

# 10 Data Science and Emerging Technologies

---

# Table of Contents of the Module

<b>10</b>	<b>Data Science and Emerging Technologies</b>	<b>1</b>
<b>10 - 1</b>	Humanitarian objectives through data science	7
<b>10 - 2</b>	Role of a volunteer data science team	11
<b>10 - 3</b>	What a Data Scientist job description might look like	12
<b>10 - 4</b>	Sustainable implementation processes and improve maturity	13
<b>10 - 5</b>	Data, Emerging Technologies & Sustainable Goals	15
<b>10 - 6</b>	Social Diversity, Equity and Inclusion	21
<b>10 - 7</b>	Mapping Aspects of Data Science Work	26
<b>10 - 8</b>	Why Do We Need Data Science	29
<b>10 - 9</b>	Help Decisions Given Specific Settings	33
<b>10 - 10</b>	High-Level Overview of Typical Data Science Platform and Processes	38
<b>10 - 11</b>	How to Implement a Data Science Model	49
<b>10 - 12</b>	Red Cross Churn analysis Jupyter Notebook	52

# Data Science and Emerging Technologies

---

**This module introduces teams to basic concepts of data science and its connection to humanitarian objectives and goals. It is meant to guide teams who would like to explore the possibilities of Data Science and Emerging technologies in their daily work towards innovation and transformation.**

## Questions this module explores

- ▶ How can we effectively adopt and use data science and emerging Technology in our work?
- ▶ How can we understand and support the use of data science and emerging technologies in our work?

## Learning Objectives

- ▶ Explore how teams understand and use data science and emerging technologies in their work
- ▶ Guide teams towards critical success factors they need to know and key questions they need to ask when adoption data science projects
- ▶ Assess the required level of awareness about emerging technologies

## Module Topics

- ▶ Why use emerging technologies in our work
- ▶ How to use data and the fundamental principles to address strategic humanitarian challenges and transformations
- ▶ What are the initiatives and perspectives of data science and emerging technologies in our work

## Recipes

### A suggested step-by-step process to achieve learning objectives

- 1 Start with identifying **Humanitarian objectives through data science (10 - 1)** to get an understanding on why emerging technologies are introduced in humanitarian work.
- 2 Next discuss **Sustainable implementation processes and improve maturity (10 - 4)** that is illustrated in the emerging technologies learning canvas.
- 3 Take a closer look at **Data, Emerging Technologies & Sustainable Goals (10 - 5)** and how that is represented in the framework for sustainable goals
- 4 Allow groups to **Mapping Aspects of Data Science Work (10 - 7)** they find in their work and discuss why they need machine learning.
- 5 Guide participants on **How to Implement a Data Science Model (10 - 11)**.
- 6 Present the groups with **Help Decisions Given Specific Settings (10 - 9)**.

- In a group discussion, ask participants to share if they will be integrating data science projects to assist in their humanitarian work

## Ingredients

Pick and choose ingredients to create your own recipe. Do you have an ingredient we're missing? Send an email to [data.literacy@ifrc.org](mailto:data.literacy@ifrc.org)

## Exercises

### Presentations to be used and/or adapted:

- ▶ Opportunities in addressing humanitarian objectives through data science
- ▶ Framework for sustainability goals: a closer look at sustainability in AI
- ▶ Principles: social diversity, equity, inclusion in data science
- ▶ Mapping different aspects of data science work
- ▶ Why do we need machine learning
- ▶ Questions to help guide decisions on platforms, tools and skill selection
- ▶ How to implement a data science model

## Slide Decks

### Draft presentations to be revised/used:

- ▶ Sustainable implementation processes to improve maturity. Provides context on the emerging technologies learning canvas.

## Checklists/Handouts/Materials

### For documentation of essential elements of the learning experience:

- ▶ Role of a volunteer data science team (Canvas). Material to be used with **Why Do We Need Data Science (10 - 8)** (exercise).
- ▶ What a data scientist's job description should look like (Terms of Reference). Material to be used with **Why Do We Need Data Science (10 - 8)** (exercise).

- ▶ Data Science platforms (Slides). Provides context on granular components that are included in data science tools and their use. Material could be referred to in **Mapping Aspects of Data Science Work (10 - 7)** (exercise).
- ▶ Red cross churn analysis (Jupyter Notebook). Material to be used with **How to Implement a Data Science Model (10 - 11)** (exercise).

---

## External Resources

- ▶ [An Abstraction Framework for Reducing Complexity in AI Governance \(Article\)](#)
- ▶ [Knowledge sharing and collaboration: Example of data and digital week \(Blog Post\)](#)
- ▶ [Examples from Red Cross Red Crescent experiences: Go Github data science projects \(GitHub Repository\)](#)
- ▶ [Artificial Intelligence Suitability Framework \(Slides\)](#)
- ▶ 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.

## Next Steps

---

### Relevant modules in the Data Playbook

Refer to Module **Responsible Data Practices and Protecting Data (7)** to get a better understanding on how to mitigate risks and to the **Cognitive Biases in Humanitarian Analysis (6 - 3)** (Handout) in Module **Understanding and Analysing Data (6)**.

### Credit

---

Mahendra Samarawickrama, Jenny Paola Yela-Bello, and IFRC V1 Sprint and Data Playbook contributors

# 10 - 1 Humanitarian objectives through data science

---

When applying data and related emerging technologies on humanity, it is important that we have identified opportunities, ethics, values and social justice. In this perspective, teams need to ensure that they understand and grasp why data science and emerging technology is being used in their work. This exercise will allow teams to collectively look and discuss some frameworks. These frameworks assist teams capture opportunities that address specific humanitarian objectives within their work.

- ▶ **People:** 2 to 8 people
- ▶ **Time:** 40 Minutes
- ▶ **Difficulty:** Medium
- ▶ **Virtual Materials:** virtual meeting platform, shared document/writing space
- ▶ **In Person materials:** Flipcharts/noteboards, sticky notes, markers

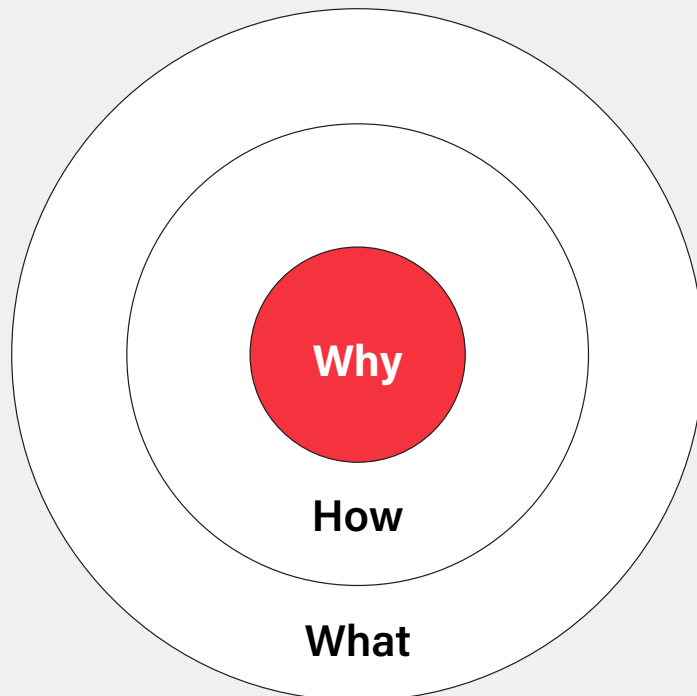
## Exercise

### Introduction

In this exercise we address why data, AI and emerging technologies are important to address in humanitarian perspectives. The goal of the exercise is to find answers to why data, AI, and emerging technologies should be considered as a component of strategy information, and why they play a big role in humanitarian work.

### Part 1: Explore

In teams please review the following



For this exercise, the IFRC strategy 2030 and the United Nations 17 Sustainable Development Goals are used as frameworks. Teams can work collectively or be split into



groups to discuss questions on why data science and emerging technology should be a key component in strategy formation. After discussions have been made, the facilitator can merge the ideas that were shared into a broader light on why the data science opportunities apply in their work and how they can be applied.

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

## Part 2: Discuss

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

This can be discussed for 10 minutes

[IFRC Strategy 2030](#)



## 3 Goals

**1**  
People anticipate, respond to, and quickly recover from crisis

**2**  
People lead safe, healthy and dignified lives and have opportunities to thrive

**3**  
People mobilise for inclusive and peaceful communities

## 5 Challenges

**1**  
Climate and environmental crisis

**2**  
Evolving crisis and disasters

**3**  
Growing gaps in health and well-being

**4**  
Migration and identity

**5**  
Values, power and inclusion

## 7 Transformations

**1**  
Supporting and developing National Societies as strong and effective local actors

**2**  
Inspiring and mobilising volunteerism

**3**  
Ensuring trust and accountability

**4**  
Working effectively as a distributed network

**5**  
Influencing humanitarian action

**6**  
Undergoing a digital transformation

**7**  
Financing the future

Question 2: Do you know the [United Nations 17 Sustainable Development Goals \(SDGs\)](#)? Do you know 79% of these goals can be supported by AI? Can you map some of them to our mission? The team can then discuss the diagram for 10 minutes.




### Credit

For development of the exercise Mahendra Samarawickrama

# 10 - 2 Role of a volunteer data science team

---

 [Download Resources](#)

# 10 - 3 What a Data Scientist job description might look like

---

 [Download Resources](#)

# 10 - 4 Sustainable implementation processes and improve maturity

# Emerging technologies learning canvas: Tool for briefing the business challenge

## Problem & opportunity

**Donor value proposition/project vision:** This defines 'what and why' we are doing this.  
**Donor problem:** what problems we are solving  
**Business Opportunity:** what value will this provide to the business

## Donor segments or personas

List your target donors & users for whom we are solving the problems and providing the solutions.

## Supporting context

**Existing research and insights:** key outputs from initial user testing, donor research or any qualitative research, surveys. Can also include any previous validated research from previous iterations. This supports the 'what and why' in above opportunity section.

## Goal

Success: What does success mean and how will we know we are delivering value? What is the key metric for success?

## Hypotheses

**Learning:** the idea or proposed solution that will deliver the value and take us close to the goal.  
 Which key strategies will drive the most value? Or what validated learning is important to confirm our path to success? This is the 'how' to the goals. A good hypotheses captures what needs to be tested through measurement and experimentation (to be proved or disproved)

## Timeline

What are the critical milestones?

## Team/stakeholders

Business Owner  
 Product Owner

# 10 - 5 Data, Emerging Technologies & Sustainable Goals

The Red Cross Red Crescent mission is to prevent or reduce human suffering, wherever it is found. We are part of the world's largest humanitarian movement, operating in over 192 countries. We have over 14.8 million volunteers worldwide. RCRC plays an auxiliary role to the government – recognised as a reliable partner for mobilising the power of humanity. With the implementation of the 2030 Strategy and increasing the use of digital tools in RCRC missions, there needs to be a closer look at how digital tools (Artificial Intelligence being one of them) can be sustained and reflective of the RCRC values, strengths and culture.

- ▶ **People:** 2 to 8 people
- ▶ **Time:** 60 Minutes
- ▶ **Difficulty:** Medium
- ▶ **Virtual Materials:** virtual meeting platform, shared document/writing space
- ▶ **In Person materials:** Flipcharts/noteboards, sticky notes, markers

## Exercise

In this exercise, the team with the guidance of the facilitator will talk about aspects that are important to highlight and work towards when driving AI towards sustainability.

### Part 1: Explore and Review

It will be a plenary discussion in 4 parts.

The session will cover 4 main points coupled with diagrams and an overview that can help lead the discussion.

### Why do Data Science and AI need to be governed ethically?

The Australian Red Cross is working on AI governance for sustainability because AI has a huge potential for humanity and significant risk if not properly governed. It is predicted that AI can contribute as much as 15.7 trillion to the world economy by 2030. At the same time, AI creates great risks for humanity such as autonomous weapons, automation-spurred job loss, socio-economic inequality, privacy violations, Deepfakes, and bias caused by data and algorithms. Further, it has been predicted that by 2022, 85% of AI projects will fail due to bias in data, algorithms, or the teams responsible for managing them.

One way of addressing sustainable AI for sustainability is to enhance AI ethics and governance while following best practices. Can you think of other ways that can be integrated and used?



