérées comme une composante de l'information stratégique, et pourquoi elles jouent un rôle important dans le travail humanitaire.

Partie 1 : Explorer

En équipe, veuillez examiner les points suivants



Pour cet exercice, la stratégie 2030 de la FICR et les 17 objectifs de développement durable des Nations Unies sont utilisés comme cadres. Les équipes peuvent travailler collectivement ou être divisées en groupes pour discuter des questions sur les raisons pour lesquelles la science des données et les technologies émergentes devraient être un élément clé dans la formation de la stratégie. Après les discussions, l'animateur peut regrouper les idées qui ont été partagées dans une perspective plus large sur les raisons pour lesquelles les opportunités de la science des données s'appliquent à leur travail et sur la manière dont elles peuvent être mises en œuvre.

Note : [The role of artificial intelligence in achieving the Sustainable Development Goals.](#)

Partie 2 : Discuter

Question 1: Keeping the Strategy 2030 in mind, identify the opportunities that data science and Artificial Intelligence bring into humanity?

Cette question peut être discutée pendant 10 minutes

[FICR Strategie 2030](#)

3 Objectifs



1
Tout le monde peut anticiper les crises, y faire face et s'en relever rapidement

2
Tout le monde mène une vie sûre et saine dans la dignité et a des possibilités de prospérer

3
Tout le monde se mobilise pour bâtir des communautés inclusives et pacifiques

5 Défis

1
Changements climatiques et crise environnementale

2
Évolution des crises et des catastrophes

3
Lacunes croissantes dans l'accès à la santé et au bien-être

4
Migration et identité

5
Valeurs, pouvoir et inclusion

7 Transformations

1
Soutenir et développer les Sociétés nationales en tant qu'acteurs locaux forts et efficaces

2
Inspirer et favoriser le volontariat

3
Garantir la confiance et la redevabilité

4
Travailler efficacement en tant que réseau

5
Influencer l'action humanitaire

6
Assurer la transformation numérique

7
Construire l'avenir du financement

Question 2 : Connaissez-vous les **17 objectifs de développement durable (ODD) des Nations unies** ? Savez-vous que 79 % de ces objectifs peuvent être soutenus par l'IA ? Pouvez-vous faire correspondre certains d'entre eux à notre mission ? L'équipe peut ensuite discuter du diagramme pendant 10 minutes.



Crédit

Pour le développement de l'exercice Mahendra Samarawickrama

10 - 2 Rôle d'une équipe bénévole en science des données

[Télécharger Ressources](#)

10 - 3 À quoi pourrait ressembler une description du poste de Data Scientist

[Télécharger Ressources](#)

10 - 4 Processus de mise en œuvre durable et amélioration de la maturité

Tableau d'apprentissage des technologies émergentes :

Outil d'information sur le défi commercial

Problème & opportunité

Proposition de valeur du donateur / vision du projet : Elle définit "ce que nous faisons et pourquoi nous le faisons".

Problème du donateur : Quels sont les problèmes que nous résolvons ?

Opportunité pour l'entreprise : Quelle valeur cela apportera-t-il à l'entreprise ?

Segments de donateurs ou personas

Dressez la liste de vos donateurs et utilisateurs cibles pour lesquels nous résolvons les problèmes et fournissons des solutions.

Contexte de soutien

Recherches et idées existantes : Résultats clés des tests initiaux auprès des utilisateurs, de la recherche sur les donateurs ou de toute autre recherche qualitative, des enquêtes. Il peut également s'agir de toute recherche antérieure validée lors d'itérations précédentes. Ces éléments viennent à l'appui du "quoi et pourquoi" dans la section "opportunité" ci-dessus.

Objectif

Le succès : Que signifie la réussite et comment saurons-nous que nous apportons de la valeur ? Quel est l'indicateur clé de la réussite ?

Hypothèses

Apprentissage : l'idée ou la solution proposée qui apportera de la valeur et nous rapprochera de l'objectif.

Quelles sont les stratégies clés qui produiront le plus de valeur ? Ou quel apprentissage validé est important pour confirmer notre chemin vers le succès ? C'est le "comment" des objectifs. Une bonne hypothèse permet de déterminer ce qui doit être testé par le biais de mesures et d'expériences (pour être prouvé ou réfuté).

Chronologie

Quelles sont les étapes critiques ?

Équipe/parties prenantes

Business Owner

Product Owner

10 - 5 Données, technologies émergentes et objectifs de développement durable

La mission de la Croix-Rouge et du Croissant-Rouge est de prévenir ou de réduire la souffrance humaine, où qu'elle se trouve.

Nous faisons partie du plus grand mouvement humanitaire au monde, opérant dans plus de 192 pays. Nous comptons plus de 14,8 millions de volontaires dans le monde. La CRCR joue un rôle d'auxiliaire du gouvernement - reconnu comme un partenaire fiable pour mobiliser le pouvoir de l'humanité. Avec la mise en œuvre de la Stratégie 2030 et l'utilisation croissante d'outils numériques dans les missions de la CRCR, il convient d'examiner de plus près comment les outils numériques (l'intelligence artificielle étant l'un d'entre eux) peuvent être durables et refléter les valeurs, les forces et la culture de la CRCR.

- ▶ **Personne** : 2 à 8 personnes
- ▶ **Durée** : 60 Minutes
- ▶ **Difficulté** : Moyenne
- ▶ **Matériel virtuel** : plate-forme de réunion virtuelle, espace de rédaction/documentation partagé
- ▶ **Matériel en personne** : Tableaux, post-it, marqueurs

Exercice

Dans cet exercice, l'équipe, avec l'aide du facilitateur, discutera des aspects qu'il est important de mettre en évidence et de travailler pour que l'IA devienne durable.

Partie 1 : Explorer et réviser

Il s'agira d'une discussion plénière en 4 parties.

La session couvrira 4 points principaux accompagnés de diagrammes et d'une vue d'ensemble qui peuvent aider à diriger la discussion.

Pourquoi la science des données et l'IA doivent-elles être gouvernées de manière éthique ?

La Croix-Rouge australienne travaille sur la gouvernance de l'IA pour la durabilité parce que l'IA a un énorme potentiel pour l'humanité et un risque important si elle n'est pas correctement gouvernée. Selon les prévisions, l'IA pourrait contribuer à l'économie mondiale à hauteur de 15,7 trillions de dollars d'ici à 2030. Dans le même temps, l'IA crée des risques importants pour l'humanité, tels que les armes autonomes, les pertes d'emploi dues à l'automatisation, les inégalités socio-économiques, les violations de la vie privée, les "Deepfakes" et les préjugés causés par les données et les algorithmes. En outre, il a été prédit que d'ici 2022, 85 % des projets d'IA échoueront en raison de biais dans les données, les algorithmes ou les équipes chargées de les gérer.

L'une des façons d'aborder l'IA durable pour le développement durable est de renforcer l'éthique et la gouvernance de l'IA tout en suivant les bonnes pratiques. Pouvez-vous imaginer d'autres façons de l'intégrer et de l'utiliser ?

Quelles sont nos valeurs, notre objectif et notre force lorsque nous nous concentrons sur la science des données et l'IA ? Ouvrez la discussion après avoir jeté un coup d'œil au diagramme ci-dessous.

Veuillez consulter l'[animation](#)

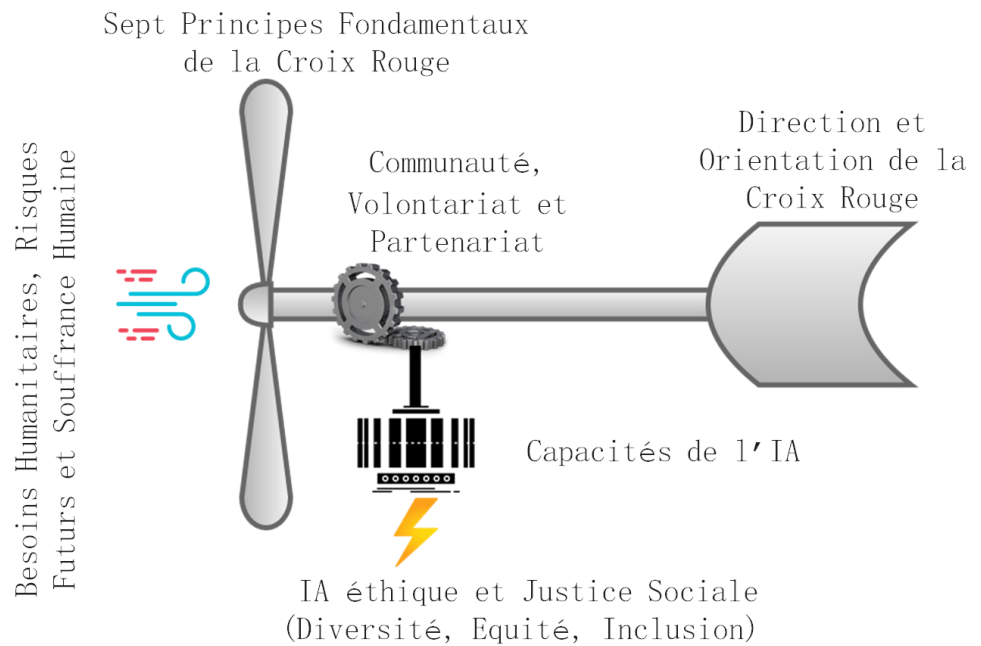


Fig.: Un modèle conceptuel pour gérer l'IA de manière éthique. (Source: <https://hyperight.com/social-justice-and-sustainability-by-leveraging-data-science-and-ai-interview-with-dr-mahendra-samarawickrama/>)

Les personnes, la culture et la mission de l'IA. Que pouvons-nous dire de l'IA sur la base de notre personnel, de notre culture et de notre mission ? Discutez de ces attributs et de la figure illustrée ci-dessous.

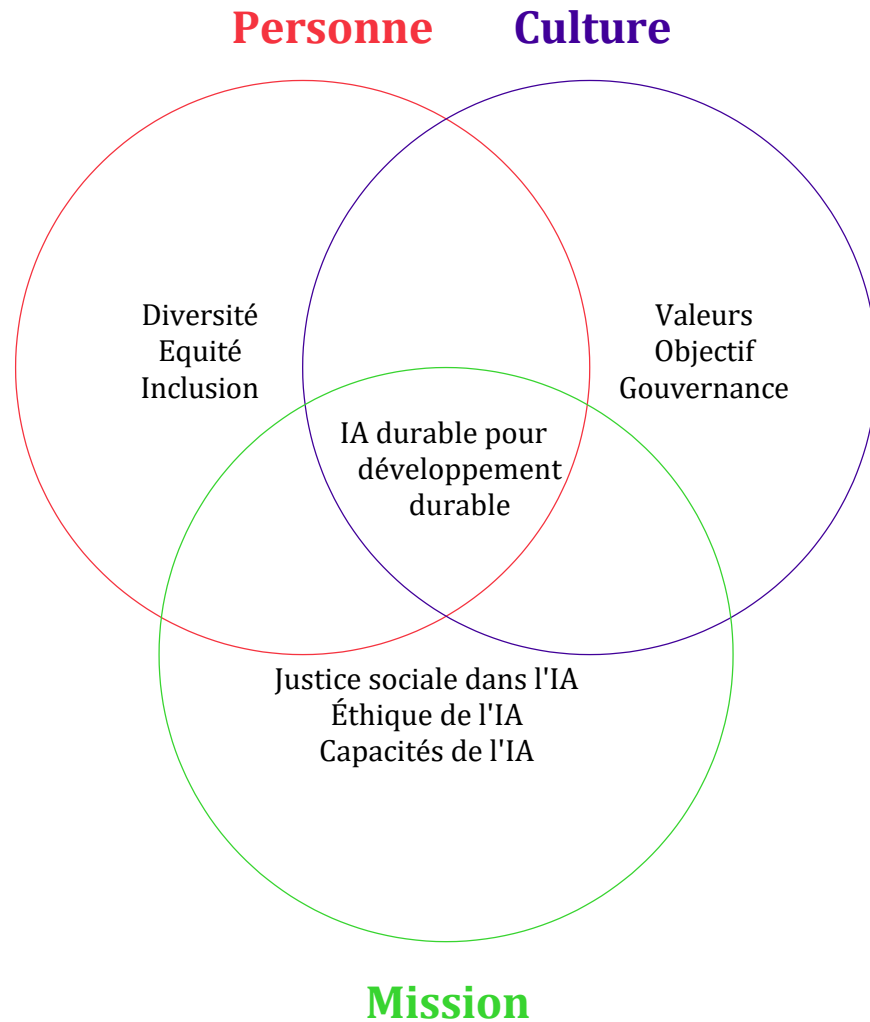


Fig.: Les aspects des personnes, de la culture et de la mission sur l'IA durable pour la durabilité.

