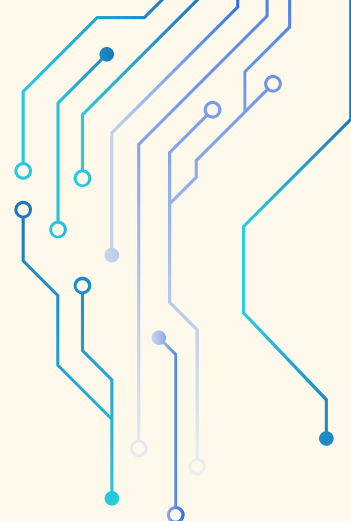# Responsible use of AI for public policy:
Project formulation manual

Gabriela Denis
María Paz Hermosilla
Claudio Aracena
Roberto Sánchez Ávalos
Natalia González Alarcón
Cristina Pombo

IDB

Responsible use of AI for public policy:
Project formulation manual

Gabriela Denis
María Paz Hermosilla
Claudio Aracena
Roberto Sánchez Ávalos
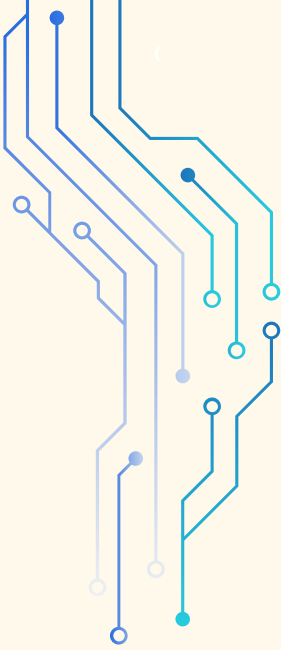Natalia González Alarcón
Cristina Pombo

August 2021

https://www.iadb.org/

Acknowledgment

CONTENTS

ABOUT THE MANUAL

fAIr LAC Initiative

The Inter-American Development Bank (IDB), in collaboration with partners and strategic allies, leads the fAIr LAC[1] initiative through which it seeks to promote the responsible adoption of artificial intelligence (AI) and decision support systems. This is to improve the provision of social services and create development opportunities to reduce social inequality. This manual is part of a series of documents and tools aimed at guiding policymakers[2] and their technical teams in mitigating the challenges inherent in AI-based decision support systems and in promoting their responsible adoption (Cabrol et al., 2020).

Why this manual?

This manual is intended to help those responsible for formulating projects with AI-based support systems to carry out their planning and design, as well as to subsequently lead their execution and monitoring. AI is a very powerful tool that can help solve complex problems, as long as it is contextualized within the public policy problem that it seeks to solve, and the ethical and legal problems involved in the application of automatic decision-making tools are addressed.

Who is this manual for?

This manual is aimed at decision-makers (managers, directors or professionals not necessarily experienced in data science) of public institutions who lead AI projects from the design phase to their implementation. The document is divided into two main parts: **planning and design**, whose tasks will be in charge of the project manager (responsible for making decisions), and **execution**, where the latter plays a critical role in the tasks that will be carried out jointly with a multidisciplinary team made up of the technical team of AI model developers, sector experts and legal specialists, among others.

This manual complements the Responsible AI Technical Manual aimed primarily at technical teams, available at https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial.

---

1   For more information, see https://fairlac.iadb.org/

2   For purely stylistic reasons, the inclusive generic masculine is used in this document, regardless of grammatical gender. The positions and functions will correspond indistinctly to people of either sex.

INTRODUCTION

Today, artificial intelligence (AI)-based decision support systems can process massive amounts of data to generate recommendations, predictions or classifications that can be used to improve different processes. We are increasingly familiar with their applications in our daily lives: from suggestions about possible friendships on our social networks or about series or movies we would like to watch on streaming platforms, to advertisements specifically designed to respond to our interests and tastes, to the use of facial recognition techniques to unlock our electronic devices.

However, AI can also be used to solve public policy problems, and institutions are increasingly implementing solutions based on it that seek to generate a positive impact on the well-being of society.

AI tools, and the capacity of modern computers to process information, have allowed them to be applied in the different tasks of institutions, solving both internal and external problems. Mentioned below are some of the uses of AI in public entities dealing with various fields of activity in society:

- Assignment of teachers or students to schools.[3]
- Timely prevention or treatment of diseases through early detection.[4]
- Matching vacancies with job candidates using more complete information.[5]
- Automatic response to requests from users and beneficiaries of an institution.
- Targeting the delivery of subsidies and benefits to the target population.
- Supervision of the use of permits for the exploitation of water sources through image analysis.
- Supervision of pollution produced by companies based on data from monitoring stations.
- Targeting of police patrols.
- Prediction of travel and waiting times in public transport.