**What are our values, purpose and strength when focusing on data science and AI? open the discussion after having a look at the diagram below.**

Please visit the whiteboard [animation](#)

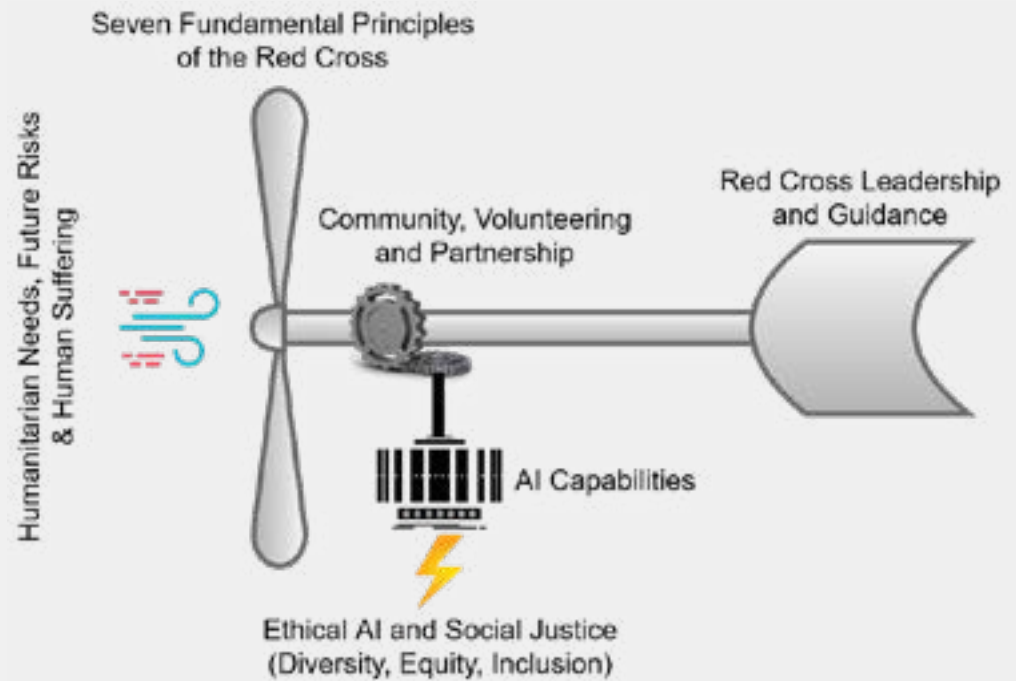


Fig.: A conceptualised model to drive AI ethically. (Source: <https://hyperight.com/social-justice-and-sustainability-by-leveraging-data-science-and-ai-interview-with-dr-mahendra-samarawickrama/>)

People, culture and Mission in AI. What can we reflect about AI based on our people, culture and mission? Discuss these attributes and the illustrated figure below.

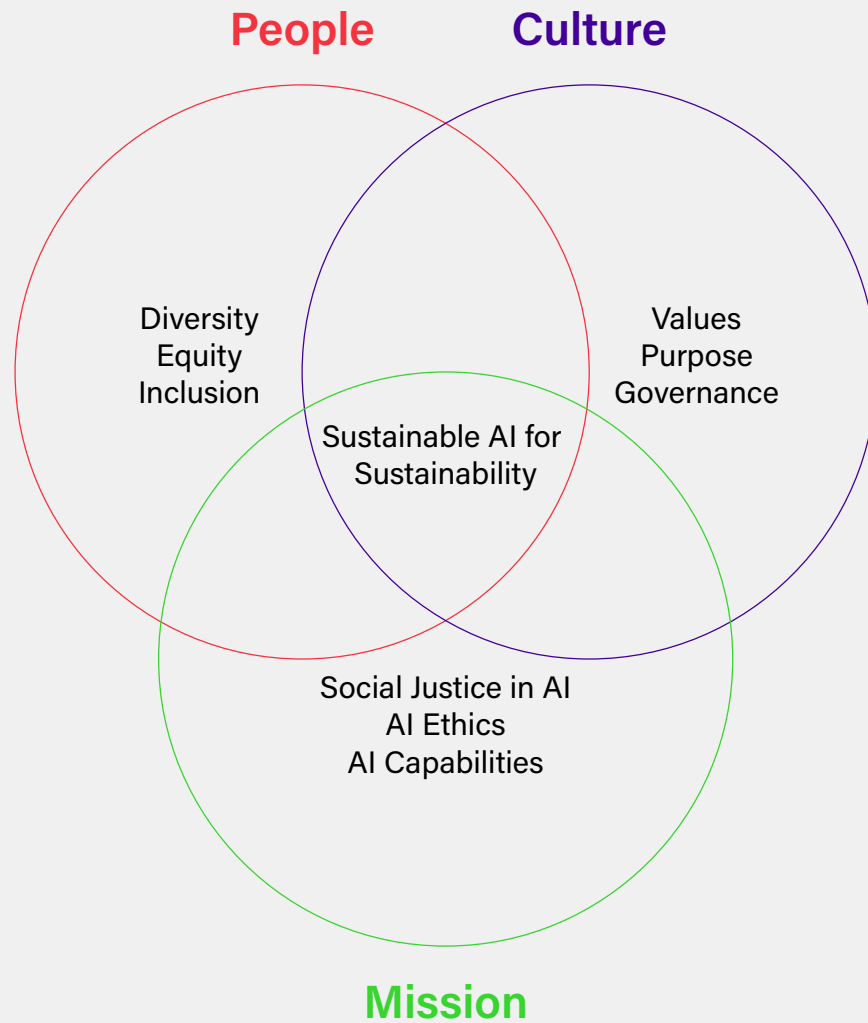


Fig.: The aspects of people, culture and mission on sustainable AI for sustainability.

How can we make Data Science and AI as core competency?  
 Discuss ways of integrating AI as part of the core competencies, while looking at the diagram below.



Fig.: Develop AI as a core competency to drive business.

## Part 2: Feedback

Ask people to reflect on the exercise and cite 1 or 2 observations or insights.

### Extra credit

---

The Australian Red Cross keynotes related to AI ethics and governance:

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](#)

### Credit

---

For developing the exercise Mahendra Smarawickrama and Paola Yela.

# 10 - 6 Social Diversity, Equity and Inclusion

---

To address social diversity, equity and inclusion, there needs to be a wider understanding on emerging technology governance and how that plays a role in sustainability and humanity within our network.

- ▶ **People:** 2 to 8 people
- ▶ **Time:** 30 Minutes
- ▶ **Difficulty:** Medium to Hard
- ▶ **Virtual Materials:** virtual meeting platform, shared document/writing space
- ▶ **In Person materials:** Flipcharts/noteboards, sticky notes, markers

## Exercise

### Part 1: Explore and Review

#### AI and Emerging Tech Governance for Sustainability and Humanity

When bringing diversity, equity and inclusion (DEI) to data and emerging technologies, leaders and contributors should be aware of different perspectives of why, how and what.

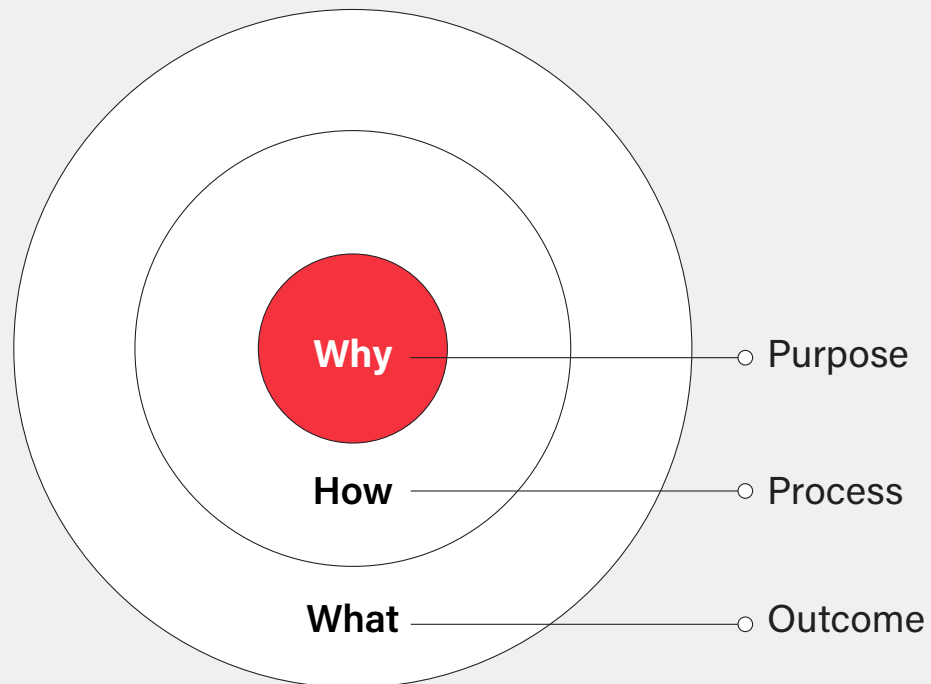
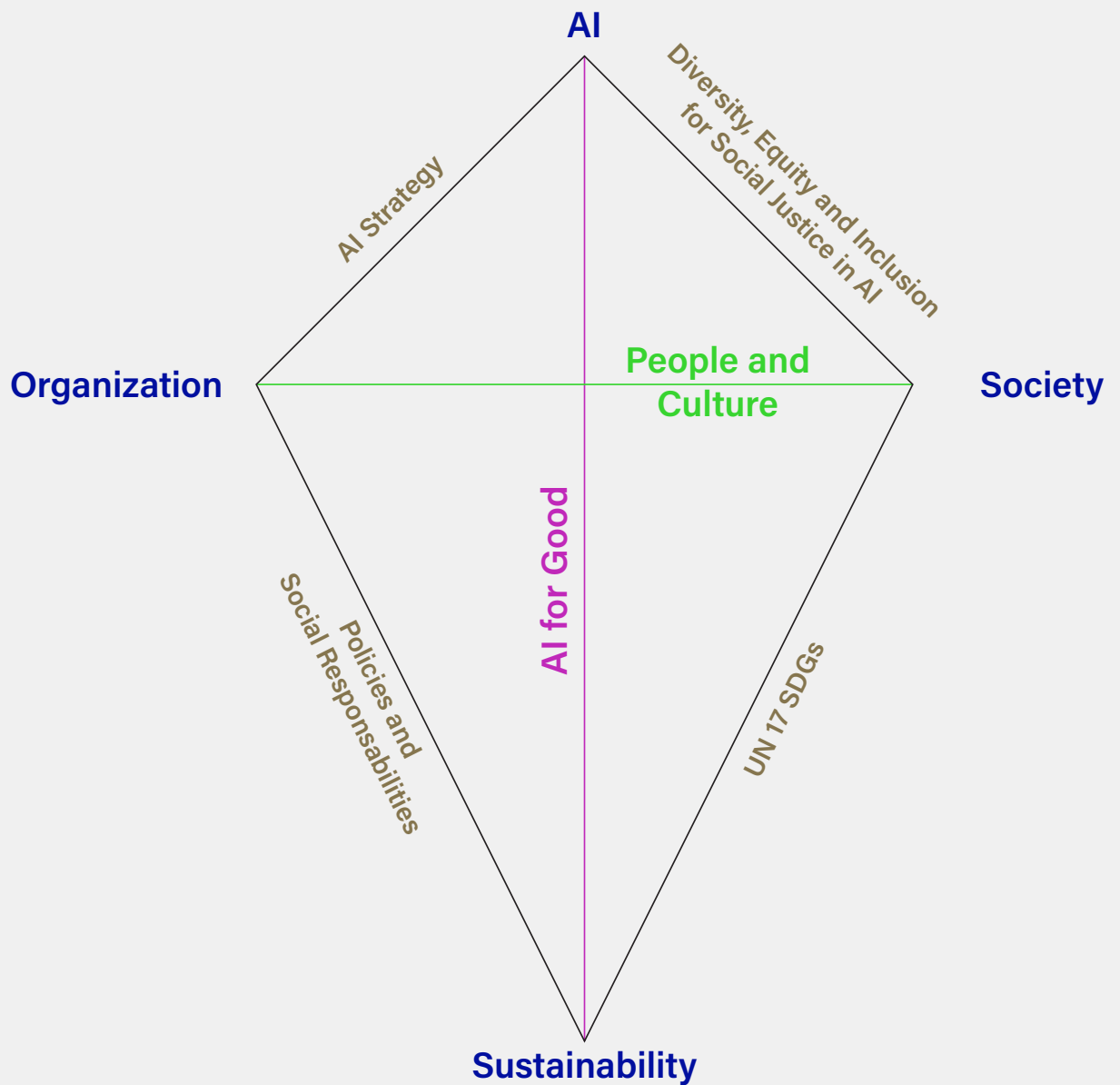


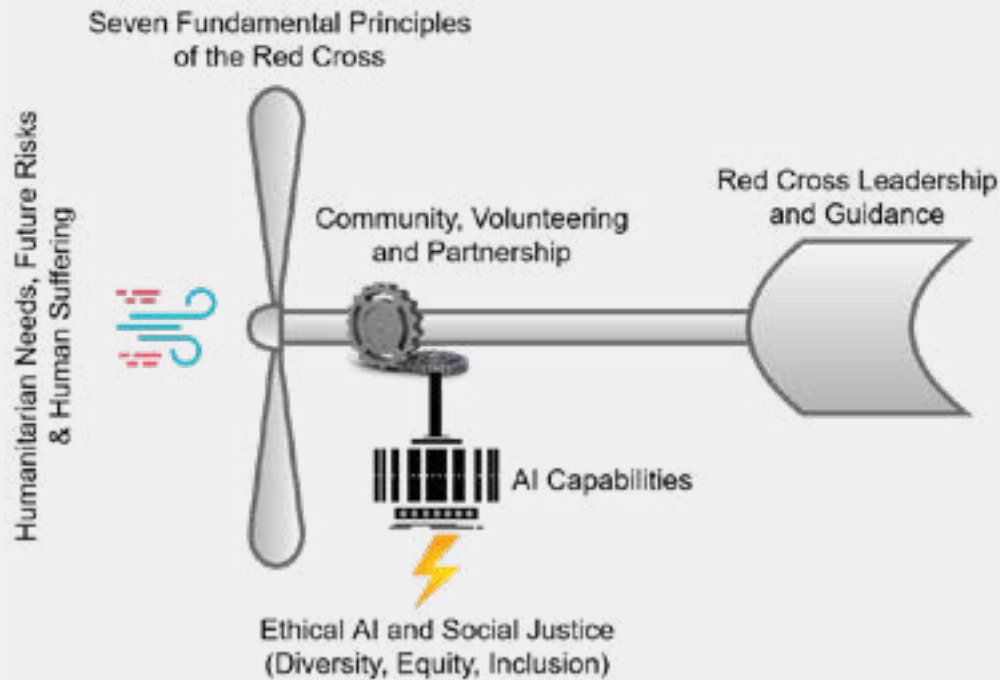
Fig: Golden Circle

For success in DEI, understanding the importance of people, culture and mission towards emerging technologies and mitigating the risk of failure due to biases are important. Please refer to the following whiteboard animations:

What are the important aspects of emerging technologies and AI governance? (Understanding of “Why” questions; such as why data and AI play a key role and why AI governance and sustainability need to be considered). Please, play: [The KITE Abstraction Framework for AI Governance and ESG for Sustainability](#).