Comment faire de la science des données et de l'IA des compétences de base ? Discutez des moyens d'intégrer l'IA dans les compétences de base, tout en examinant le diagramme ci-dessous.

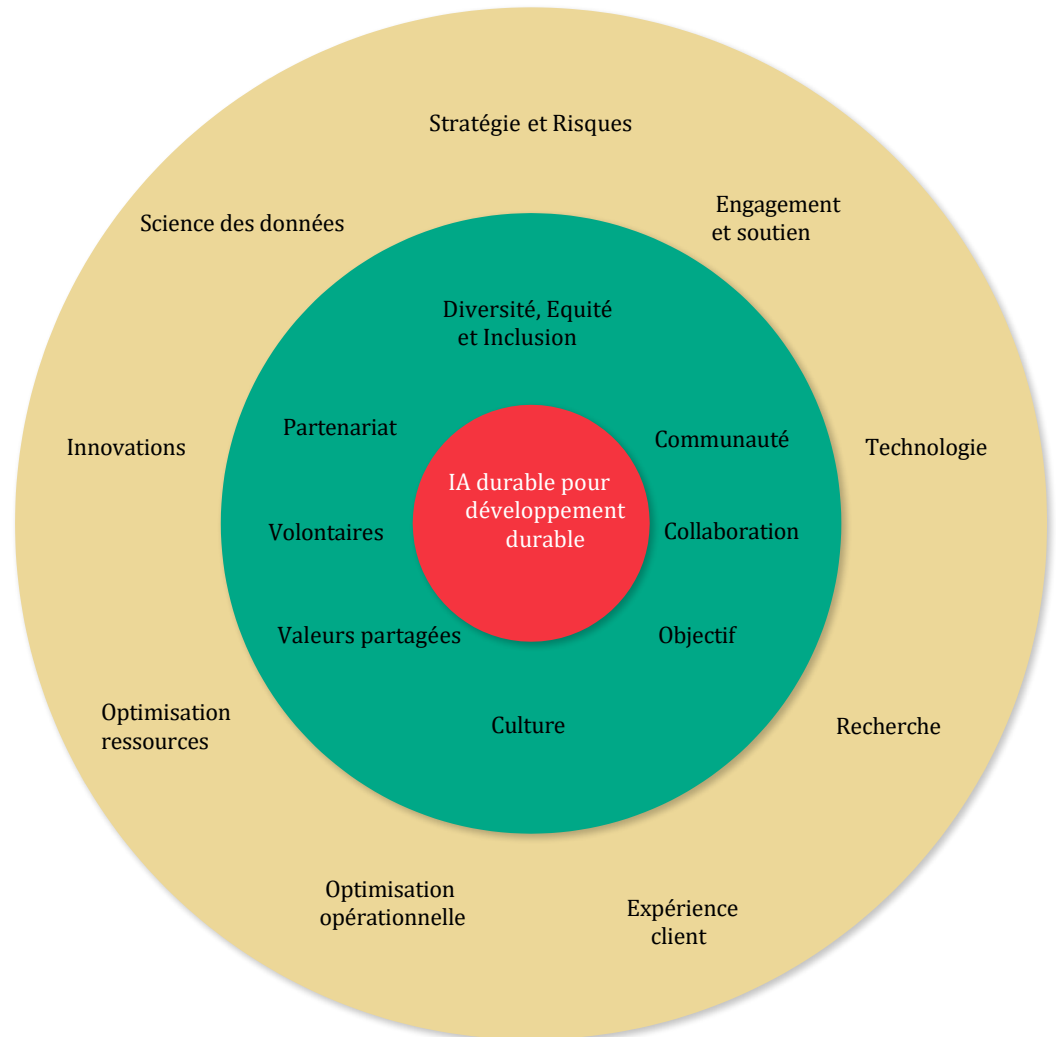


Fig.: Faire de l'IA une compétence de base pour stimuler l'activité.

Partie 2 : Feedback

Demandez aux participants de réfléchir à l'exercice et de citer une ou deux observations ou idées.



Bonus

Les exposés de la Croix-Rouge australienne sur l'éthique et la gouvernance de l'IA :

Data Innovation Summit 2021 | AI for ESG Toolkit | [A Unified AI Governance Toolkit for ESG](#)

Ai4 Conference 2021 | [A Unified AI-Governance Collaborative Framework for Sustainability](#)

Crédit

Pour l'élaboration de l'exercice, Mahendra Smarawickrama et Paola Yela.

10 - 6 Diversité sociale, équité et inclusion

Pour aborder la question de la diversité sociale, de l'équité et de l'inclusion, il faut une compréhension plus large de la gouvernance des technologies émergentes et de la manière dont elles jouent un rôle dans la durabilité et l'humanité au sein de notre réseau.

- ▶ **Personne** : 2 à 8 personnes
- ▶ **Durée** : 30 Minutes
- ▶ **Difficulté** : Moyen à Difficile
- ▶ **Matériel virtuel** : plate-forme de réunion virtuelle, espace de rédaction/documentation partagé
- ▶ **Matériel en personne** : Tableaux, post-it, marqueurs

Exercice

Partie 1 : Explorer et réviser

Gouvernance de l'IA et des technologies émergentes pour la durabilité et l'humanité

Lorsqu'il s'agit d'intégrer la diversité, l'équité et l'inclusion (DEI) dans les données et les technologies émergentes, les dirigeants et les collaborateurs doivent être conscients des différentes perspectives concernant le pourquoi, le comment et le quoi.

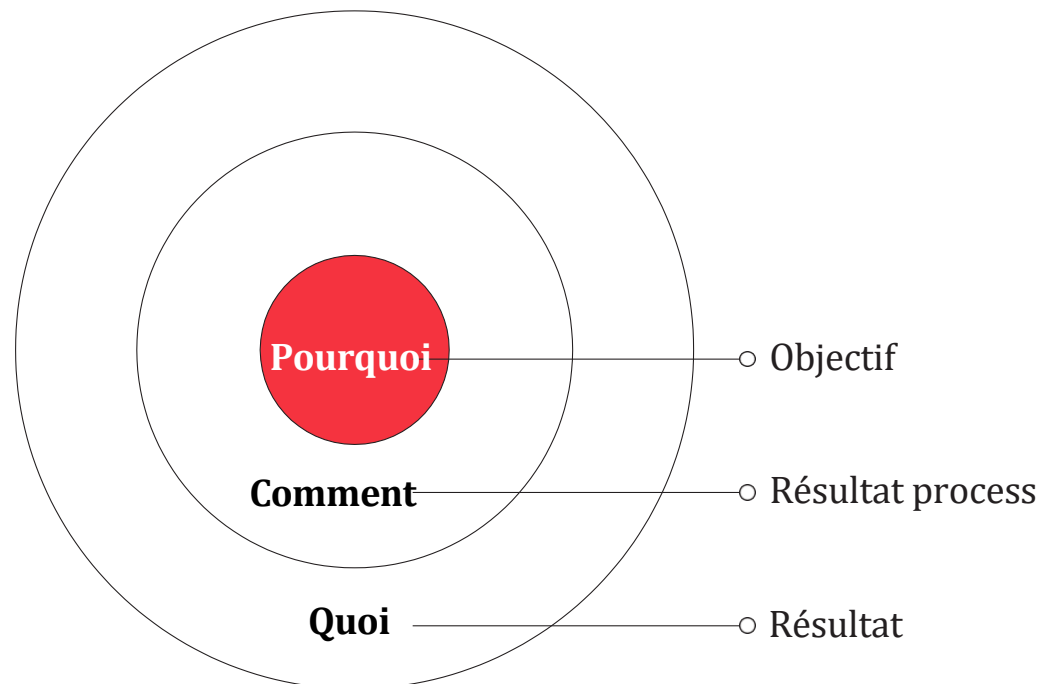
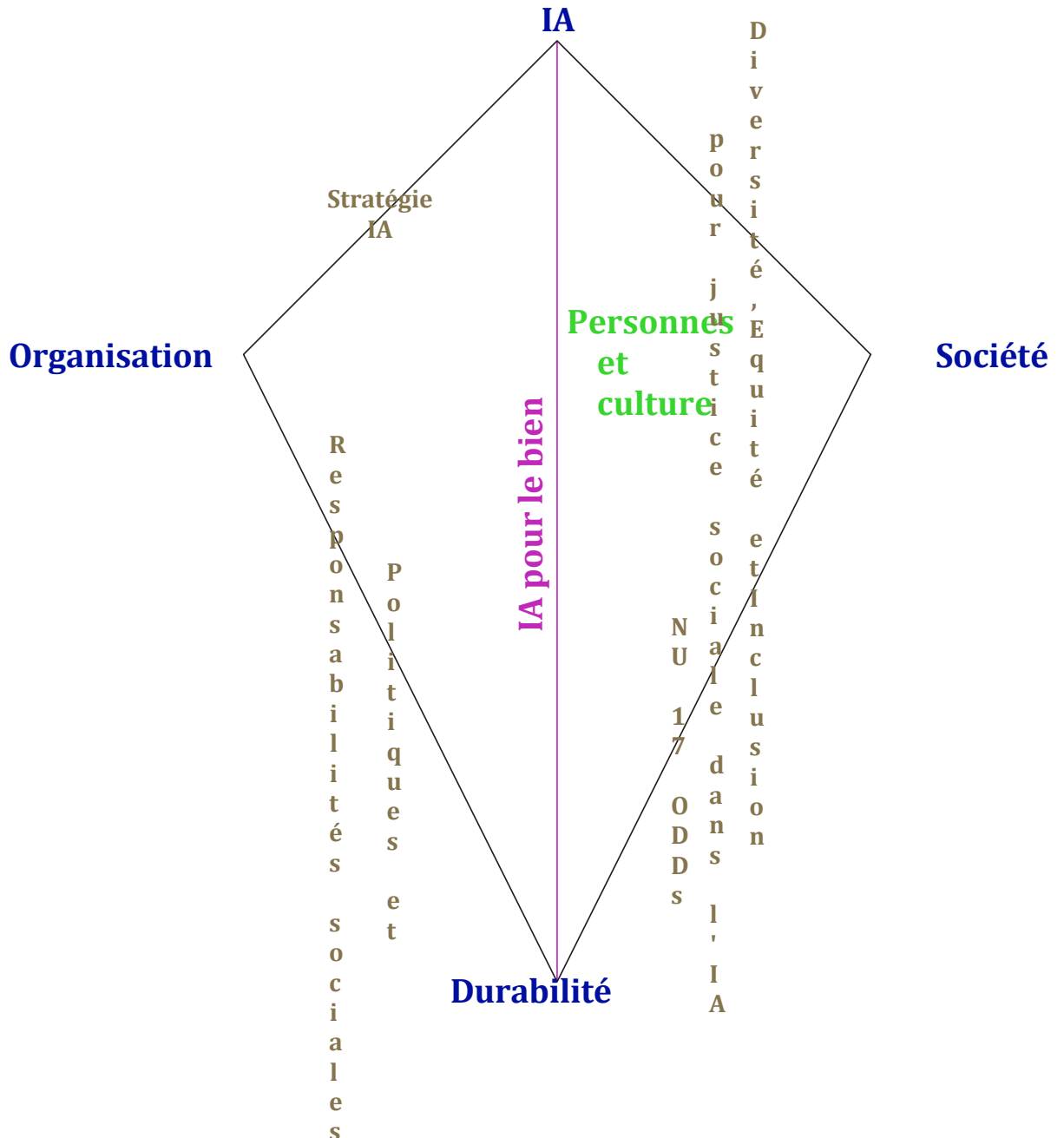


Fig: Cercle doré

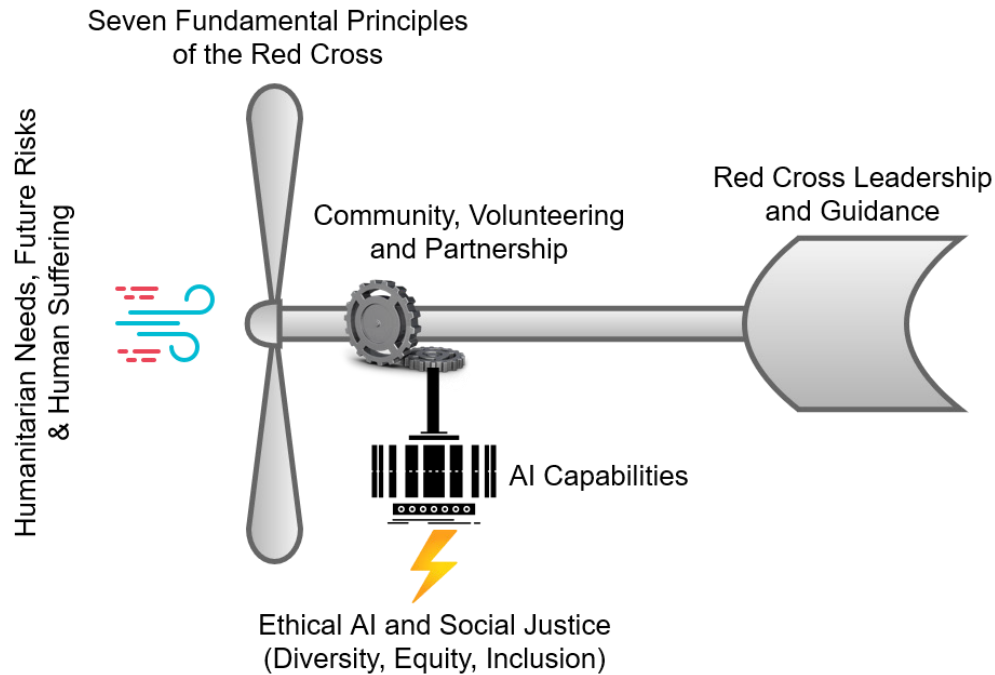
Pour réussir dans la DEI, il est important de comprendre l'importance des personnes, de la culture et de la mission vis-à-vis des technologies émergentes et d'atténuer le risque d'échec dû à des préjugés.