A detailed description of some AI application initiatives for social good in the region can be found in the document "La inteligencia artificial al servicio del bien social en América Latina y el Caribe: panorámica regional e instantáneas de doce países" by the fAIr LAC[6] initiative (Gómez, del Pozo, Martínez, and Martín del Campo, 2020).

---

3   For more information about the application of AI in education, see here.
4   For more information about the application of AI in health, see here.
5   For more information on the use of AI in job matching systems, see here.
6   For more information on the initiative, see https://fairlac.iadb.org/

Decision-making and/or support systems and machine learning

The OECD describes decision support systems as "computer systems that can, for a given set of human-defined goals, make predictions and recommendations or make influencing decisions in real or virtual environments". These systems are designed to operate with different levels of autonomy (OECD, 2019).

This manual seeks to address the most common aspects regarding the use of support and decision-making systems from the perspective of the people in charge, including determining the viability of the project, detecting biases and evaluating the possibility of producing undesirable results for society or a particular institution.

Although machine learning (ML) methods are not the only type of algorithm that can be used by AI systems, they are the ones that have seen the greatest growth in recent years. They are a set of techniques that enable a system to learn behaviors automatically through patterns and inferences and not through explicit or symbolic instructions introduced by human beings (OECD, 2019). Two archetypes of the use of machine learning in decision-making processes are considered (González, Ortiz and Sánchez Ávalos, 2020): decision support systems and decision-making systems.

Decision support systems: As they are related to the concept of assisted or augmented intelligence, they include systems where the information generated by AA models is used as input for decision-making by human beings.

Decision-making systems: As they are related to the concept of automated and autonomous intelligence, the final decisions and the actions that arise from them are taken without direct human intervention. This means that the system goes on to perform tasks previously executed by people. In many contexts, the acronym ADM (Automated Decision-Making) is used to refer to these systems.

Components of an AI system for public policies

AI, or the creation of a decision-making and/or support system based on it, does not replace public policies since AI by itself does not solve any social problem. It is a tool that is used during the life cycle of public policies to provide information in the form of prediction, classification and/or segmentation, among other possibilities, in the context of formulating an intervention or action of a social nature, as seen in Figure 1:

Figure 1. The life cycle of public policy with AI

PUBLIC POLICY LIFECYCLE

AI LIFE CYCLE

Source: González, Ortiz and Sánchez Ávalos, 2020.

The public policy life cycle is a simplified tool that seeks to represent how policies should be developed, and that serves to plan and analyze the different phases of the process. The cycle begins with the identification and definition of a problem or issue to be resolved and then goes on to formulate the different courses of action. Consequently, the government must evaluate the proposed alternatives, including maintaining the status quo, that is, refraining from taking any action.

Throughout this process, technology acts as an instrument that can be used to develop the proposed public policy alternatives based on the context and feasibility of the solution. Once the government chooses the alternative and the instrument to develop it—for example, through an AI-based decision-making or support system—, the implementation phase will begin. Finally, in the final stage, the aim is to evaluate the effectiveness of the policy in terms of its objectives, results and expected impacts (Giorgi, 2017).

The AI life cycle is activated when it is identified as the appropriate instrument to incorporate into public policies that will respond to the selected problem. The stages in the AI life cycle are the following:

1. **Planning and design:** It includes the key points that decision-makers need be clear about before starting a project.

2. **Data collection and processing:** It includes data cleaning and processing, as well as the identification of deficiencies and biases that may jeopardize the development of the model.

3. **Model building and validation:** It includes the key concepts to follow in order to have robust and validated AI systems.

4. **Deployment and monitoring:** It includes the evaluation of the tool once the implementation has begun.

5. **Accountability:** It is associated with the need to provide information and transparency to foster public understanding of AI.

In the interrelation of these two cycles, an important set of challenges is generated to achieve a robust and responsible AI, which must be evaluated and considered during the development and use of such systems. There are also cross-cutting challenges, including transparency and accountability, as well as personal data governance, security and protection. Finally, there are other challenges related to the design of public policies and the definition of the intervention, which are also related to the application of criteria of necessity and proportionality in the use of AI, which is presented throughout the life cycle of the AI.