**How to bring the Red Cross values to emerging technologies and AI governance? (Understanding “What” and “How” questions). Please play the Wind Turbine conceptual model whiteboard animation.**



**Can you discuss how social diversity, equity and inclusion is brought to your emerging technologies or AI project? How confident are you about the social justice of your project? Did you contribute to sustainability (i.e., social or environmental)?**

- ▶ Please check the following success story recently presented in IFRC Planet:Red Summit?
- ▶ Session Link: <https://planetredsummit.com/session/UJSGJU-1>
- ▶ Streamed Video session: <https://youtu.be/psySmdBc0IE>
- ▶ Please share your story as well.
- ▶ Additional resource: Samarawickrama, M. (2022). [Keeping AI Honest. Company Director, 38\(2\), 52-53](#)



Your Society	Public link to the Success Story

## Part 2: Feedback

Ask people to reflect on the exercise and cite 1 or 2 observations or insights

### Extra credit

How to bring social diversity, equity and inclusion from volunteers.

- ▶ If time is available we can talk about the formation of a volunteer data science team in your society. Please use the Emerging Tech Innovation Canvas for this.
- ▶ Please use the canvas populated by the Australian Red Cross to establish their diversified volunteer data science team.
- ▶ [Canvas](#)
- ▶ [Job description](#)
- ▶ [Australian Red Cross Volunteer Data science Team](#)

This could be used with Module **Responsible Data Practices and Protecting Data (7)** and **Humanitarian Values & Data Protection (handout) (7 - 8)** (exercise).

### Credit

Mahendra Samarawickrama and Paola Yela for developing the exercise

# 10 - 7 Mapping Aspects of Data Science Work

---

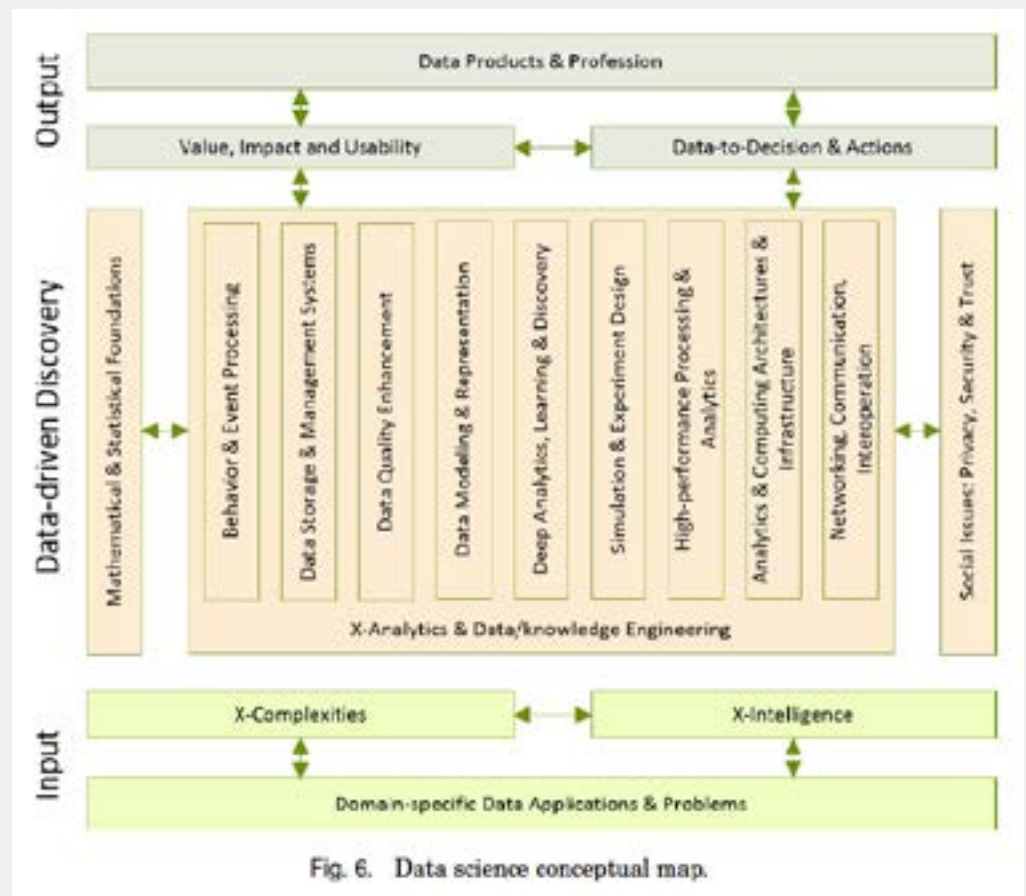
**Clarifying what we understand by data science. Mapping different aspects of the data science work, why do we need data science and illustrating the type of products/analysis we can expect from it. This exercise allows teams to verify if their conceptualization of data science is aligned. By the end of this session, we hope everybody will have an overview of data science, its different perspectives, and can understand their role in the data science work spectrum.**

- ▶ **People:** 2 to 8 people
- ▶ **Time:** 30 Minutes
- ▶ **Difficulty:** Easy to Medium
- ▶ **Virtual Materials:** virtual meeting platform, shared document/writing space
- ▶ **In Person materials:** Flipcharts/noteboards, sticky notes, markers

## Exercise

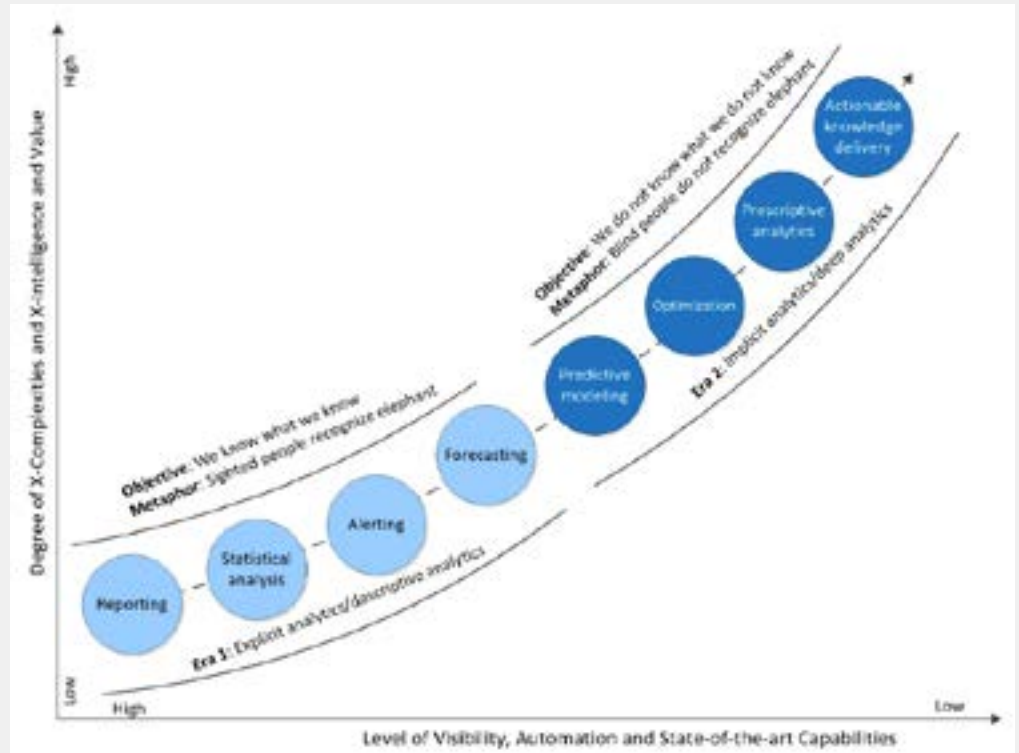
### Part 1: Explore and Review

Please discuss how a data science team can be established by referring to the below diagram and where do their roles best fit in the spectrum?



(i) Taken from here: Source: <https://arxiv.org/abs/2007.03606>

Discuss the main components of data science illustrated by the following diagram and reflect back on whether these match the needs of your National Society in forming their data team. ([Data Science: A Comprehensive Overview](#))



## Part 2: Feedback

Ask people to reflect on the exercise and cite 1 or 2 observations or insights.

### Extra credit

The facilitator can follow up on this discussion by referring to Why do we need Data Science exercise in Module 10 through linking the key takeaways from this session to exploring the capabilities of a data science team.

### Credit

For developing the exercise Mahendra Samarawickrama and Paola Yela

# 10 - 8 Why Do We Need Data Science

---

**Clarifying what we understand by data science. Mapping different aspects of the data science work, why do we need data science and illustrating the type of products/analysis we can expect from it. This exercise explores the need of establishing a data science team and helps uncover some need-to-know key aspects when working with data science teams at RCRC begins.**

- ▶ **People:** 2 to 8 people
- ▶ **Time:** 45 – 60 Minutes
- ▶ **Difficulty:** Medium to Hard
- ▶ **Materials in person:** Flipchart paper and markers
- ▶ **Virtual Material:** Virtual meeting platform, shared document/writing space

## Exercise

### Part 1: Explore and Review

Case study: The senior leadership team of the ABC Red Cross society decided to establish a data science team that can drive data science and analytics to drive business.

Referring to the scenario, allow groups to discuss a series of questions that can help them explore the need of establishing a data science team and if that brings them a step closer to identifying the use of machine learning in the data science spectrum.

### Questions that will be used for discussion:

- 1 What is data science?
  - (prompt for facilitator - answers can include: data science = statistics + informatics + computing + communication + sociology + management | data + environment + thinking)
- 2 Why does senior leadership need a data science team?
  - (prompt for facilitator - answers can include: to assist data informed decision making, to personalize customer experience, to assist in making predictions and risk management, to diversify the services and strategy, etc.)
- 3 What are the current capabilities that the senior leadership team should be aware of?
  - (prompt for facilitator - answers can include: platforms, skills of people, tools being used, the data culture, etc.)
- 4 What investments the leadership team should be ready to spend on?
  - (prompt for facilitator - answers can include: training staff, acquiring technology such as tools and platforms, etc.)
- 5 What skills does the leadership team include in the new job description?

- (prompt for facilitator - answers can include: Red Cross values, communication skills, machine learning skills, knowledge in statistics, knowledge in Python and R, etc.)
- 6 What would be the team composition?
    - (prompt for facilitator - answers can include: gender and diversity in the team, try and promote women in AI, etc.)
  - 7 Other than the fundraising campaign, what are the socially impactful data sets that Red Cross is interested in?
    - (prompt for facilitator - answers can include: climate change data, migration data, health and wellbeing data, volunteer data, etc. -Can refer to the Red Cross Strategy 2030)
  - 8 What risk management was currently put in place to manage the data risks?
    - (prompt for facilitator - answers can include: IT cyber security policy, IT data governance policies, data compliance guidelines, legal compliance, etc.)
  - 9 Your society at the moment is using Tableau, Power-BI or any other business intelligence tool. At which stage you are on the data science road map?
    - (prompt for facilitator - answers can include: reporting and exploratory data analytics, etc.)
  - 10 Which stages of data road map machine learning bring to the organisation?
    - (prompt for facilitator - answers can include: predictive and prescriptive analytics, etc.)
  - 11 What does machine learning do?
    - (prompt for facilitator - answers can include: Identify patterns in the data and create a model in which we can make predictions in advance)

---

## Part 2: Feedback

Ask people to reflect on the exercise and cite 1 or 2 observations or insights.

### Extra credit

---

To dive deeper into the integration of machine learning in data science projects, the groups have the option to answer the following questions:

- 1 Have you heard about machine learning algorithms?

- 2 Can you explain them to a non-technical person?
- 3 What does garbage in garbage out means in data science modelling?
- 4 What is a hypothesis in Data Science modelling?
- 5 Can you explain the bias and variance in data science modelling using the hypothesis?

Option for the facilitator to refer to the Role of a volunteer data science team and What a Data Scientist job description should look like additional material provided in Module 10.

### Credit

---

For developing the exercise Mahendra Samarawickrama and Paola Yela, Edited by Melissa El Hamouch

Extra resource: Eggleton, M., 2022. Award-winner warns of the failures of artificial intelligence. [The Australian Financial Review - Special Report of Reimagination 2022](#), pp.S4-S5.



# 10 - 9 Help Decisions Given Specific Settings

---

**When various options of data science processes are available, it is not always easy to identify which option is the best to implement for a given project. This exercise will help teams discuss their available options based on their resources and capacities.**

- ▶ **People:** 2 to 8 people
- ▶ **Time:** 45 – 60 Minutes
- ▶ **Difficulty:** Hard
- ▶ **Virtual Materials:** virtual meeting platform, shared document/writing space
- ▶ **In Person materials:** Flipcharts/noteboards, sticky notes, markers

## Exercise

### Part 1: Explore and Review

This exercise allows teams to answer questions about their resources, capacities, and time in order to understand what is the most suitable option for them to select for a particular objective. The facilitator can either come up with a theoretical project task or discuss one that the team proposes.