Veillez-vous référer aux visualisations suivantes :

Quels sont les aspects importants des technologies émergentes et de la gouvernance de l'IA ? (Compréhension des questions "Pourquoi" ; par exemple pourquoi les données et l'IA jouent un rôle clé et pourquoi la gouvernance de l'IA et la durabilité doivent être prises en compte). Veuillez consulter : [Le cadre d'abstraction KITE pour la gouvernance de l'IA et l'ESG pour la durabilité.](#)



Comment appliquer les valeurs de la Croix-Rouge aux technologies émergentes et à la gouvernance de l'IA ? (Comprendre les questions "Quoi" et "Comment"). Veuillez lire la visualisation du modèle conceptuel de l'éolienne.



Pouvez-vous expliquer comment la diversité sociale, l'équité et l'inclusion sont prises en compte dans votre projet de technologies émergentes ou d'IA ? Dans quelle mesure êtes-vous convaincu de la justice sociale de votre projet ? Avez-vous contribué à la durabilité (sociale ou environnementale) ?

- ▶ Veuillez consulter la success story suivante, récemment présentée au IFRC Planet:Red Summit
- ▶ Lien de la session : <https://planetredsummit.com/session/UJSGIU-1>
- ▶ Vidéo de la session : <https://youtu.be/psySmdBc0IE>
- ▶ Veuillez également partager votre histoire.
- ▶ Ressource supplémentaire : Samarawickrama, M. (2022). [Keeping AI Honest. Company Director, 38\(2\), 52-53](#)

Votre Société	Lien public vers la success story

Partie 2 : Feedback

Demandez aux participants de réfléchir à l'exercice et de citer une ou deux observations ou idées.

Bonus

Comment apporter de la diversité sociale, de l'équité et de l'inclusion de la part des bénévoles.

- ▶ Si du temps est disponible, nous pouvons parler de la formation d'une équipe volontaire de science des données dans votre société. Veuillez utiliser le Canvas d'innovation des technologies émergentes pour cela.
- ▶ Veuillez utiliser le canvas alimenté par la Croix-Rouge australienne pour établir leur équipe bénévole de science des données.
- ▶ [Canvas](#)
- ▶ [Description du poste](#)
- ▶ [Équipe de volontaires en science des données de la Croix-Rouge australienne](#)

Cela peut être utilisé avec le module **Pratiques responsables et protection des données (7)** et **Valeurs humanitaires et protection des données (polycopié) (7 - 8)** (exercice).

Crédit

Mahendra Samarawickrama et Paola Yela pour l'élaboration de l'exercice.

10 - 7 Cartographie des aspects du travail en science des données

Clarifier ce que nous entendons par science des données. Cartographier les différents aspects du travail de science des données, pourquoi avons-nous besoin de la science des données et illustrer le type de produits/analyses que nous pouvons en attendre. Cet exercice permet aux équipes de vérifier si leur conceptualisation de la science des données est alignée. A la fin de cette session, nous espérons que chacun aura une vue d'ensemble de la science des données, de ses différentes perspectives, et pourra comprendre son rôle dans le spectre de travail de la science des données.

- ▶ **Personne** : 2 à 8 personnes
- ▶ **Durée** : 30 Minutes
- ▶ **Difficulté** : Facile à Moyen
- ▶ **Matériel virtuel** : plate-forme de réunion virtuelle, espace de rédaction/documentation partagé
- ▶ **Matériel en personne** : Tableaux, post-it, marqueurs

Exercice

Part 1 : Explorer et réviser

Veillez discuter de la manière dont une équipe de science des données peut être constituée en vous référant au diagramme ci-dessous et où leurs rôles s'inscrivent-ils le mieux dans l'éventail ?

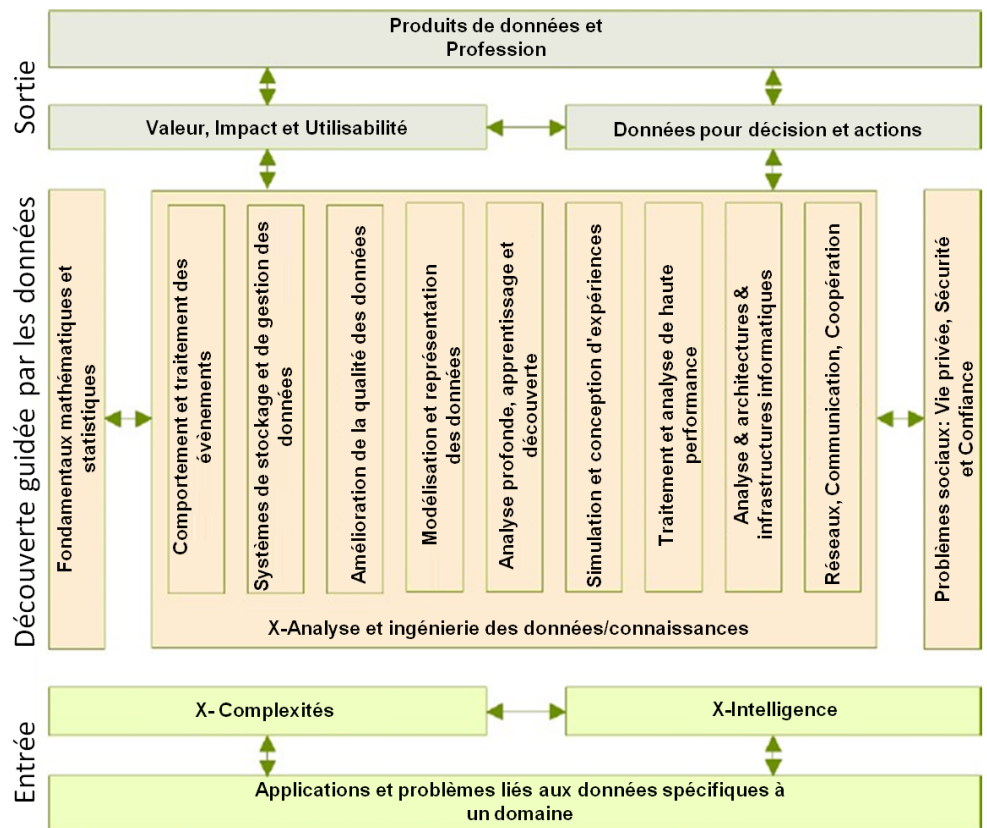
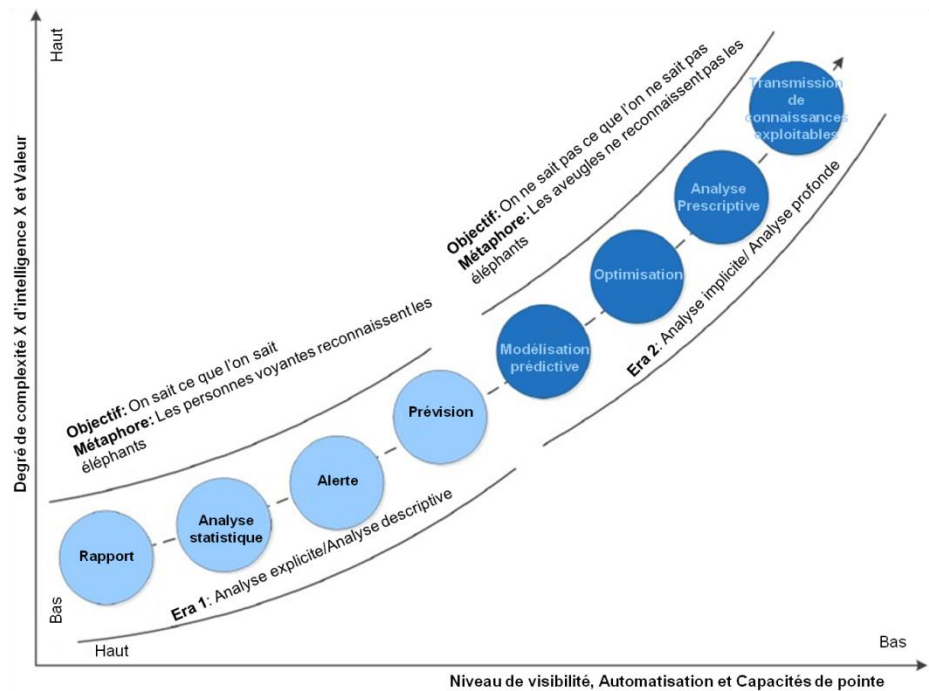


Fig. 6 Plan conceptuel de la science des données

(i) Tiré d'ici : Source: <https://arxiv.org/abs/2007.03606>

Discutez des principales composantes de la science des données illustrées par le diagramme suivant et demandez-vous si elles correspondent aux besoins de votre Société nationale lors de la formation de son équipe chargée des données. ([Science des données : une vue d'ensemble](#))



Partie 2 : Feedback

Demandez aux participants de réfléchir à l'exercice et de citer une ou deux observations ou idées.

Bonus

L'animateur peut poursuivre cette discussion en se référant à l'exercice Pourquoi avons-nous besoin de la science des données du module 10, en reliant les principaux enseignements de cette session à l'exploration des capacités d'une équipe de science des données.

Crédit

Pour l'élaboration de l'exercice Mahendra Samarawickrama et Paola Yela

10 - 8 Pourquoi avons-nous besoin de la science des données ?

Clarifier ce que nous entendons par science des données. Cartographier les différents aspects du travail de la science des données, pourquoi avons-nous besoin de la science des données et illustrer le type de produits/analyses que nous pouvons en attendre. Cet exercice explore la nécessité de créer une équipe de science des données et aide à découvrir certains aspects clés à connaître lorsque l'on commence à travailler avec des équipes de science des données au sein du RCRC.

- ▶ **Personnes** : 2 à 8 personnes
- ▶ **Durée** : 45 – 60 Minutes
- ▶ **Difficulté** : Moyen à Difficile
- ▶ **Matériel en personne** : Tableau et marqueurs
- ▶ **Matériel virtuel** : Plateforme de réunion virtuelle, document partagé/espace d'écriture.

Exercice

Partie 1 : Explorer et réviser

Étude de cas : L'équipe de direction de la Croix-Rouge ABC a décidé de mettre en place une équipe de science des données capable de mettre la science des données et l'analyse au service de l'activité.

En vous référant au scénario, permettez aux groupes de discuter d'une série de questions qui peuvent les aider à explorer la nécessité de mettre en place une équipe de science des données et si cela les rapproche de l'identification de l'utilisation du machine learning dans le spectre de la science des données.

Questions qui seront utilisées pour la discussion :

- 1 Qu'est-ce que la science des données ?
 - (aide pour l'animateur - les réponses peuvent inclure : science des données = statistiques + informatique + calcul + communication + sociologie + gestion | données + environnement + réflexion).
- 2 Pourquoi la direction générale a-t-elle besoin d'une équipe de science des données ?
 - (aide pour l'animateur - les réponses peuvent inclure : pour aider à la prise de décision basée sur les données, pour personnaliser l'expérience client, pour aider à faire des prédictions et à la gestion des risques, pour diversifier les services et la stratégie, etc.)
- 3 Quelles sont les capacités actuelles dont l'équipe de direction doit être consciente ?
 - (aide pour l'animateur - les réponses peuvent inclure : les plateformes, les compétences des personnes, les outils utilisés, la culture des données, etc.)
- 4 Quels sont les investissements pour lesquels l'équipe dirigeante devrait être prête à dépenser ?