The role of public policy decision-makers (project managers) is not to develop the AI-based tool but to formulate the project, communicate with the technical team, monitor risk mitigation, determine the viability of the tool and be accountable for its execution.

Figure 2. AI life cycle in light of the structure of this manual
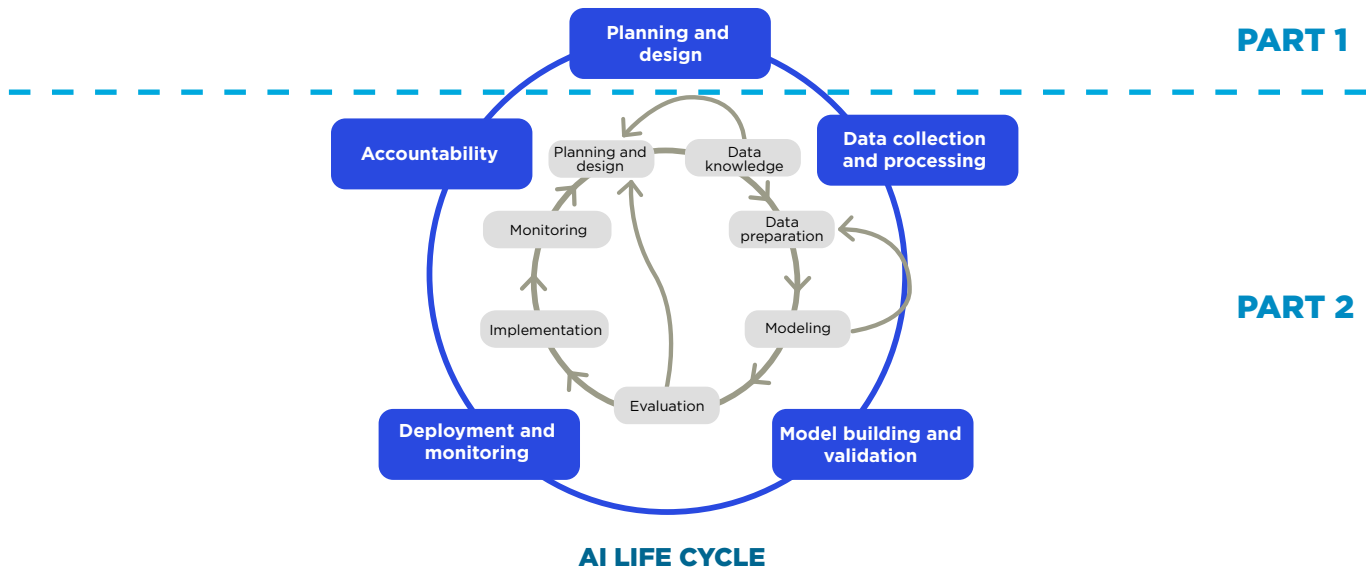
Source: Modified from González, Ortiz and Sánchez Ávalos, 2020.

To the extent that they affect the management of the institution and its users, AI projects require a person in charge of making decisions and a multidisciplinary team of professionals responsible for execution. It should be noted that, in the different stages of the life cycle of each project, knowledge of various areas that may or may not be the domain of the team in charge of the implementation will be required. In such circumstances, it will be necessary to consult experts either within the institution or externally.

The two main functions in the formulation and execution of the project can be described as follows:

- The decision-maker or the project manager is in charge of formulating public policies and will lead the project design and subsequent implementation.
- The technical team will be in charge of analyzing data sources, developing AI models and monitoring their use.

In order to join the development of a decision-making and/or support system, tools are proposed for both the project manager and the technical team. These forms (Annexes 1 to 5) must be filled out simultaneously and are part of a feedback process throughout the development cycle.

The tools for public policy decision-makers are the following:

- **Project design and feasibility sheet:** This tool seeks to identify the main key aspects of an AI project to assess its viability, determine if AI is the correct solution for the problem and collect the necessary information for the design, which must be shared with the technical team and with the entire multidisciplinary team responsible for the project (Annex 1).

- **Data Maturity Matrix:** This allows an initial approach to the quality and relevance of the data to be used (Annex 2).

- **Project manager checklist:** This tool seeks to consolidate the main concerns per risk dimension of the AI life cycle from the decision-makers perspective (Annex 3).

The tools for the technical team are the following:

- **Data profile:** It includes the main findings of the databases that will be used with the tool. It is based on data exploratory analyses conducted by the technical team[7] (Annex 4).