Note: Have teams write down different options that can be cross-checked against the resources they have and the situation in which they are in.

The facilitator can then guide the team in identifying what option will work best for their situation. A list of questions can be asked to help groups and teams think along.

#### Questions can be as follows:

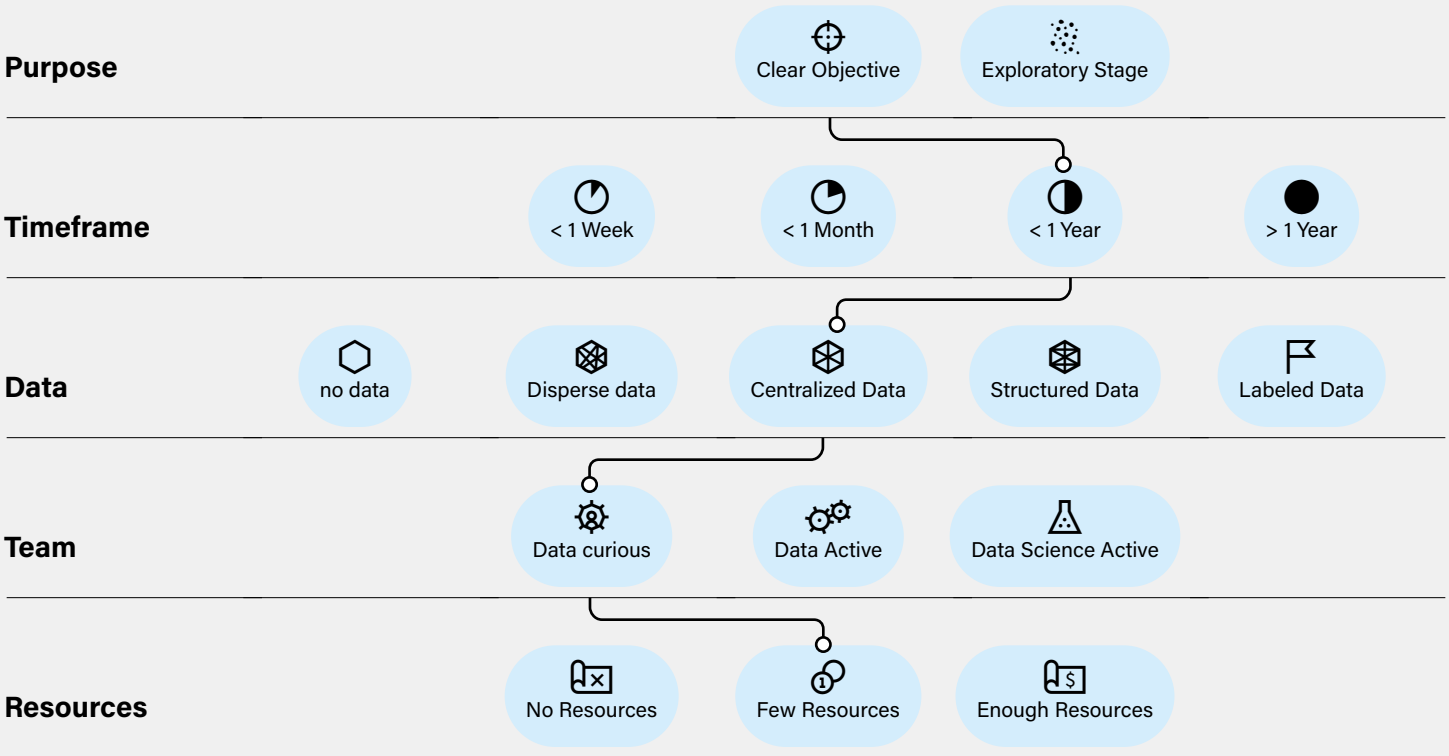
- ▶ How many resources do you have?
- ▶ How much time is available?
- ▶ How much time will it take to work on the project?
- ▶ What potential constraints might be faced?
- ▶ What is the budget allocated for the project?
- ▶ How much staff is available to work on the project?
- ▶ What skills are needed from the team?
- ▶ What is the setting of which the work will take place in (emergency setting or not)?
- ▶ What data does the team have (historic data, secondary data, no data, etc.) and what data do they need to have?

The facilitator has the option to present the team with the below cases as a starting point for discussions. They can then refer back to the set of questions presented above to get a more indepth view from the team and assess what is needed and what is available. This will help them in deciding what data science process and resources will be most efficient.

#### CASE A:

The National Society XYZ has a repository of community risk assessments that were done in several regions within their country. The assessments are unstructured documents that have similar content but are in varied formats.

The National Society risk management team would like to get key information from those documents, however they don't have the internal capacity to process all the documents. It is not an urgent task, but they have some resources that could be invested for this task.



The team in National Society XYZ decided to go for a 4 months internship at a technical institution. Would your team do the same? Which option would you have preferred?

## CASE B:

Please select your situational path:

### Purpose



Clear Objective



Exploratory Stage

### Timeframe



&lt; 1 Week



&lt; 1 Month



&lt; 1 Year



&gt; 1 Year

### Data



no data



Disperse data



Centralized Data



Structured Data



Labeled Data

### Team



Data curious



Data Active



Data Science Active

### Resources



No Resources



Few Resources



Enough Resources

Which option would you consider? Which other options would you propose for such a situation?

Internships Research partnerships Consultancies Hackathons Internal data

collection project Internal DS project Movement support Tech partnerships

Internal labeling project Data partnerships Engage volunteers Working groups



## Part 2: Feedback

Ask people to reflect on the exercise and cite 1 or 2 observations or insights.

### Extra credit

---

For further reading, teams can view the [AI Suitability Framework](#) which provides a set of questions that should be asked through the development of a data science project. It brings enough details and tries to preserve a practical perspective. The above represented graph can be seen as a first approach to it.

### Credit

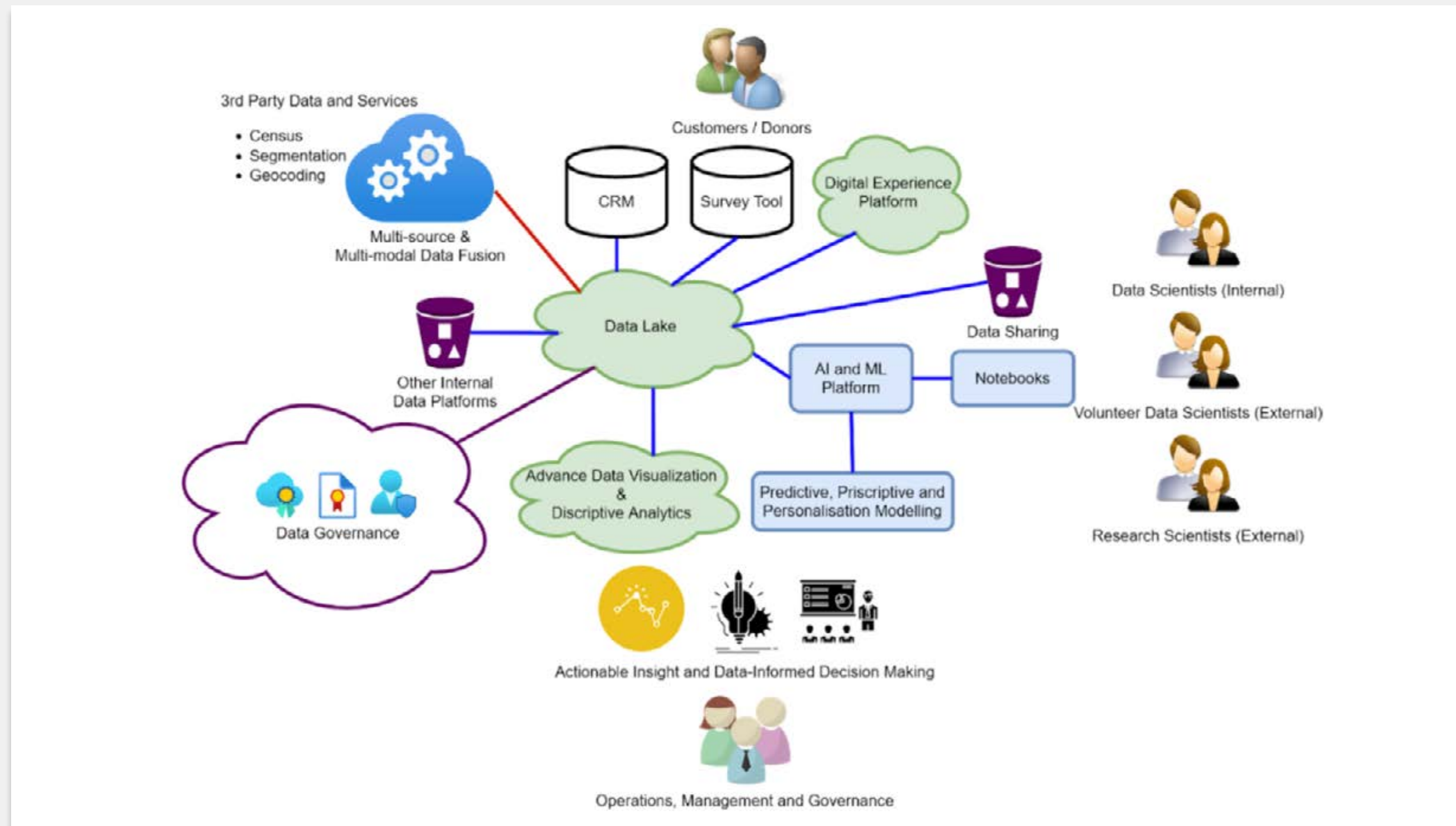
---

For exercise development Paola Yela and Melissa El Hamouch

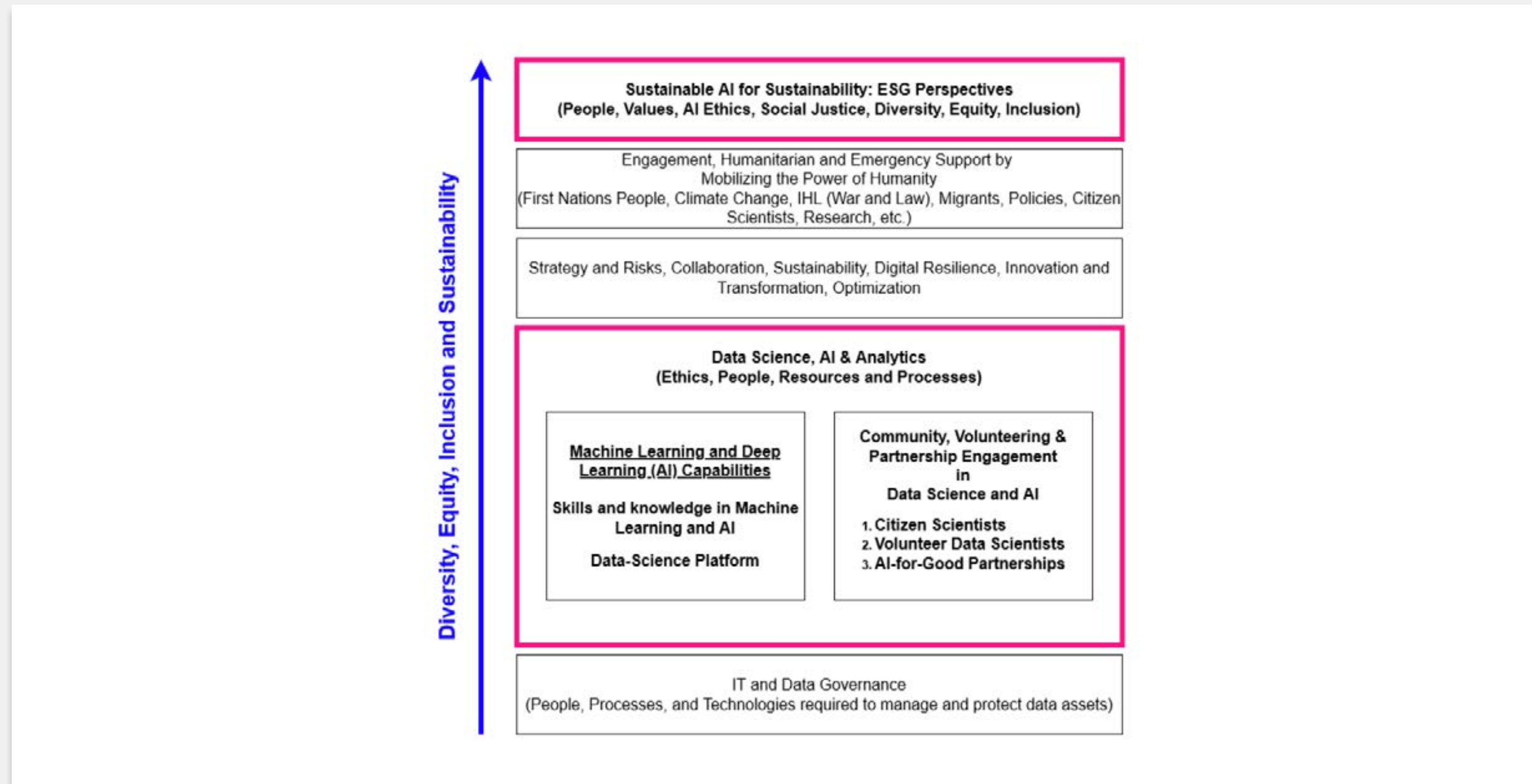
# 10 - 10 High-Level Overview of Typical Data Science Platform and Processes