- (aide pour l'animateur - les réponses peuvent inclure : Identifier des modèles dans les données et créer un modèle dans lequel nous pouvons faire des prédictions à l'avance).

Partie 2 : Feedback

Demandez aux participants de réfléchir à l'exercice et de citer une ou deux observations ou idées.

Bonus

Pour approfondir l'intégration du machine learning dans les projets de science des données, les groupes ont la possibilité de répondre aux questions suivantes :

- 1 Avez-vous entendu parler des algorithmes de machine learning ?
- 2 Pouvez-vous les expliquer à un non-technicien ?
- 3 Que signifie "garbage in garbage out" dans la modélisation de la science des données ?
- 4 Qu'est-ce qu'une hypothèse dans la modélisation en science des données ?
- 5 Pouvez-vous expliquer le biais et la variance dans la modélisation en science des données à l'aide de l'hypothèse ?

Possibilité pour l'animateur de se référer aux documents complémentaires Rôle d'une équipe bénévole de science des données et À quoi pourrait ressembler une description du poste de Data Scientist, fournis dans le module 10.

Crédit

Pour l'élaboration de l'exercice Mahendra Samarawickrama et Paola Yela, sous la direction de Melissa El Hamouch

Ressource supplémentaire : Eggleton, M., 2022. Un lauréat met en garde contre les échecs de l'intelligence artificielle. [The Australian Financial Review - Special Report of Reimagination 2022](#), pp.S4-S5.

10 - 9 Aider à la prise de décision dans des contextes spécifiques

Lorsque plusieurs options de processus de science des données sont disponibles, il n'est pas toujours facile d'identifier la meilleure option à mettre en œuvre pour un projet donné. Cet exercice aidera les équipes à discuter des options disponibles en fonction de leurs ressources et de leurs capacités.

- ▶ **Personne** : 2 à 8 personnes
- ▶ **Durée** : 45 – 60 Minutes
- ▶ **Difficulté** : Difficile
- ▶ **Matériel virtuel** : Plateforme de réunion virtuelle, espace de rédaction/documentation partagé
- ▶ **Matériel en personne** : Tableaux, post-it, marqueurs

Exercice

Partie 1 : Explorer et Réviser

Cet exercice permet aux équipes de répondre à des questions sur leurs ressources, leurs capacités et leur temps afin de comprendre quelle est l'option la plus appropriée pour un objectif particulier. L'animateur peut soit proposer une tâche de projet théorique, soit discuter d'une tâche proposée par l'équipe.

Note : Demandez aux équipes de rédiger différentes options qui peuvent être comparées aux ressources dont elles disposent et à la situation dans laquelle elles se trouvent.

L'animateur peut alors guider l'équipe dans l'identification de l'option qui conviendra le mieux à sa situation. Une liste de questions peut être posée pour aider les groupes et les équipes à réfléchir.

Les questions peuvent être les suivantes :

- ▶ De combien de ressources disposez-vous ?
- ▶ De combien de temps disposez-vous ?
- ▶ Combien de temps faudra-t-il pour travailler sur le projet ?
- ▶ Quelles sont les contraintes potentielles ?
- ▶ Quel est le budget alloué au projet ?
- ▶ Quel est le personnel disponible pour travailler sur le projet ?
- ▶ Quelles sont les compétences requises de la part de l'équipe ?
- ▶ Dans quel contexte le travail se déroulera-t-il (contexte d'urgence ou non) ?
- ▶ De quelles données l'équipe dispose-t-elle (données historiques, données secondaires, absence de données, etc.) et de quelles données a-t-elle besoin ?

L'animateur a la possibilité de présenter à l'équipe les cas ci-dessous comme point de départ des discussions. Il peut ensuite se référer à la série de questions présentées ci-dessus pour obtenir une vision plus approfondie de la part de l'équipe et évaluer ce qui est nécessaire et ce qui est disponible. Cela les aidera à décider quel processus de science des données et quelles ressources seront les plus efficaces.

Cas A :

La Société nationale XYZ dispose d'un référentiel d'évaluations des risques communautaires réalisées dans plusieurs régions de son pays. Les évaluations sont des documents non

structurés dont le contenu est similaire mais dont les formats sont variés.

L'équipe de gestion des risques de la Société nationale aimerait obtenir des informations clés à partir de ces documents, mais elle n'a pas la capacité interne de traiter tous les documents. Il ne s'agit pas d'une tâche urgente, mais elle dispose de ressources qui pourraient être investies dans cette tâche.

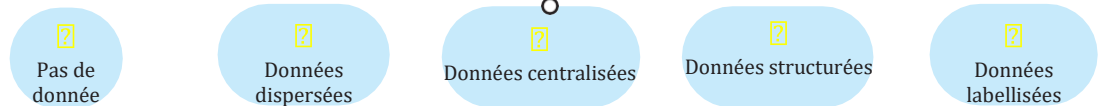
Objectif



Calendrier



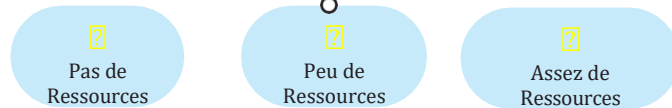
Données



Equipe



Ressources



L'équipe de la Société nationale XYZ a décidé de faire un stage de 4 mois dans une institution technique. Votre équipe ferait-elle de même ? Quelle option auriez-vous préférée ?

Cas B :

Veillez sélectionner votre parcours situationnel :

Objectif

Objectif Clair

Phase
Exploratoire

Calendrier

< 1 Semaine

< 1 Mois

< 1 Année

> 1 Année

Données

Pas de
donnée

Données
dispersées

Données centralisées

Données structurées

Données
labellisées

Equipe

Curieux des
données

Actif des
données

Data Scientist

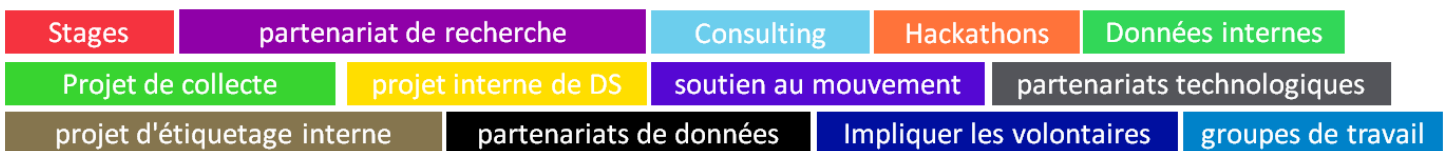
Ressources

Pas de Ressources

Peu de Ressources

Assez de
Ressources

Quelle option envisagez-vous ? Quelles autres options proposeriez-vous dans une telle situation ?



Partie 2 : Feedback

Demandez aux participants de réfléchir à l'exercice et de citer une ou deux observations ou idées.

Bonus

Pour en savoir plus, les équipes peuvent consulter le [Cadre d'adéquation de l'IA](#) qui fournit un ensemble de questions à se poser lors de l'élaboration d'un projet de science des données. Il apporte suffisamment de détails et s'efforce de conserver une perspective pratique. Le graphique ci-dessus peut être considéré comme une première approche.