- **Model card:** It is a final description of an AI model that meets the requirements of public policies and that can be carried out according to the available data[8] (Annex 5).

- **Technical team checklist:** This tool seeks to consolidate the main concerns by risk dimension of the AI life cycle from the technical team perspective[9].

As indicated at the beginning, this manual focuses mainly on the aforementioned tools corresponding to the project manager, who is in charge of making public policy decisions. However, throughout the document, the tools of the technical team will also be mentioned, to the extent that the project manager must supervise their implementation. For this reason, it is suggested that the technical team rely on the manual "IA Responsable: ciclo de vida de la inteligencia artificial" [Responsible AI: Artificial Intelligence life cycle] by González, Ortiz and Sánchez Ávalos (2020), wherein the technical tools and the main risks that arise during the development of a model and the measures to mitigate them are described in greater detail.

Throughout this manual, the DART[10] case will be used—an AI-based diabetic retinopathy early detection solution—in order to exemplify what it has to do with the planning and design of the project and its execution.

---

7 Part of Technical Manual-Artificial Intelligence Life Cycle. https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial

8 Part of Technical Manual-Artificial Intelligence Life Cycle. https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial
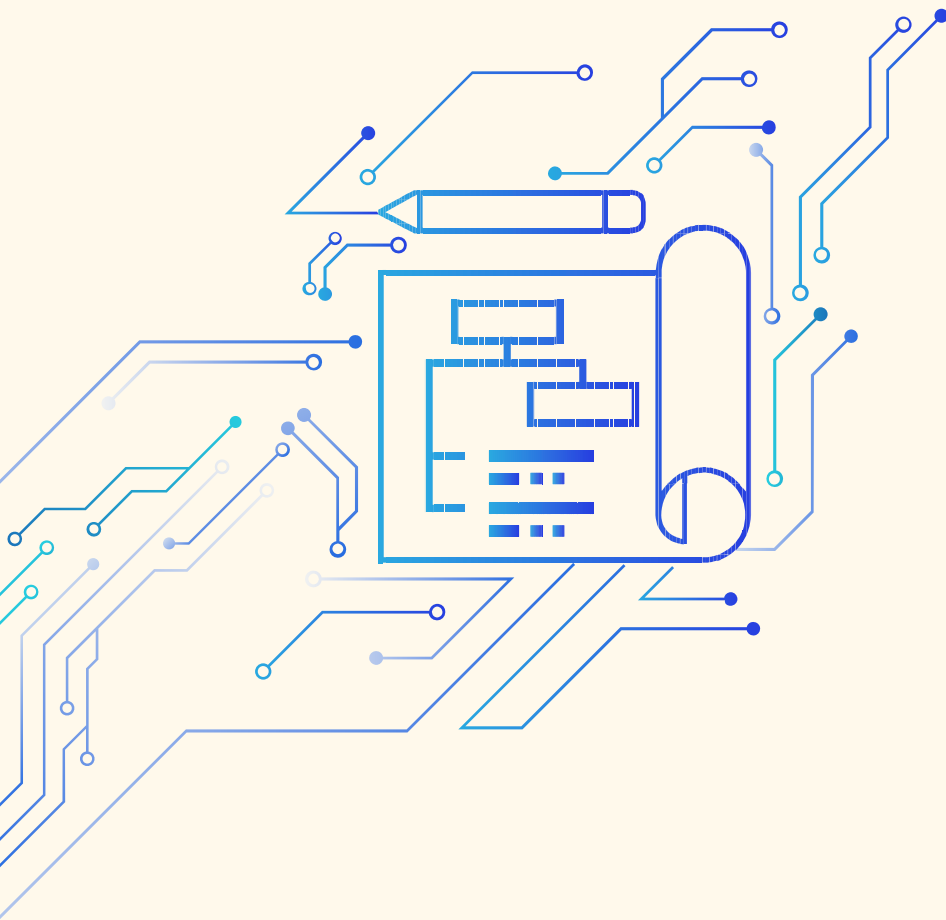
9 Part of Technical Manual-Artificial Intelligence Life Cycle. https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial

10 DART (Chile): https://www.teledx.org/dart/?lang=es

PART 1.
PLANNING
AND DESIGN

PART 1. PLANNING AND DESIGN

The first stage in the AI life cycle is the planning and design of the project, which will be in charge of its manager, that is, the person responsible for making decisions and directing the effort, which will be executed collaboratively with a multidisciplinary team from the institution. Leadership falls on that particular person since they have the expert knowledge on the subject to be addressed and the vision of the public policy that is sought to be implemented. It is also responsible for directing the execution of the project by the technical team and applying the tool to the target population.