Credit: Mahendra Samarawickrama for the IFRC Data Playbook

# Typical Data Science Platform Schematic

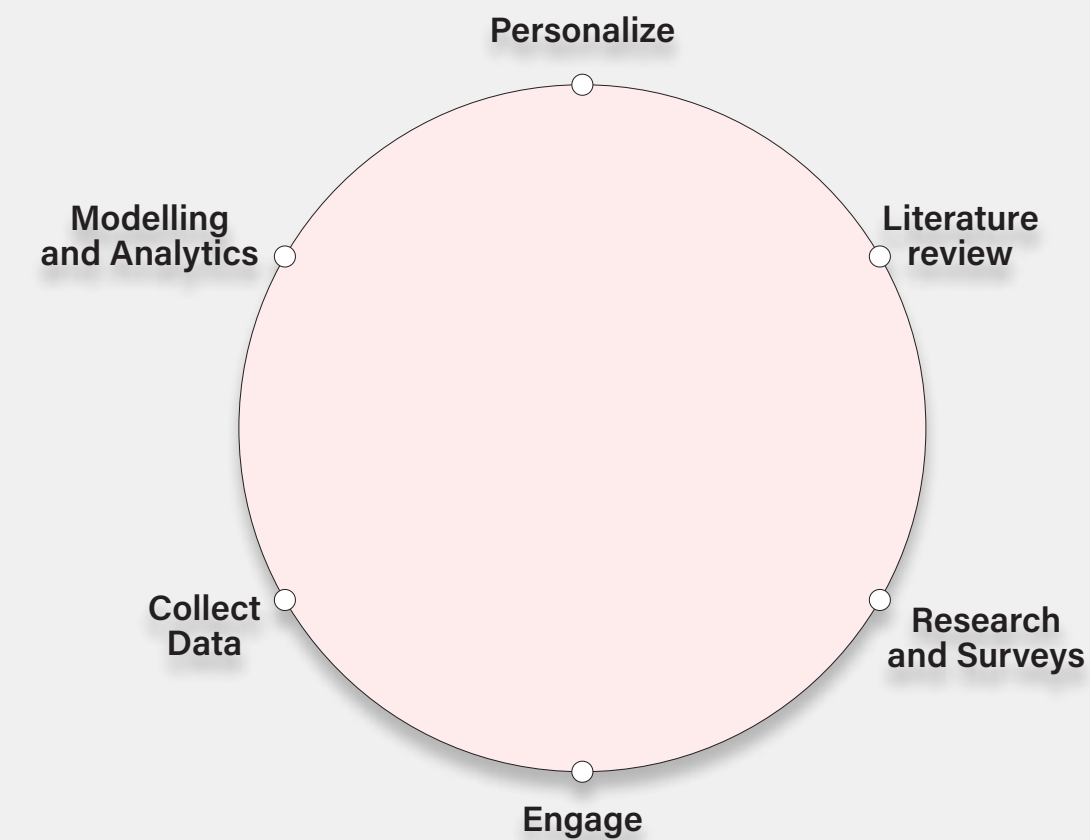
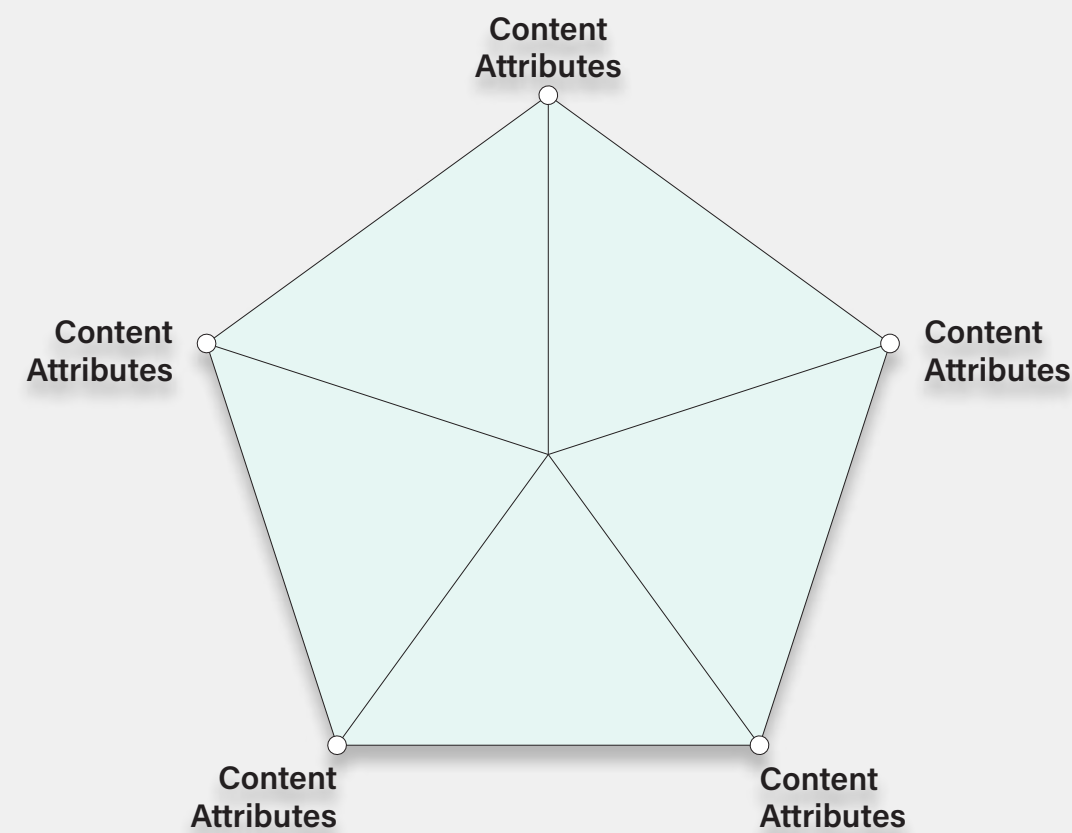
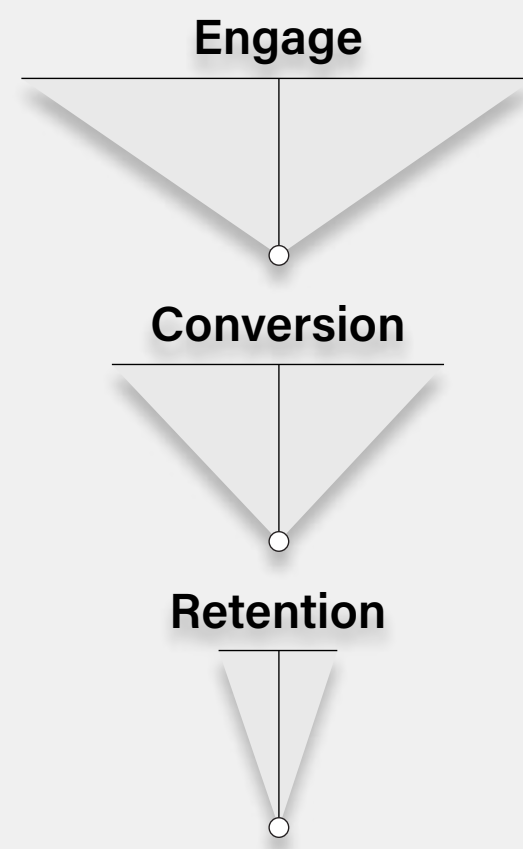


# Typical Data Science Process Focusing on Humanitarian Goals





# Data Science for Personalization



## Engagement and Support

CX

Acquisition And Loyalty

Digital

Marketing

Finance

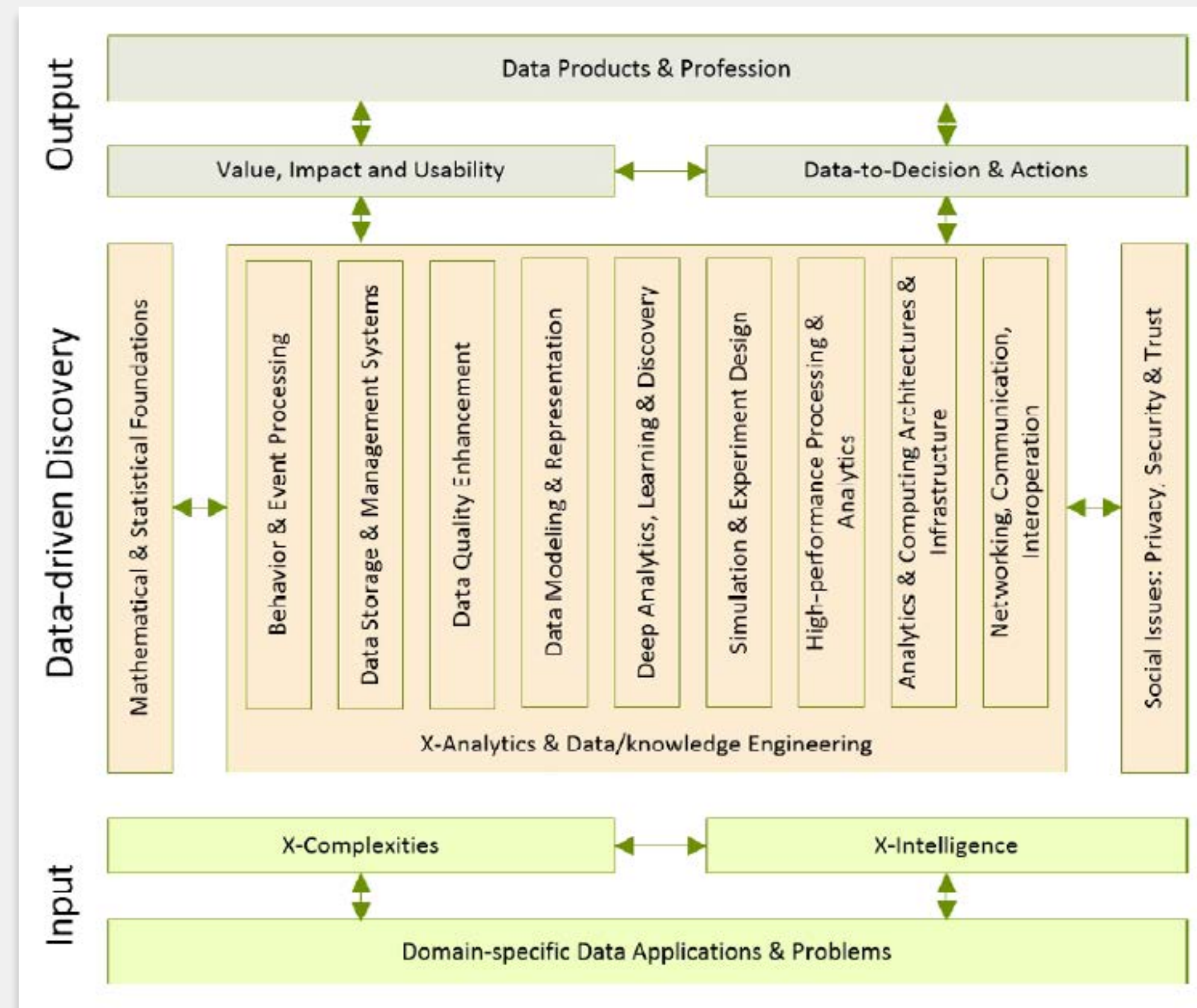
## Data Insight as a Service

Data Science (E.g., Personalization, Prediction, Advance Analytics)

Data Architecture (E.g., Infrastructure)