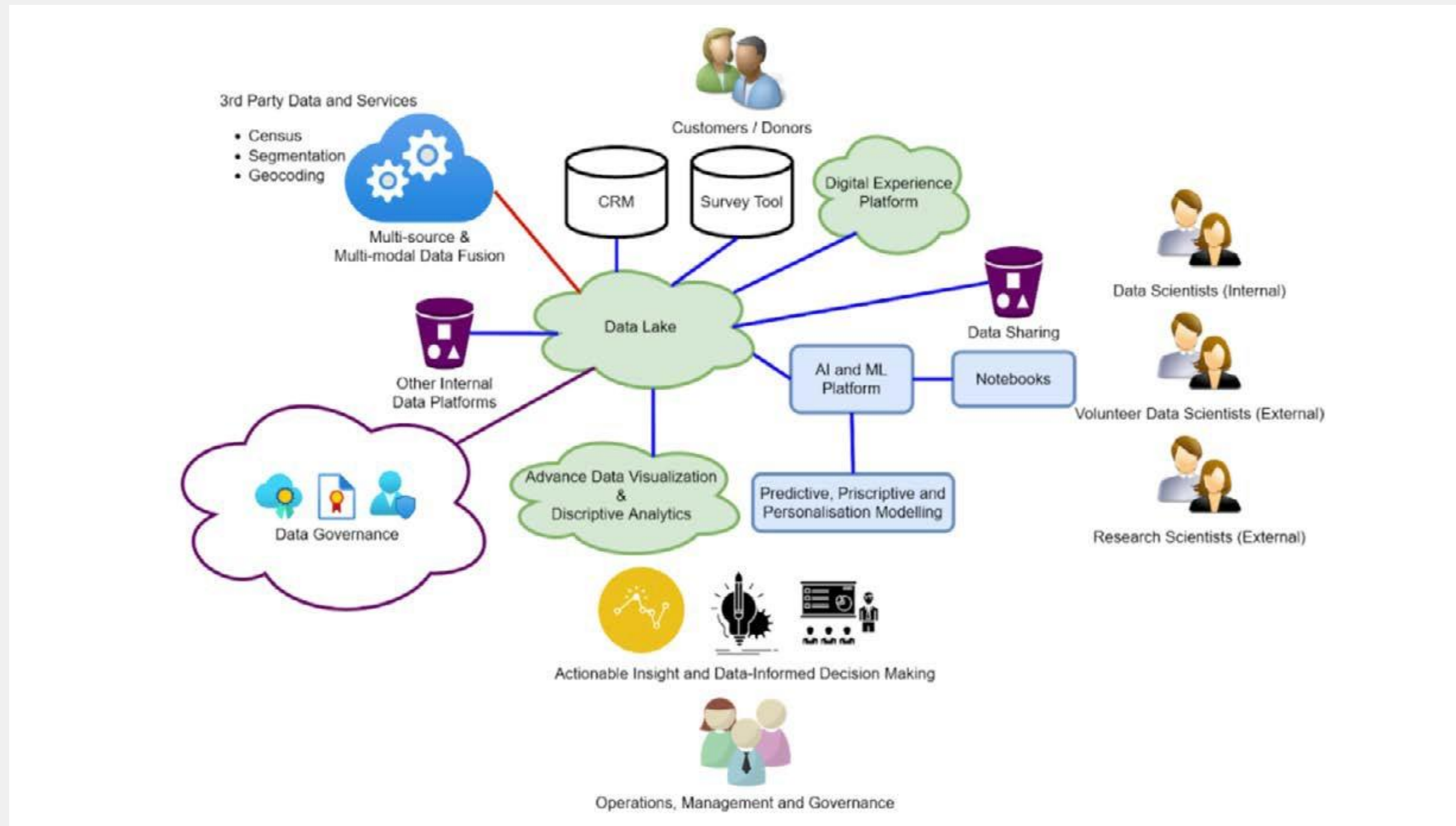
Crédit

Pour le développement de l'exercice Paola Yela et Melissa El Hamouch

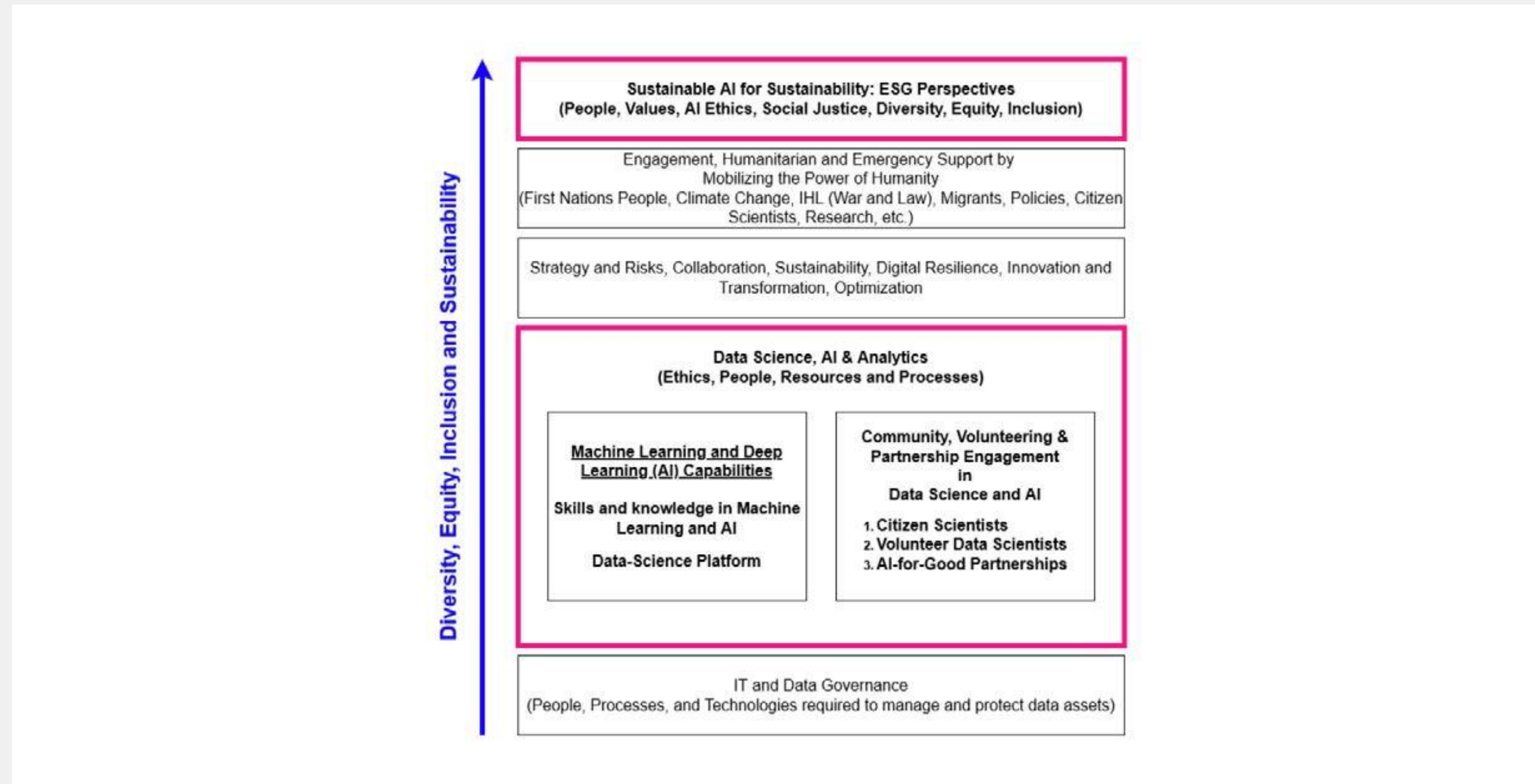
10 - 10 Vue d'ensemble de la plateforme et des processus typiques de la science des données

Crédit : Mahendra Samarawickrama pour le Data Playbook de la FICR

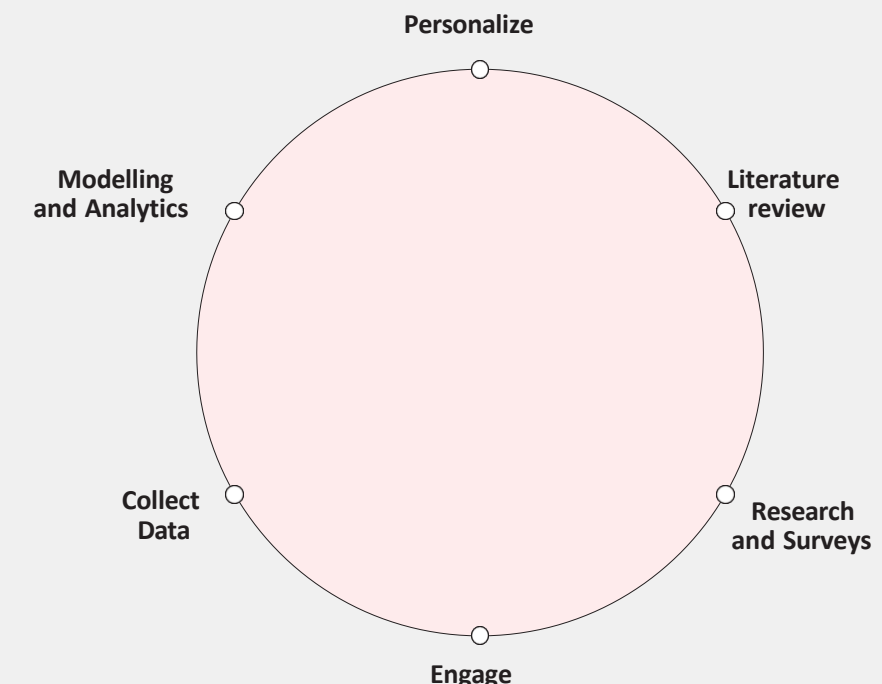
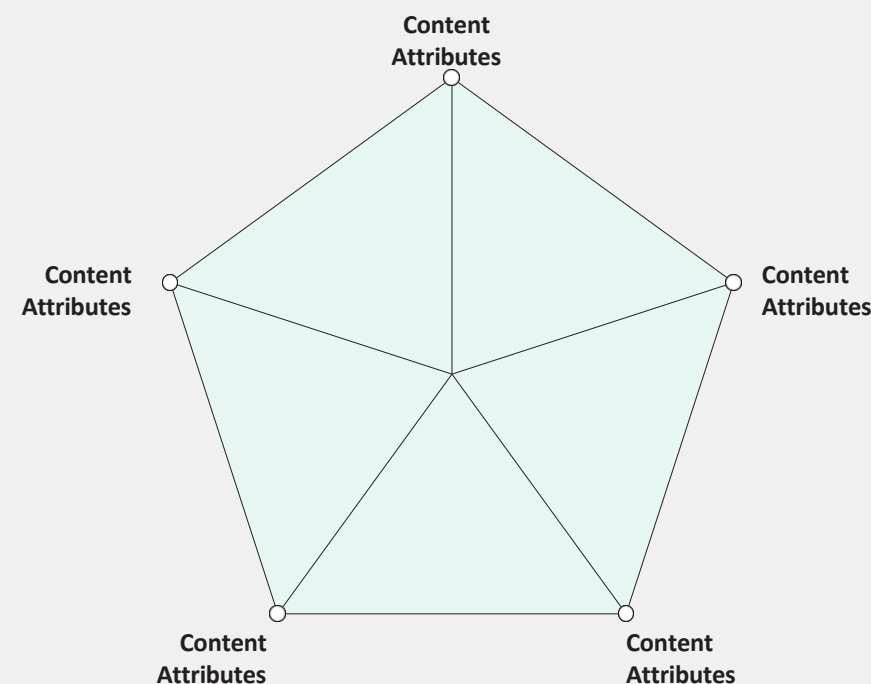
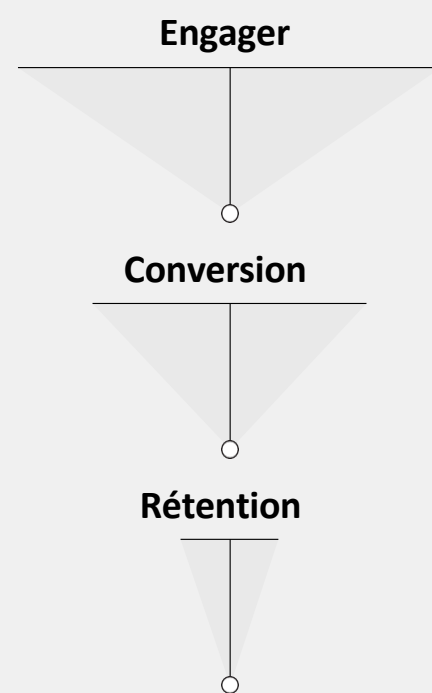
Schéma typique d'une plateforme de science des données