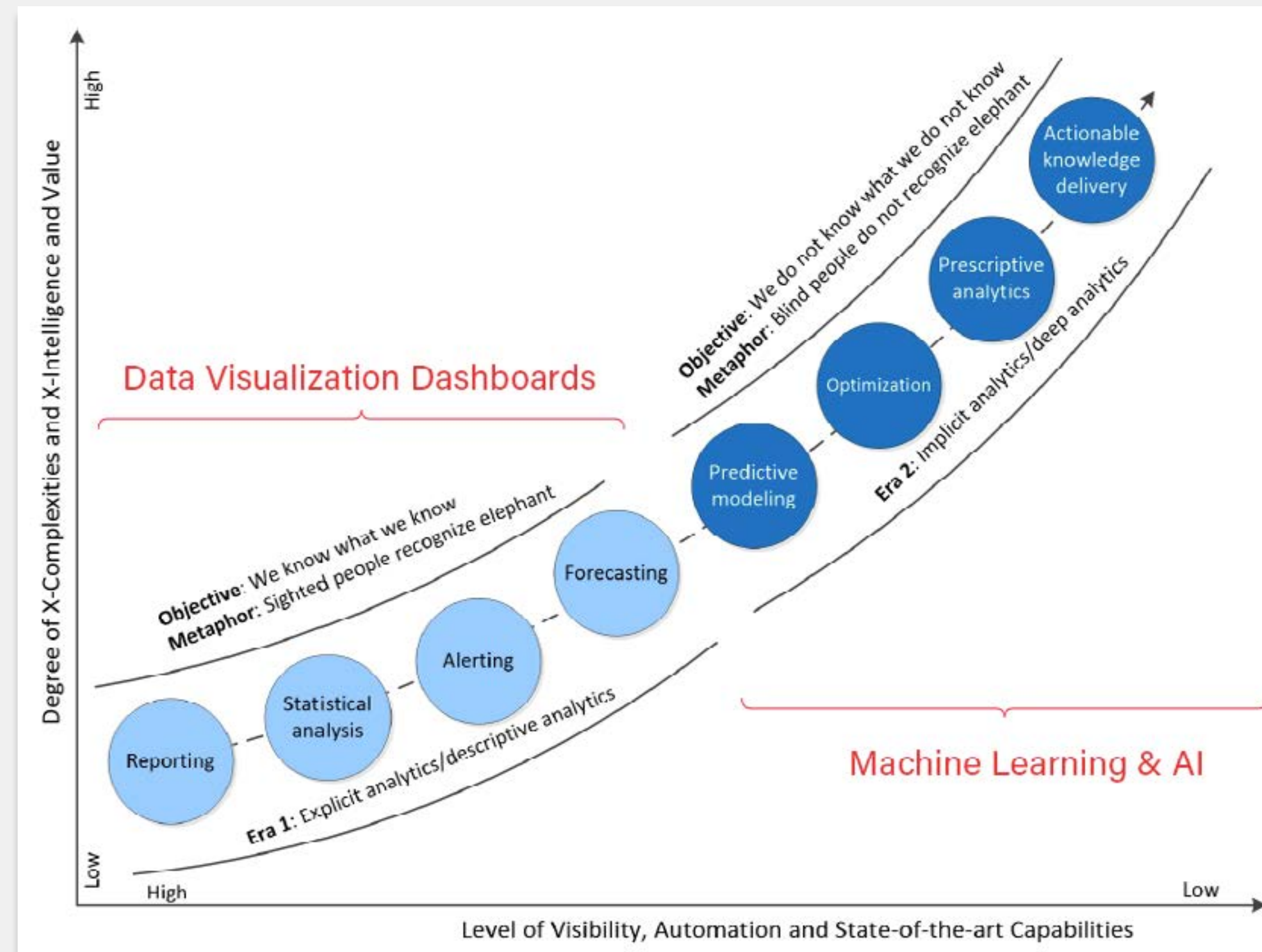
IT (E.g., Data Security)

# Data Science Conceptual Map



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

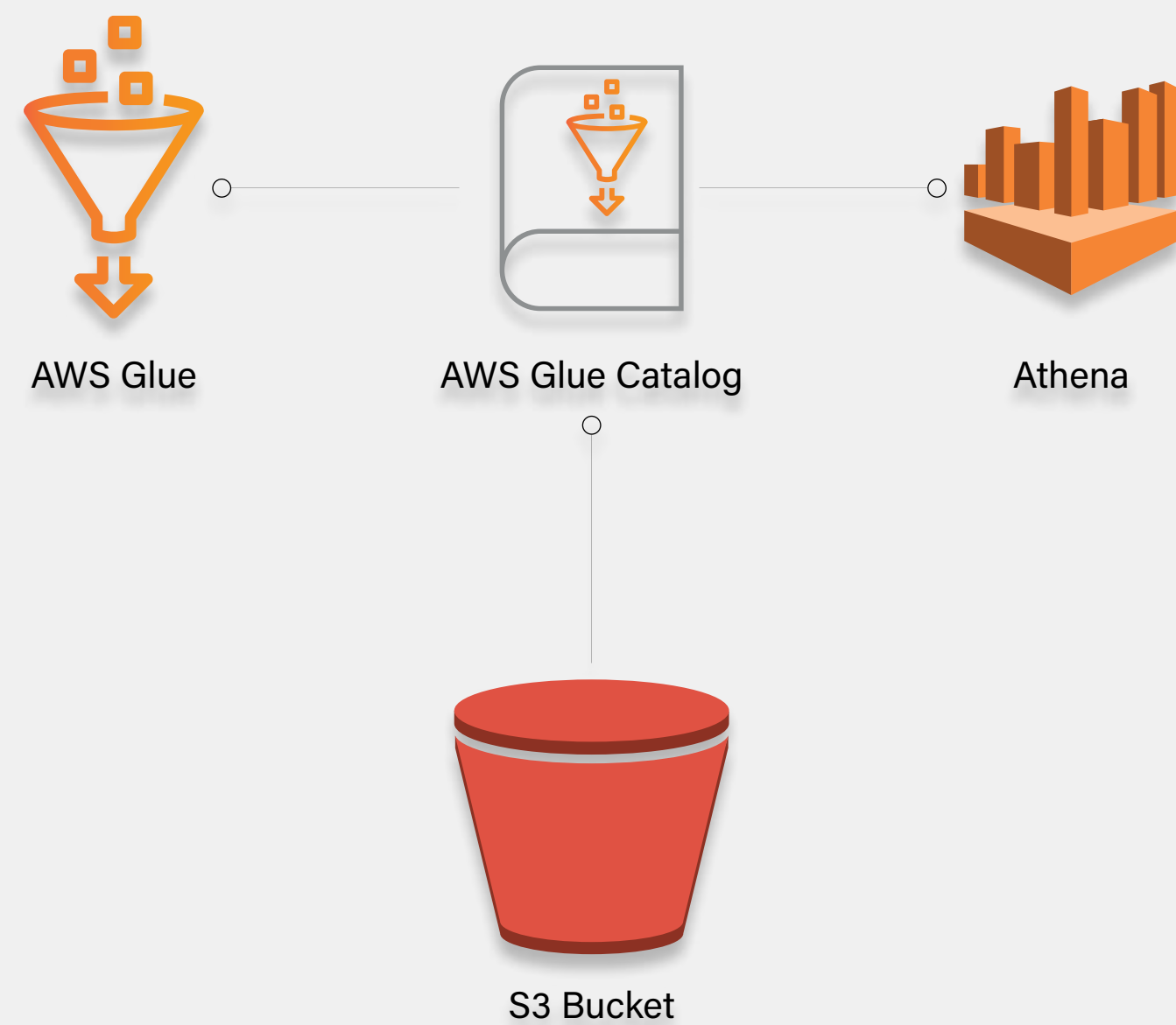
# Data Science Road Map (Value Creation)



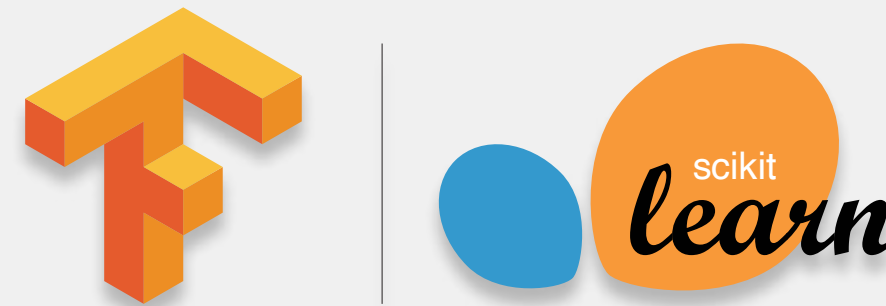
(i) Source: Longbing Cao. 2017. Data Science: A Comprehensive Overview. ACM Comput. Surv. 50, 3, Article 43 (June 2017), 42 pages. DOI: <https://doi.org/10.1145/3076253>

# Typical Data Science Platform Components

## 1. Data Lake



## 2. Machine Learning and AI Platform



**AI Services:**

Amazon Rekognition • Amazon Polly • Amazon Lex

**AI Platforms:**

Amazon Machine Learning • Amazon EMR • Spark & Spark ML

**AI Engines:**

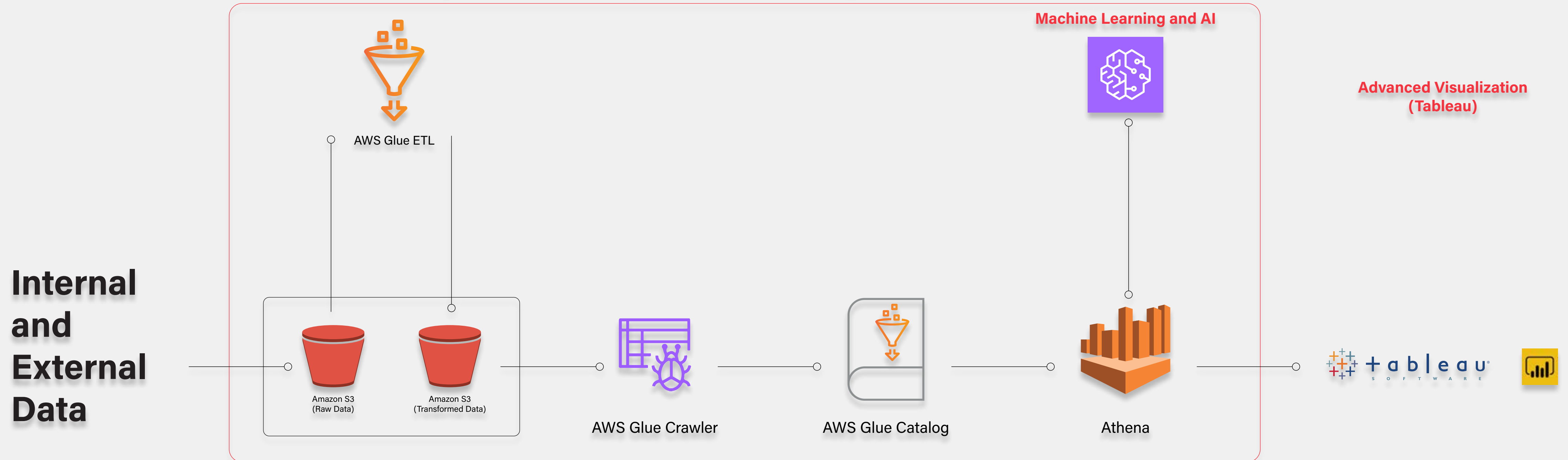
Apache MXNet • TensorFlow • Caffe • Torch • Theano • CNTK • Keras

## 3. Visualisation Framework

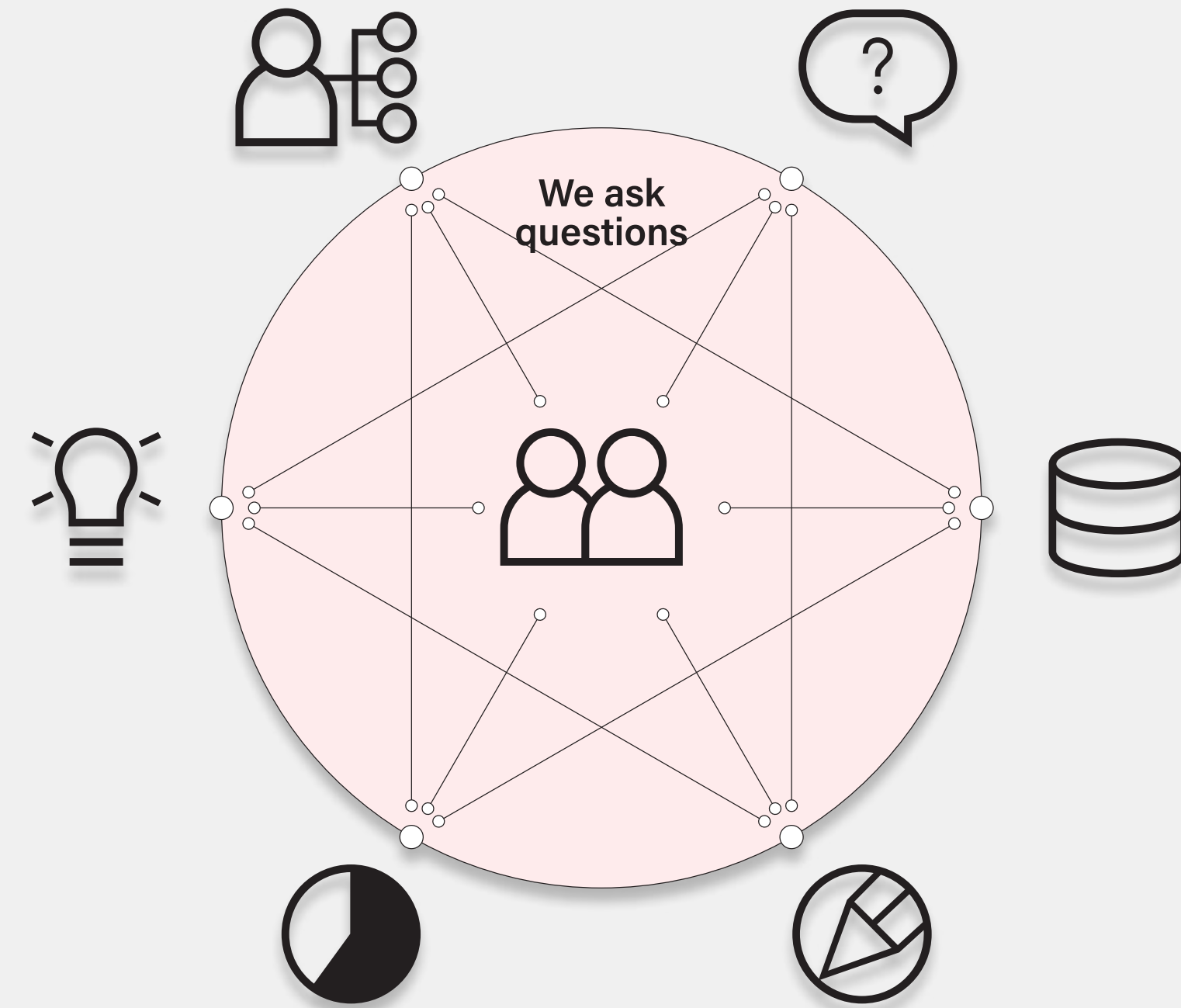
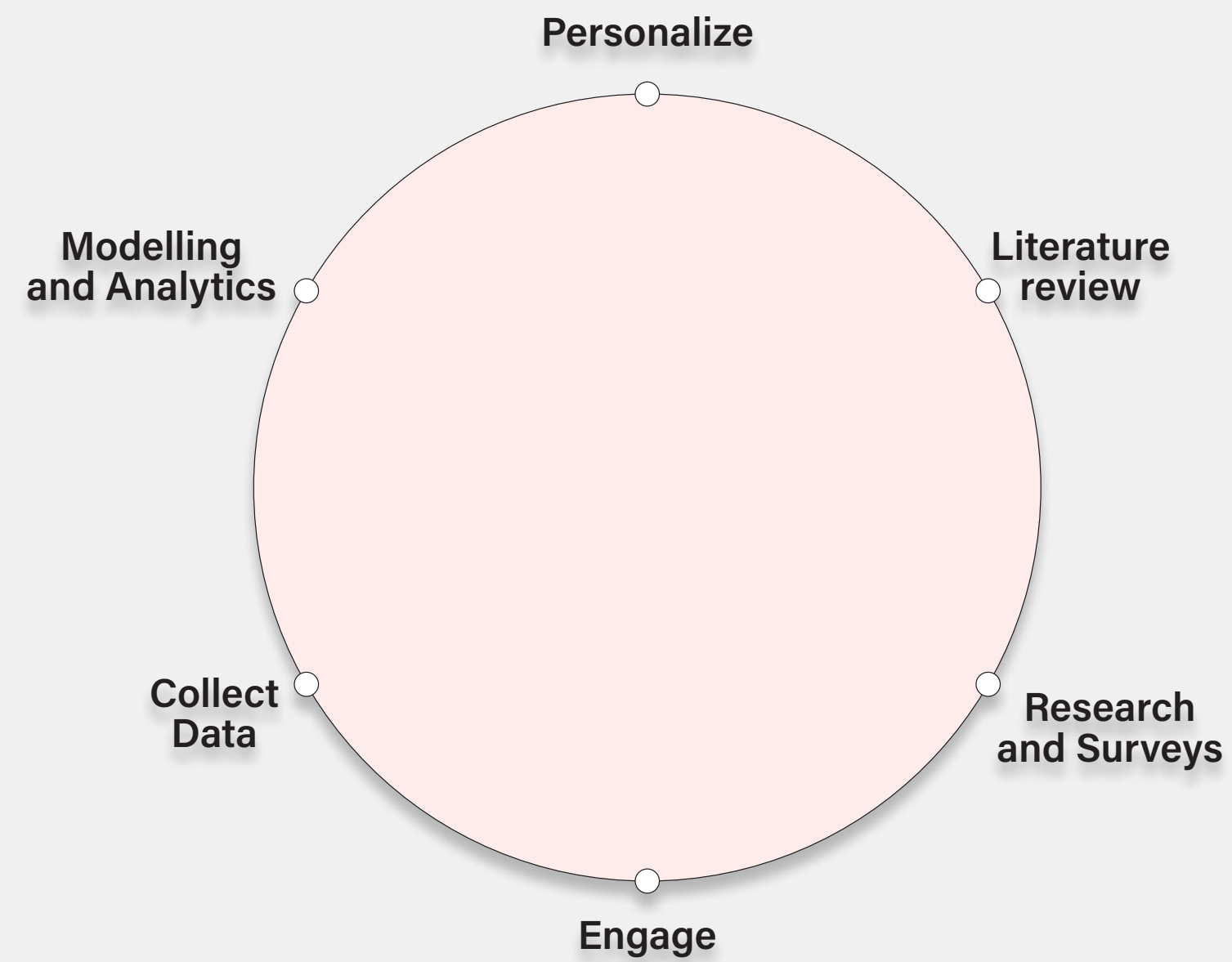
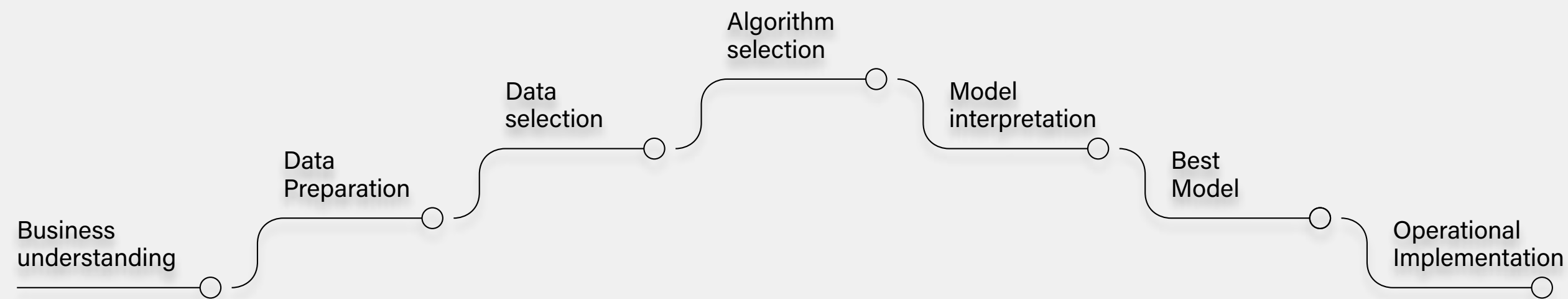


# Example for a Typical Data Science Platform (for illustration purposes only)

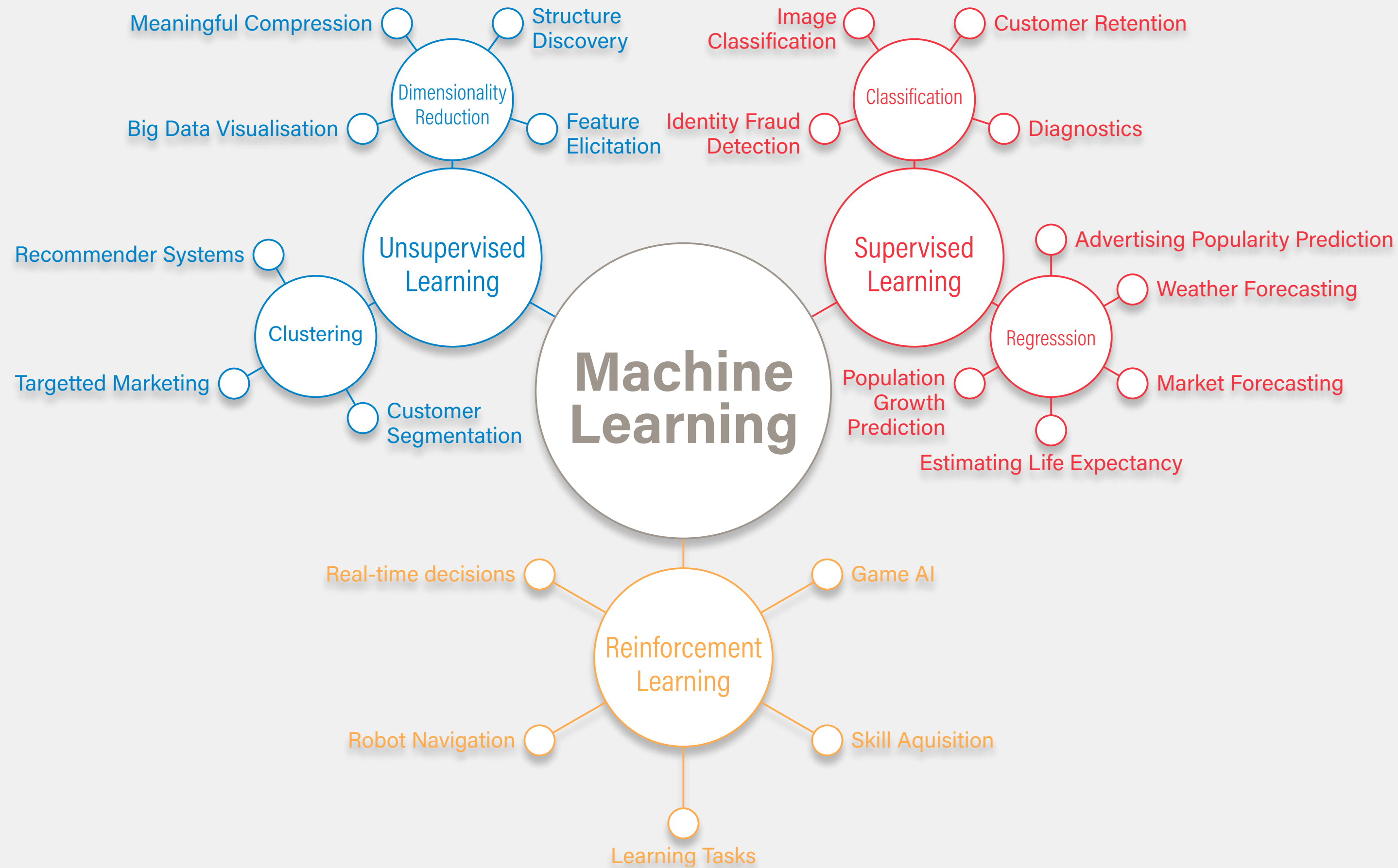
Data Lake + Machine Learning + AI + Visualization (Simple, Low Cost and Custom-designed)