Processus typique de science des données axé sur les objectifs humanitaires

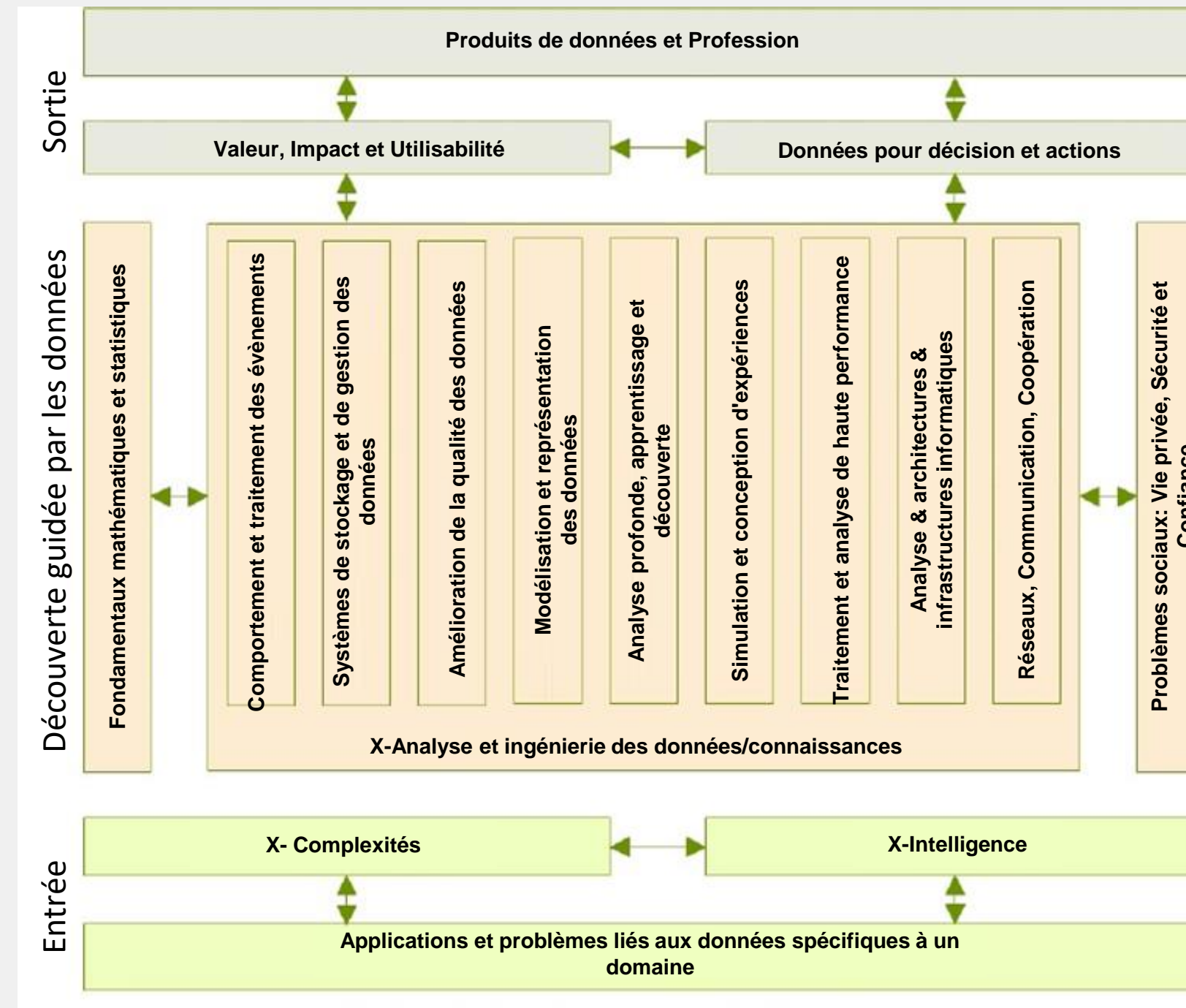


Science des données pour la personnalisation



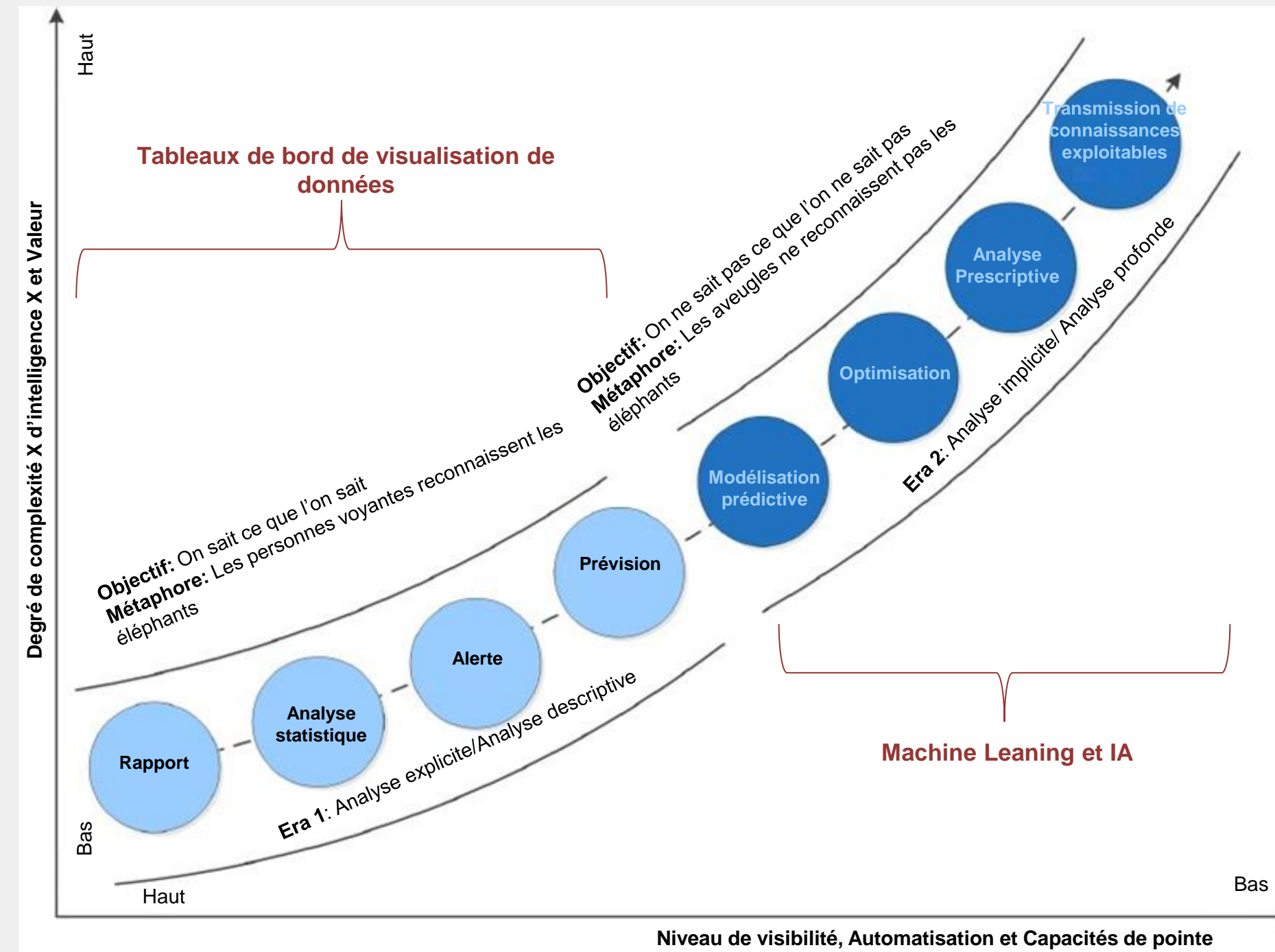
Engagement et Support				
CX	Acquisition et fidélisation	Digital	Marketing	Finance
Data Insight as a Service				
Science des données (ex. : personnalisation, prédiction, analyse avancée)				
Architecture des données (ex. : infrastructure)				
IT (E.g., Sécurité des données)				

Plan conceptuel de la science des données



(i) Source: Longbing Cao. 2017. Data Science: A Comprehensive Overview. ACM Comput. Surv. 50, 3, Article 43 (Juin 2017), 42 pages. DOI: <https://doi.org/10.1145/3076253>

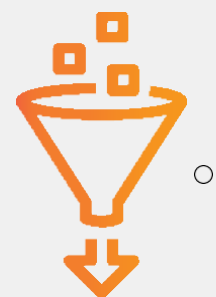
Road Map de la science des données (création de valeur)



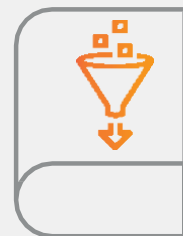
(i) Source: Longbing Cao. 2017. Data Science: A Comprehensive Overview. ACM Comput. Surv. 50, 3, Article 43 (Juin 2017), 42 pages. DOI: <https://doi.org/10.1145/3076253>

Composants typiques d'une plateforme de science des données

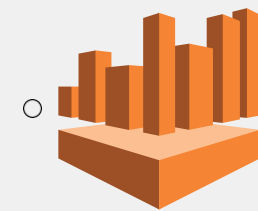
1. Lac de données



AWS Glue



AWS Glue Catalog



Athena



S3 Bucket

2. Machine Learning et plateforme d'IA



Services IA:

Amazon Rekognition • Amazon Polly • Amazon Lex

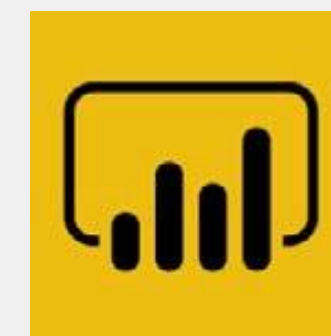
Plateformes IA:

Amazon Machine Learning • Amazon EMR • Spark & Spark ML

Engines IA :

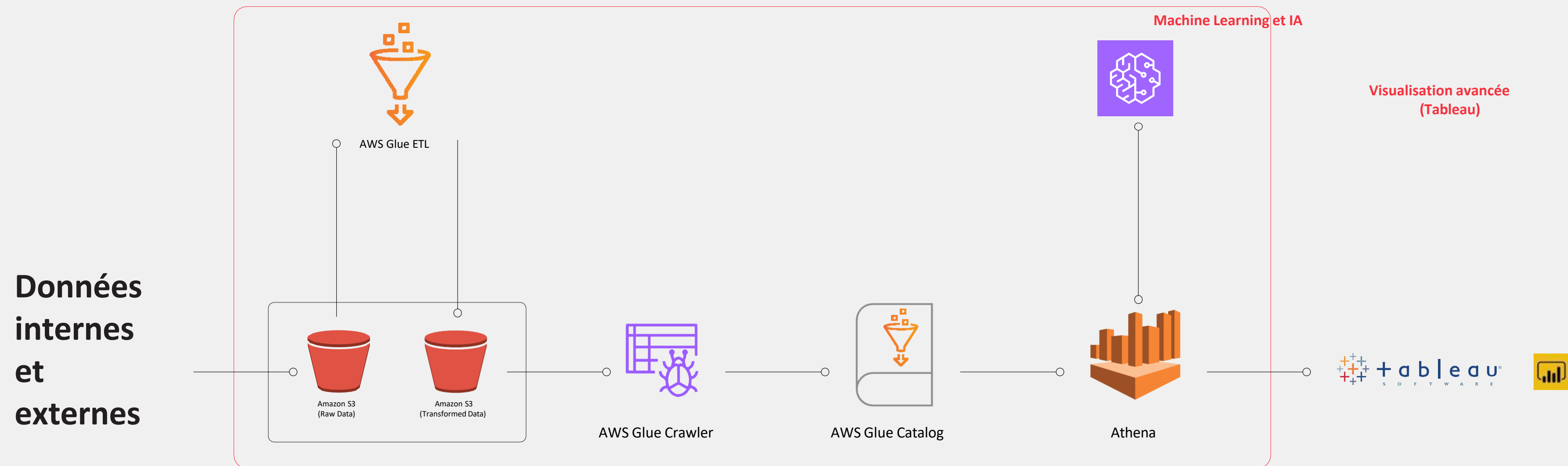
Apache MXNet • TensorFlow • Caffe • Torch • Theano • CNTK • Keras

3. Framework de visualisation

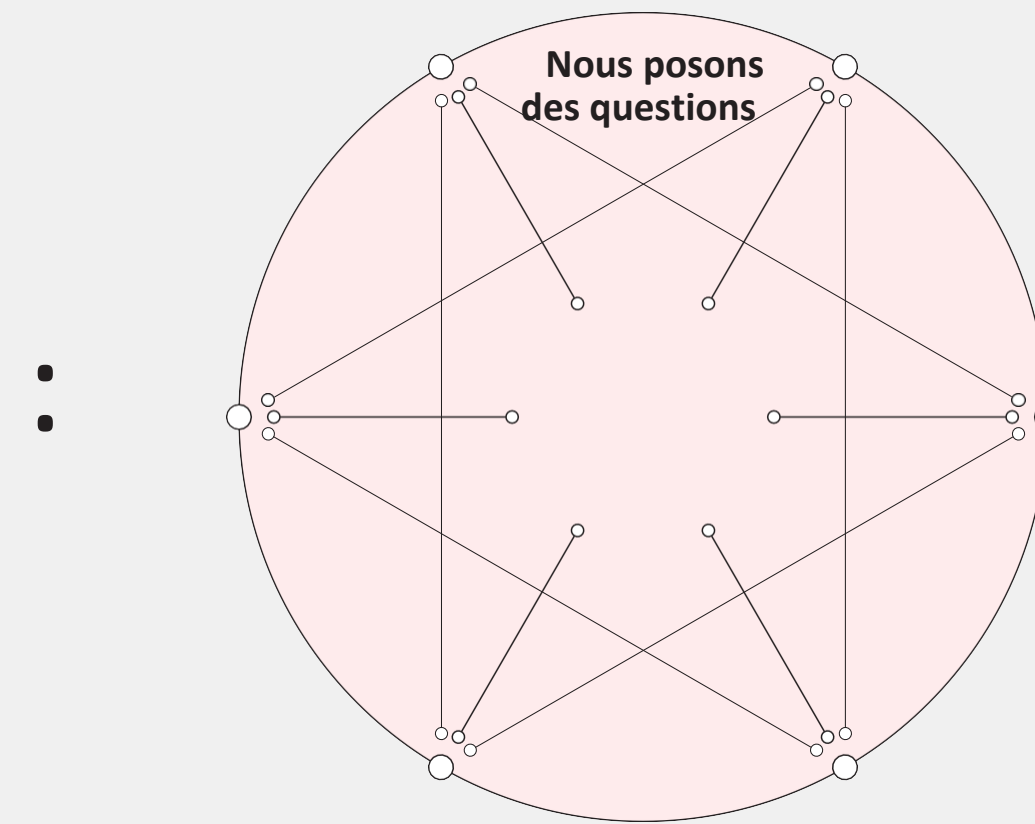
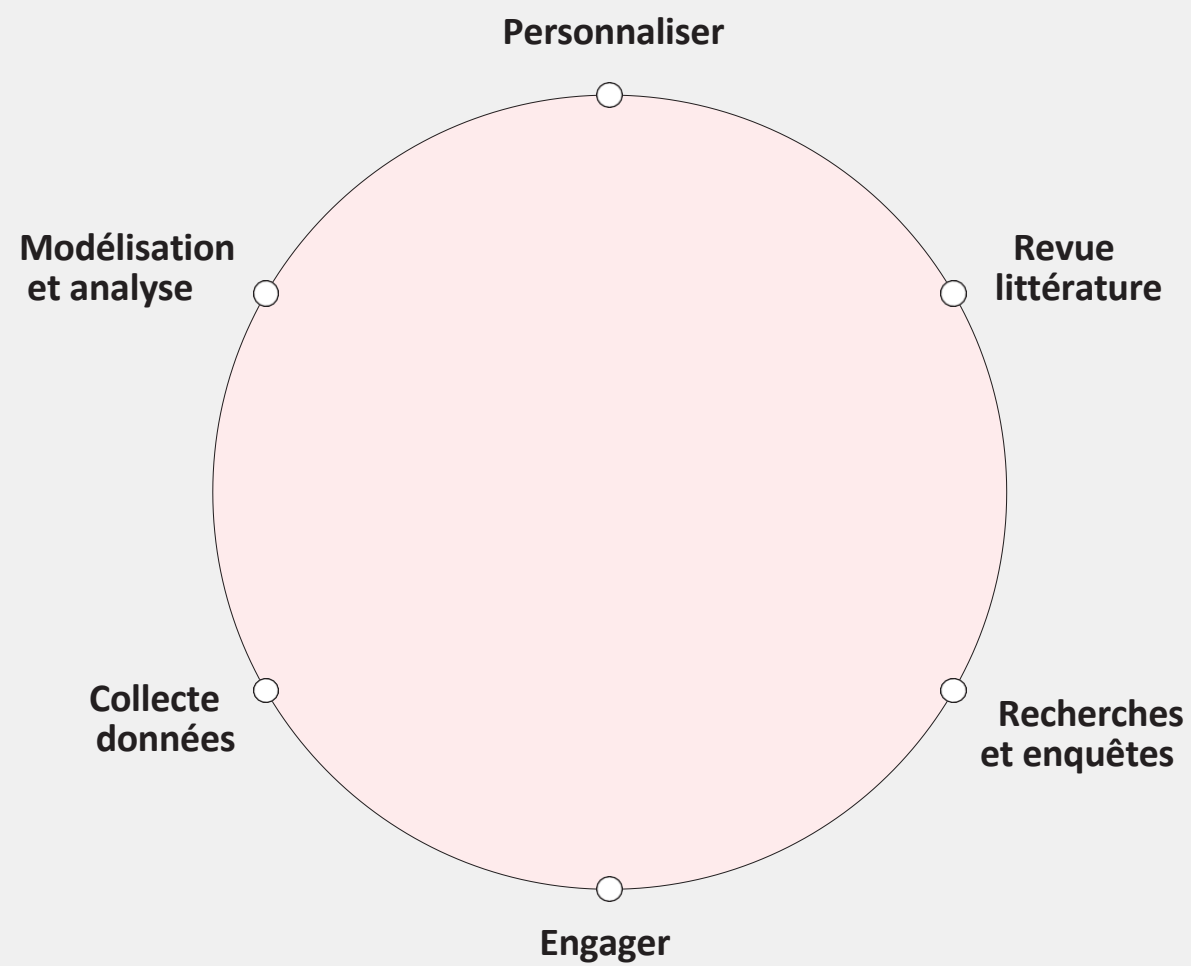
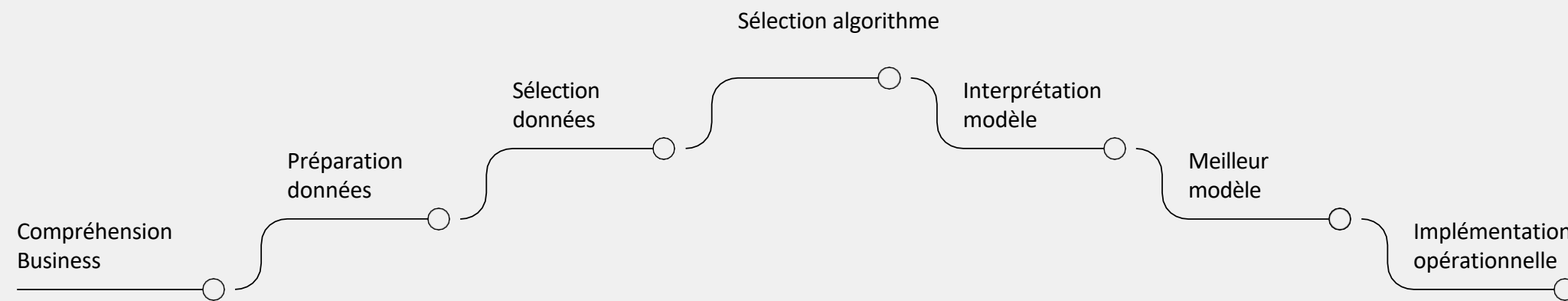