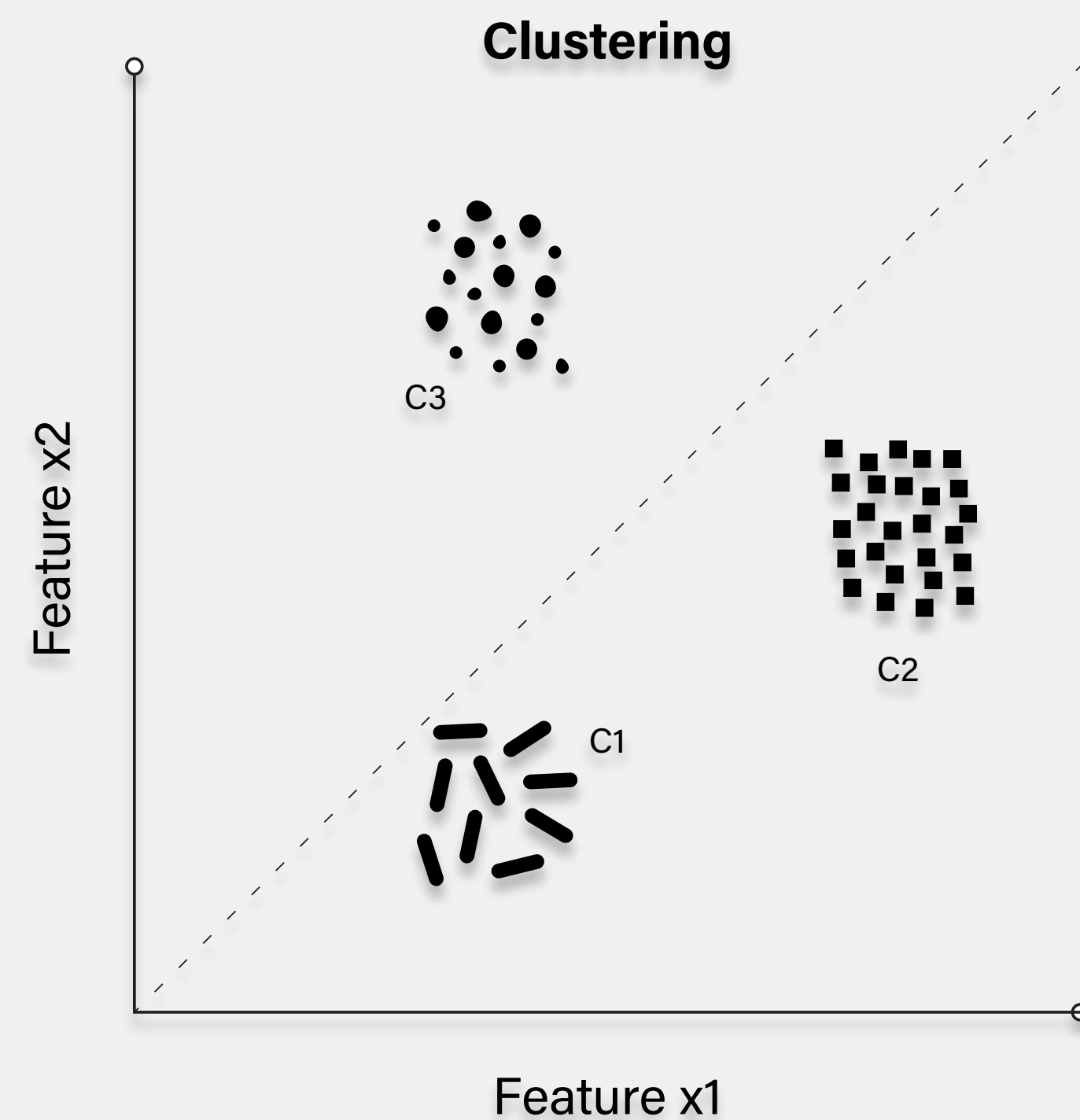
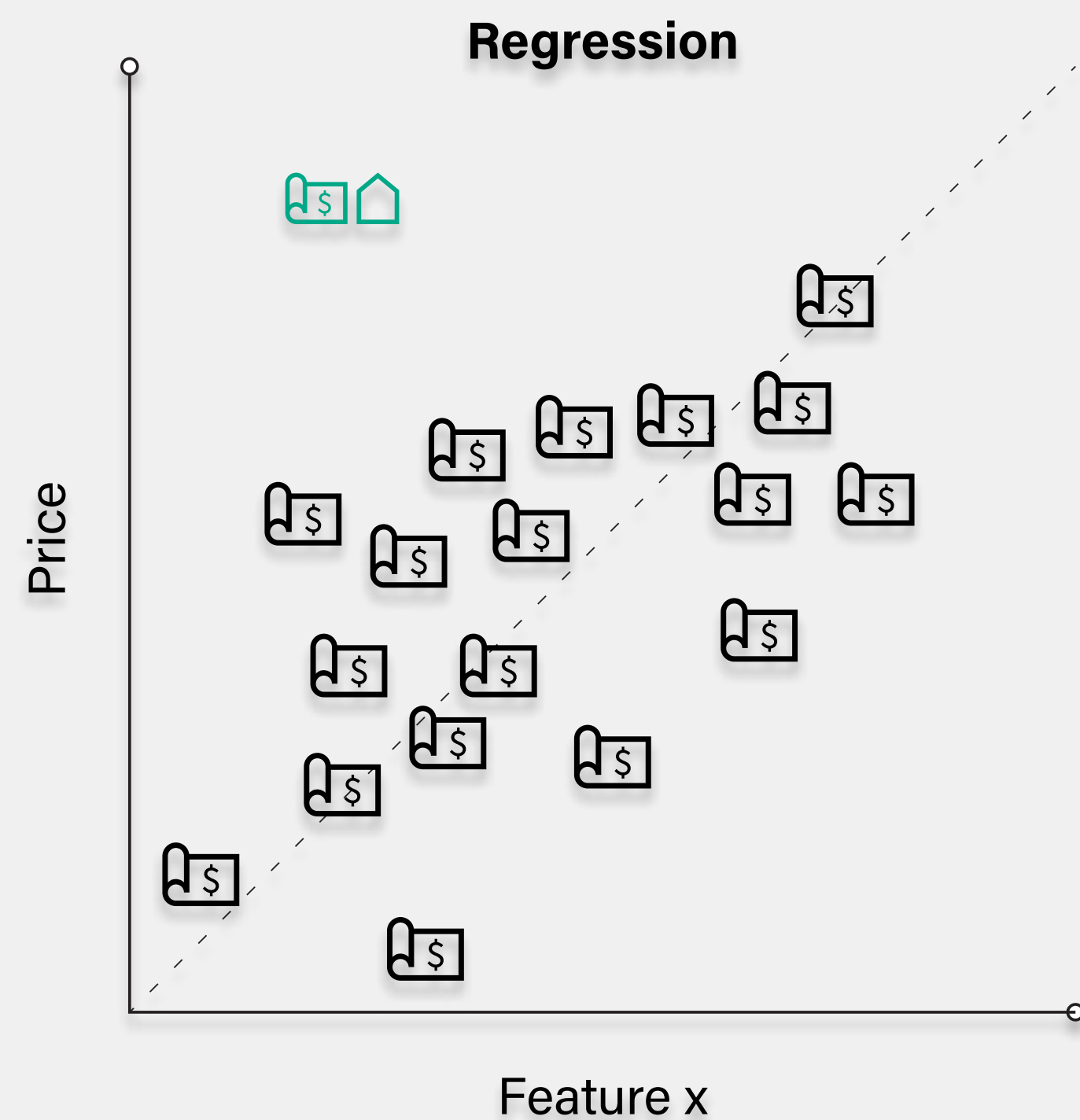
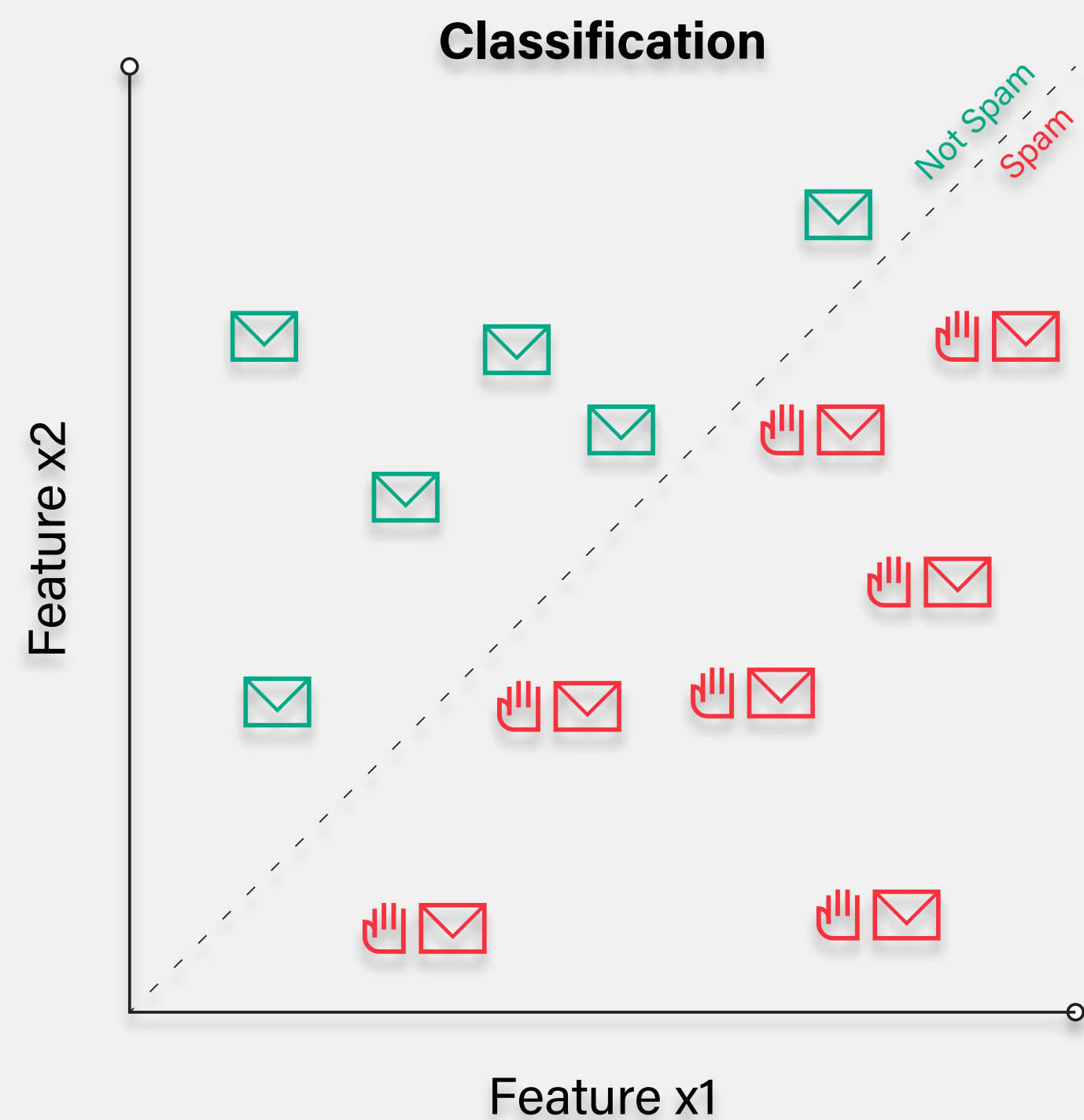
# Data Science Process (Iterative & Interactive Process)



# Machine Learning Framework



# Machine Learning: High-Level Overview





# 10 - 11 How to Implement a Data Science Model

---

**This exercise will allow teams to have hands-on experience in machine learning. By allowing teams to get a better understanding on how a data science model is implemented, they can better assess what can be integrated in their work.**

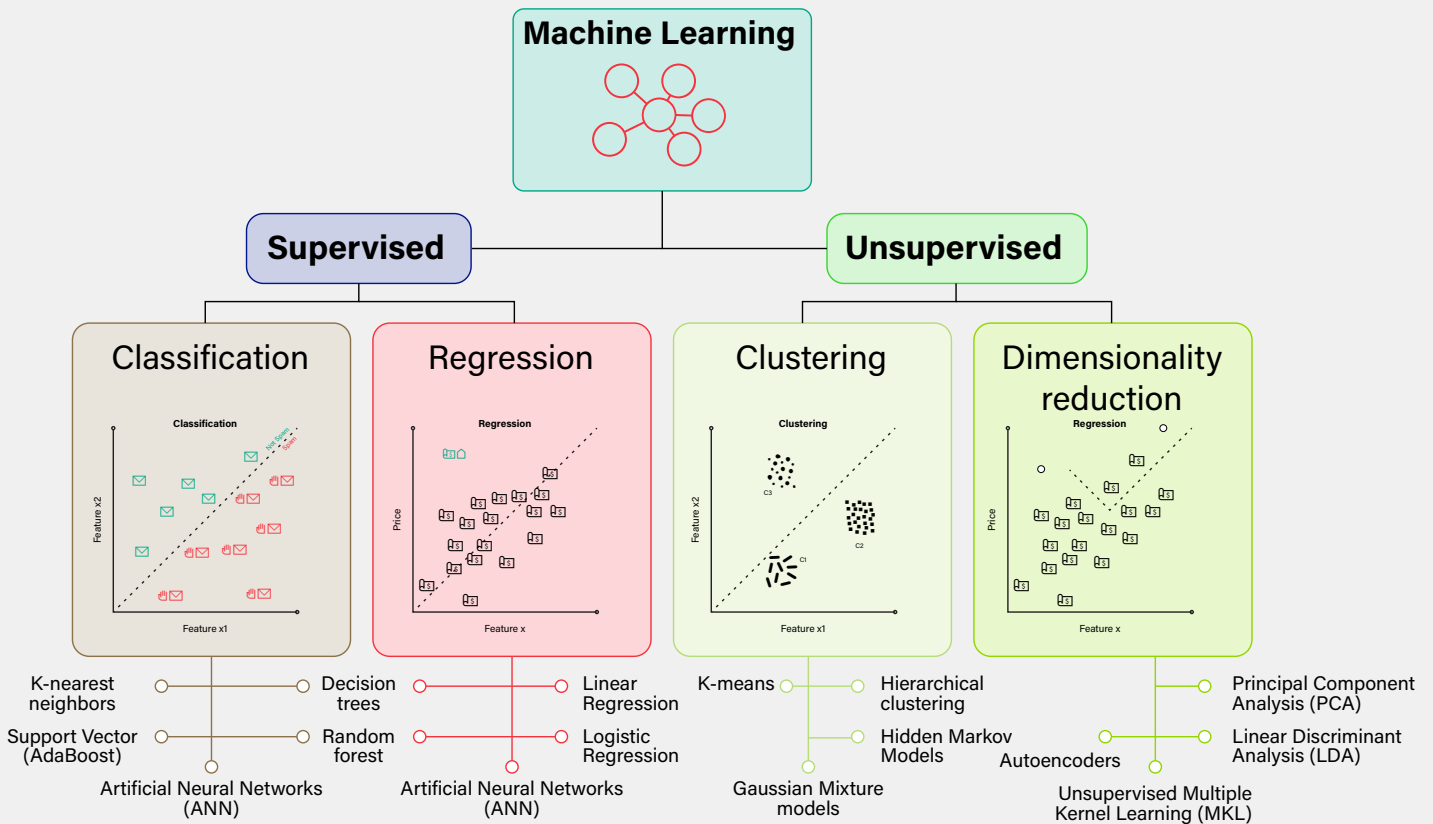
- ▶ **People:** 2 to 8 people
- ▶ **Time:** 45 Minutes
- ▶ **Difficulty:** Medium to Hard
- ▶ **Virtual Materials:** virtual meeting platform, shared document/writing space
- ▶ **In Person materials:** Flipcharts/noteboards, sticky notes, markers

## Exercise

### Part 1: Explore and Review

Machine Learning Hands-on Exercise (How to implement a data science model)

- 1 We have prepared a Jupyter Notebook to illustrate a simple machine learning classification model. [How to setup Google Colab](#). Note what is a classification model from the following diagram.



- 2 We are going to develop a simple churn model.
- 3 Because this is a binary classification model, we need to know the donor labels prior to modelling. In this case, we need to know who were active and who were not active donors. We trained this model based on past data.
- 4 Note the following
  - Data preparation stage
  - Sampling stage
  - Exploratory data analysis stage
  - Split of data to training and validation
  - Training the model (Note the LightGBM Classification Model)
  - Validation stage (Note which measure was used to validate the accuracy of the model)
  - Analysis of the model
- 5 You may further read about this kind of Machine Learning modelling:
  - <https://youtu.be/8b1JEDvenQU>
  - <https://www.youtube.com/watch?v=GM3CDQfQ4sw>
  - <https://www.youtube.com/watch?v=4jRBRDbJemM>

---

## Part 2: Feedback

Ask people to reflect on the exercise and cite 1 or 2 observations or insights.

### Extra credit

For further reading on the subject you can refer to the following links

- 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/>

### Credit

For developing the exercise Mahendra Smarawickrama and Paola Yela.

# 10 - 12 Red Cross Churn analysis Jupyter Notebook

---

 [Download resources](#)

---

 [Download resources](#) (Data Model CSV)

---

 [Download resources](#) (Data Model SQL)