Exemple d'une plateforme de science des données typique (à des fins d'illustration uniquement)

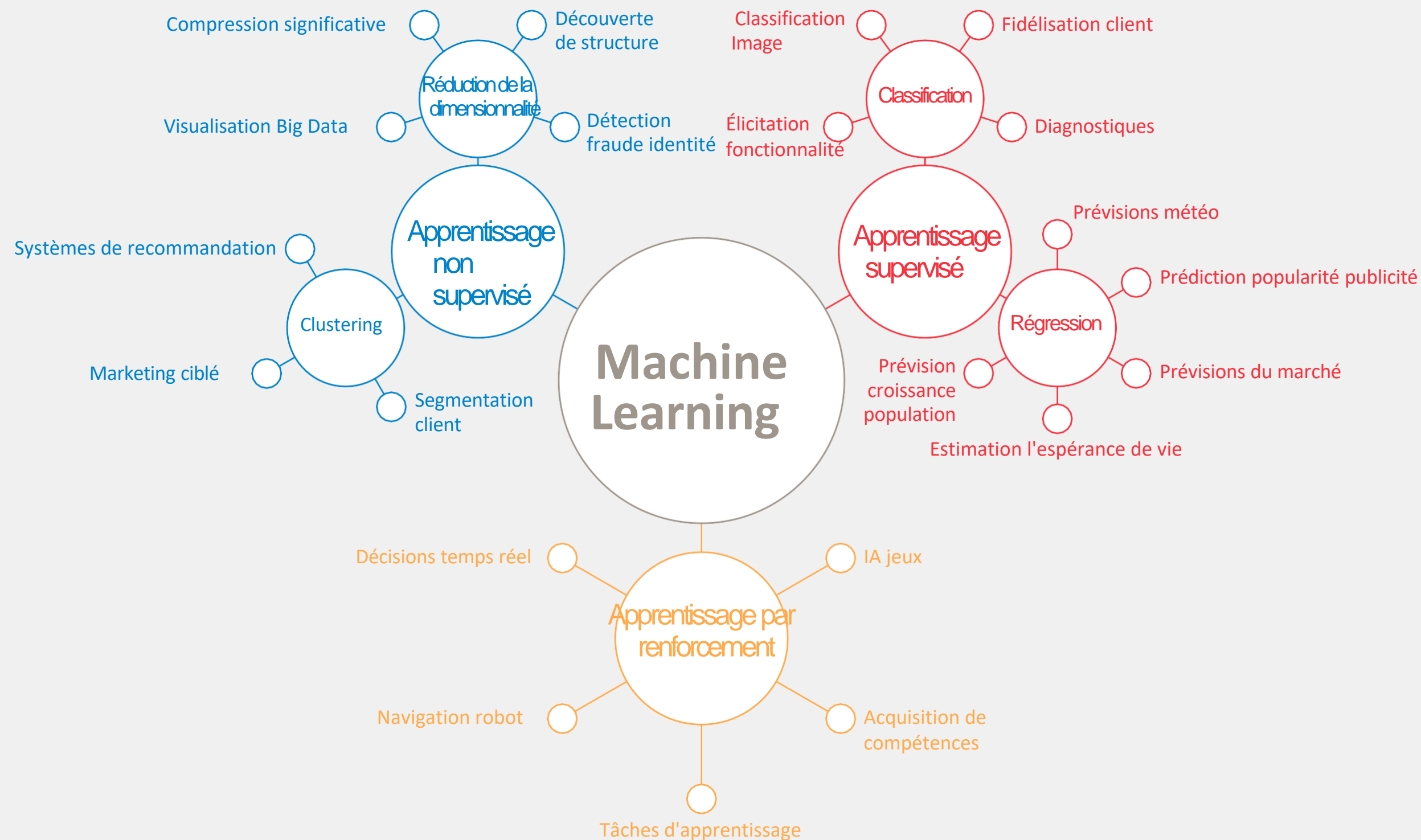
Lac de données + Machine Learning + IA + Visualisation (Simple, peu coûteux et conçu sur mesure)



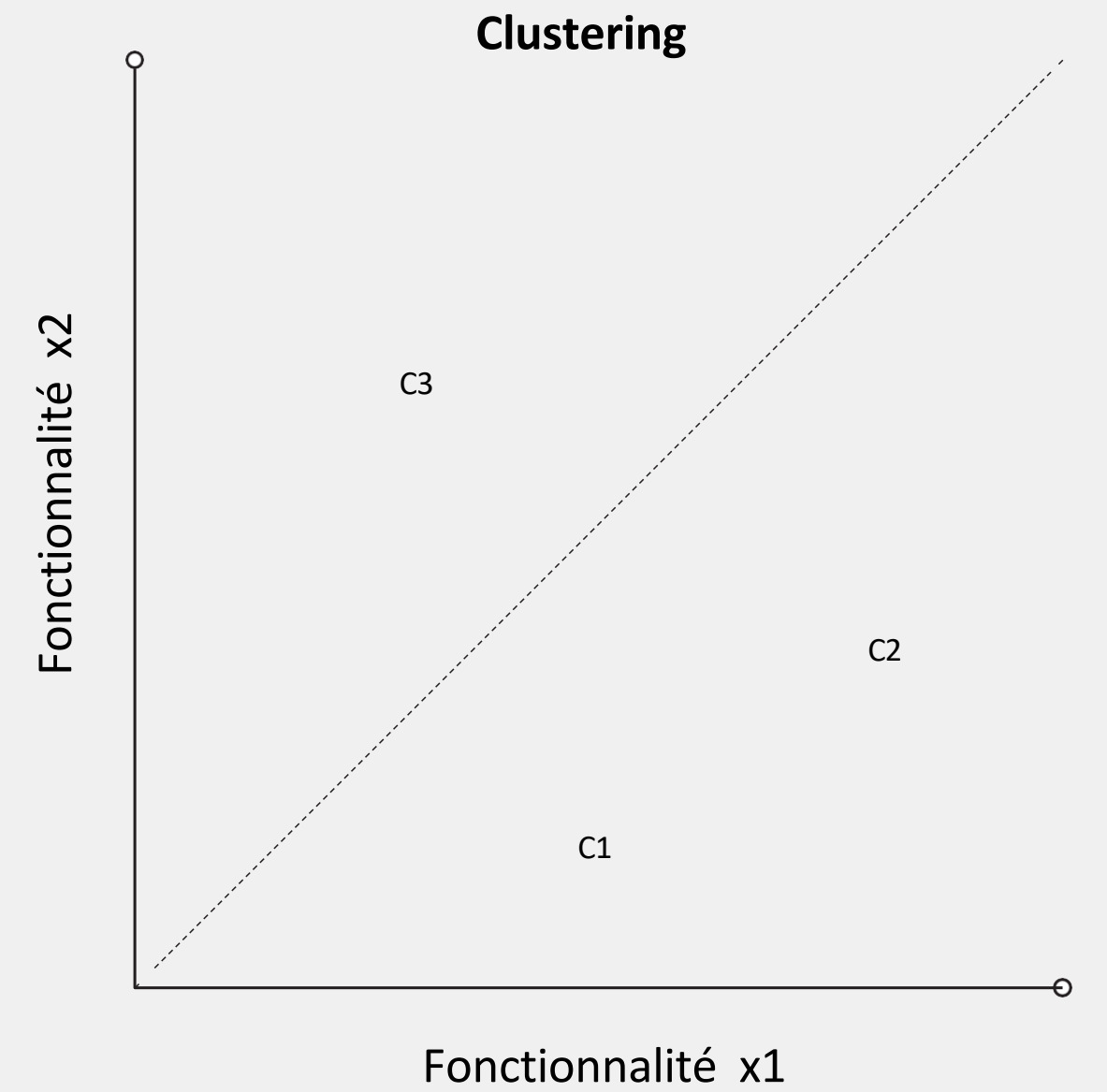
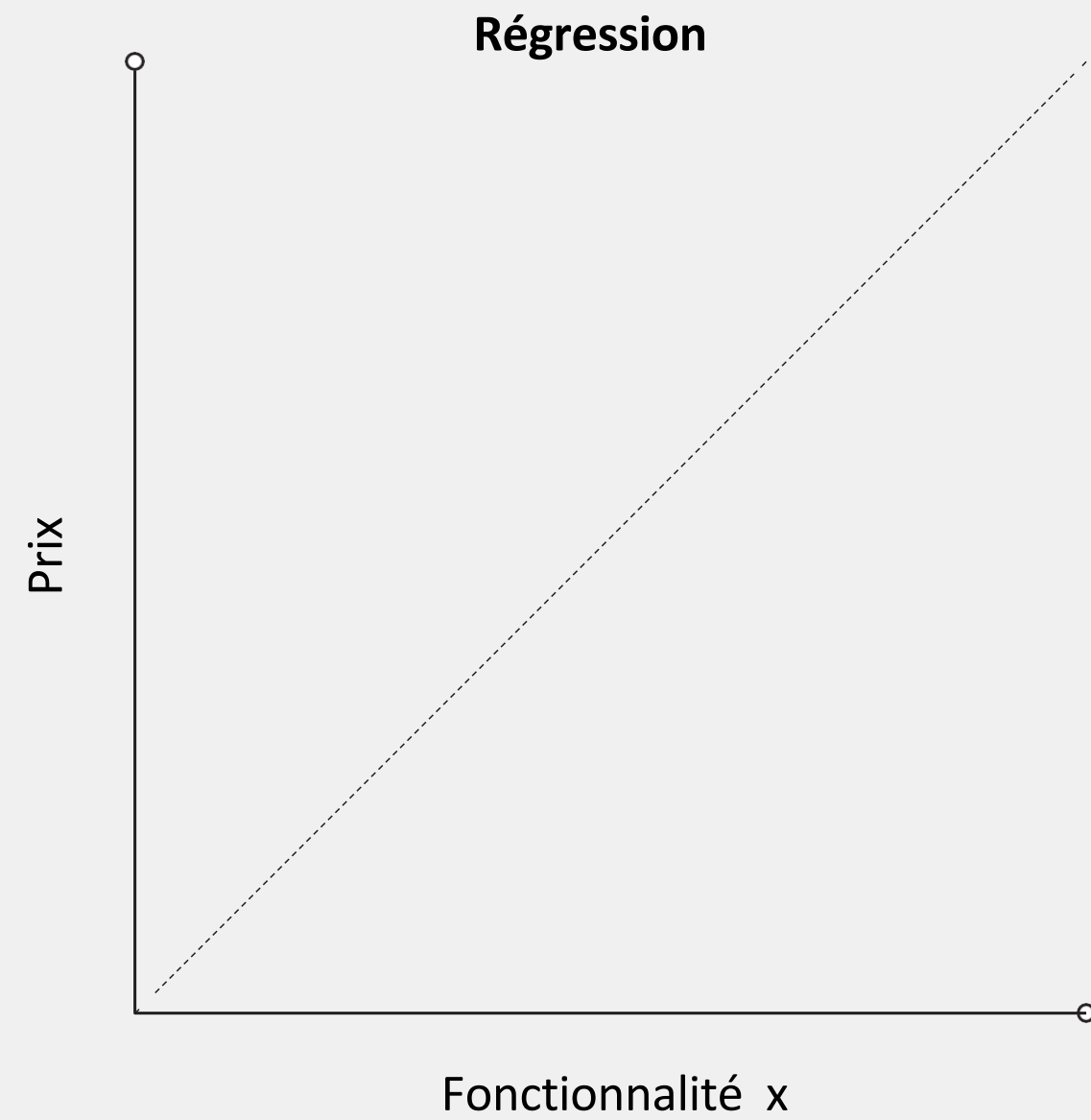
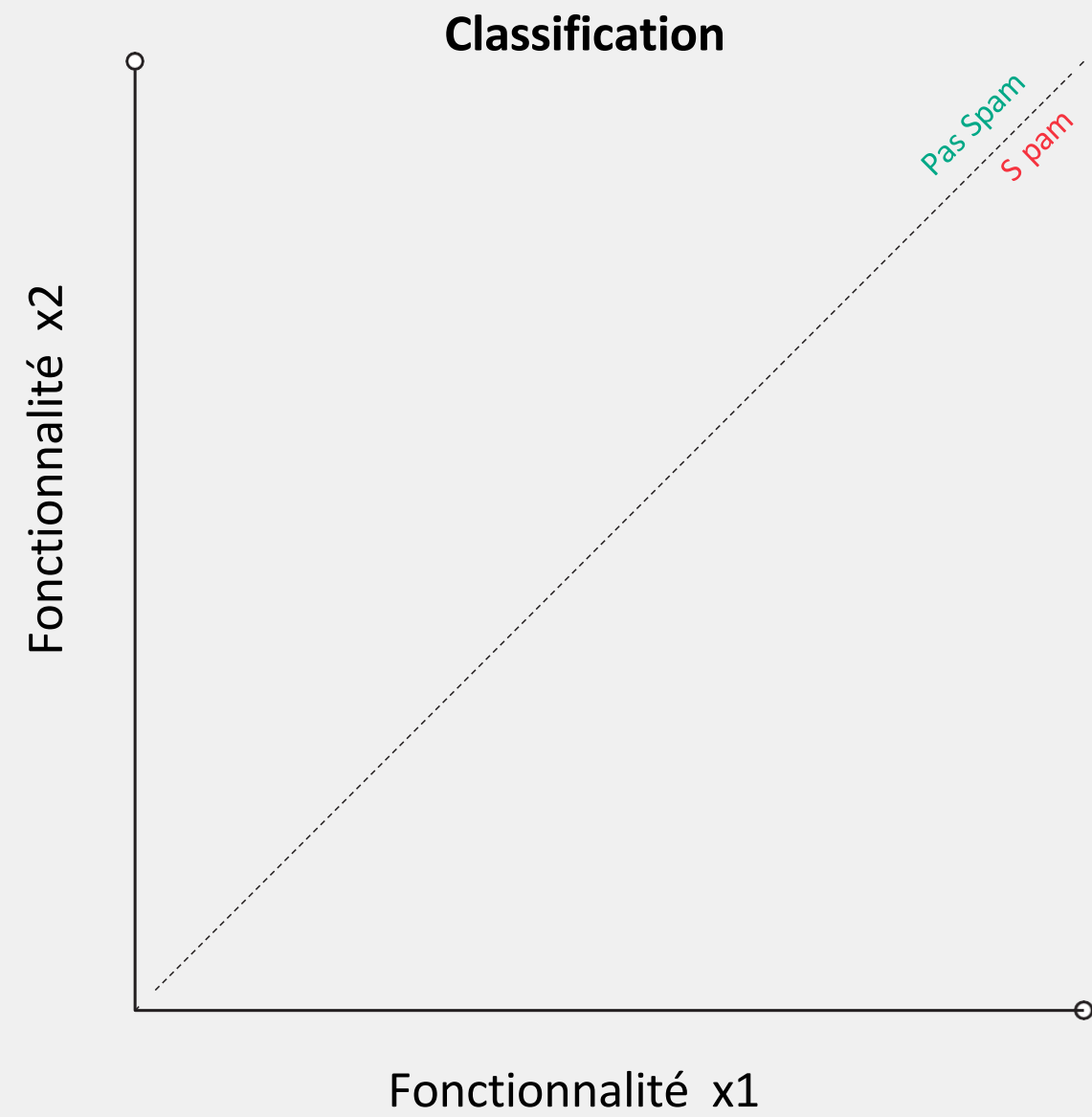
Processus de science des données (processus itératif et interactif)



Framework Machine Learning



Machine Learning : Vue d'ensemble



10 - 11 Comment mettre en œuvre un modèle de science des données

Cet exercice permettra aux équipes d'acquérir une expérience pratique du machine learning. En permettant aux équipes de mieux comprendre comment un modèle de science des données est mis en œuvre, elles peuvent mieux évaluer ce qui peut être intégré dans leur travail.

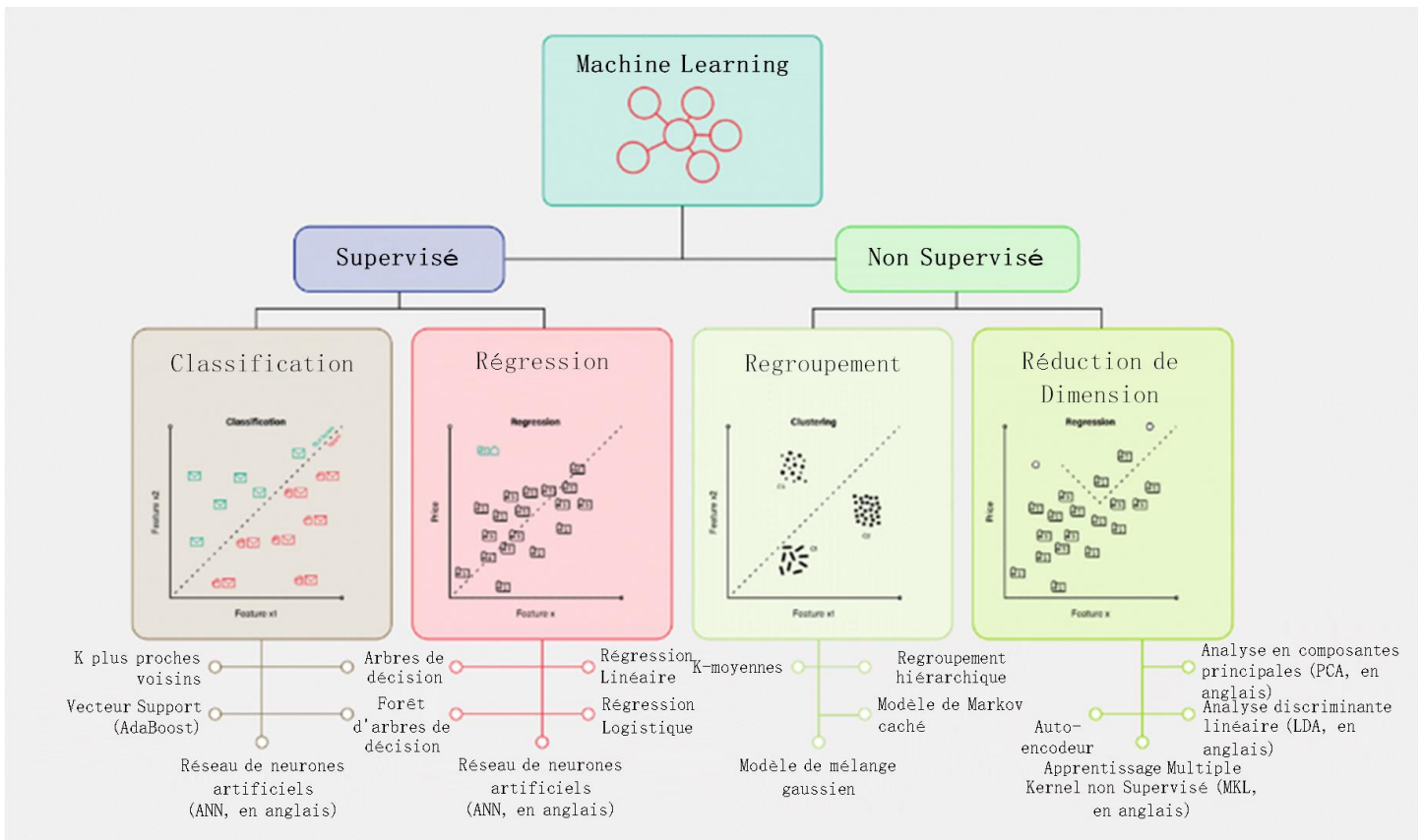
- ▶ **Personne** : 2 à 8 personnes
- ▶ **Durée** : 45 Minutes
- ▶ **Difficulté** : Moyen à Difficile
- ▶ **Matériel virtuel** : plateforme de réunion virtuelle, espace de rédaction/documentation partagé
- ▶ **Matériel en personne** : Tableaux, post-it, marqueurs

Exercice

Partie 1 : Explorer et réviser

Exercice pratique sur le machine learning (Comment implémenter un modèle de science des données)

- 1 Nous avons préparé un Notebook Jupyter pour illustrer un modèle simple de classification par machine learning. [Comment configurer Google Colab](#). Notez ce qu'est un modèle de classification dans le diagramme suivant.





- 2 Nous allons développer un modèle simple de désabonnement.
- 3 Comme il s'agit d'un modèle de classification binaire, nous devons connaître les étiquettes des donateurs avant de procéder à la modélisation. Dans ce cas, nous devons savoir qui sont les donateurs actifs et qui sont les donateurs non actifs. Nous avons entraîné ce modèle sur la base de données antérieures.
- 4 Notez ce qui suit
 - Phase de préparation des données
 - Phase d'échantillonnage
 - Phase d'analyse exploratoire des données
 - Répartition des données pour la formation et la validation
 - Entraînement du modèle (Note : le modèle de classification LightGBM)
 - Étape de validation (Noter quelle mesure a été utilisée pour valider la précision du modèle)
 - Analyse du modèle
- 5 Vous pouvez en savoir plus sur ce type de modélisation de machine learning :
 - <https://youtu.be/8b1JEDvenQU>
 - <https://www.youtube.com/watch?v=GM3CDQfQ4sw>
 - <https://www.youtube.com/watch?v=4jRBRDbJemM>

Partie 2 : Feedback

Demandez aux participants de réfléchir à l'exercice et de citer une ou deux observations ou idées.

Bonus

Pour de plus amples informations sur le sujet, vous pouvez consulter les liens suivants

- 1 <https://unstats.un.org/unsd/undataforum/blog/KITE-an-abstraction-framework-for-reducing-complexity-in-ai-governance/>
- 2 <https://hyperight.com/social-justice-and-sustainability-by-leveraging-data-science-and-ai-interview-with-dr-mahendra-samarawickrama/>

Crédit

Pour l'élaboration de l'exercice, Mahendra Smarawickrama et Paola Yela.

10 - 12 Analyse du taux de désabonnement de la Croix- Rouge - Jupyter Notebook

[Télécharger Ressources](#)

[Télécharger Ressources](#) (Modèle de données CSV)

[Télécharger Ressources](#) (Modèle de données SQL)