A good planning and design of the project will ensure its viability, sustainability and public value, and will also mitigate the risks that the application of the AI tool entails.

In the planning and design process, key steps and essential questions that the decision-maker must answer before executing an AI project can be identified.

1. **Problem definition:** The first step in any project is to clearly define the public policy problem to be addressed by using the implementation of an AI-based decision-making and/or support tool.

2. **Pre-feasibility analysis:** After defining the problem and determining if AI is the right tool to support the solution, and before proceeding with the project, there are other key questions that need to be answered (see 1.2 below). The goal of this step is to ensure the viability of the project so as not to waste the scarce resources of the institution.

3. **Definition of objectives:** Once the project is declared feasible, it is time to set the objectives and their metrics or indicators, which will serve to measure the achievements. Such metrics should reflect the expected impact of the tool application on the target population. The achievement of these objectives should help to solve the identified problem.

4. **Action description:** Actions are the activities carried out by the public institution that will materialize the public policy response to solve the problem. They can be part of public policy programs aimed at addressing specific problems, or institution regular processes such as hiring, payments or customer service. While these actions typically exist independently of the AI system, AI will help to transform them in order to achieve the project objectives.

5. **Data mapping:** It should be investigated if the necessary and sufficient data exist to carry out the project, if the institution has access to the databases or if agreements will be needed to obtain them. An AI project can be based on both internal and external, public or private data. It should be noted that, at the design and planning stage, it is not necessary to carry out a detailed analysis of the data; this will be done by the technical team once it is decided to go ahead with the project.

6. **Analysis definition and relevant tools:** At this stage, the project manager must preliminarily identify the type of analysis required to solve the problem. The type of analysis, or the tool to be implemented, will depend on the nature of the analysis and will help to improve the necessary attention or response processes. At this stage, it is about achieving an initial approximation that must later be agreed upon with the technical team.

7. **Ethical, legal and governance considerations:** Even before starting project execution, the project manager should be clear about the ethical and legal challenges that may arise during im-

plementation. This will allow to anticipate possible situations that could impose a risk and take appropriate mitigation measures.

8. **Responsible team formation:** As a project manager, the person responsible for making decisions must form the team that will be in charge of carrying out the project. AI projects not only involve decision-makers and the technical team but also involve a variety of institution areas (for example, the legal team) and even external institutions (for example, those who have useful databases for the project).

The process of conceptualizing and designing a project must be iterative. Although the idea is to start with a solid definition of the problem scope, this can change if, for example, the institution does not have the necessary capacity to act on it, or must rethink if the required data is not available or sufficient.

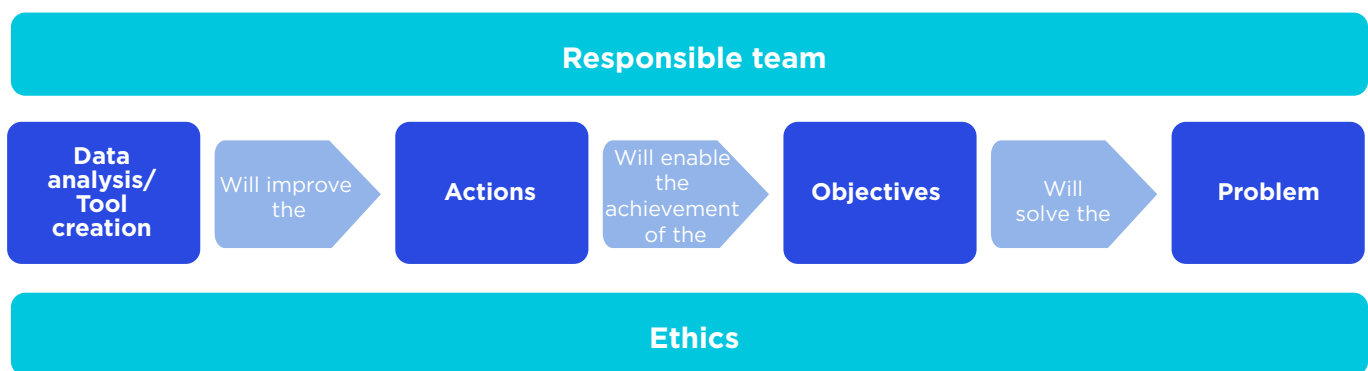**Figure 3.** Planning and design process
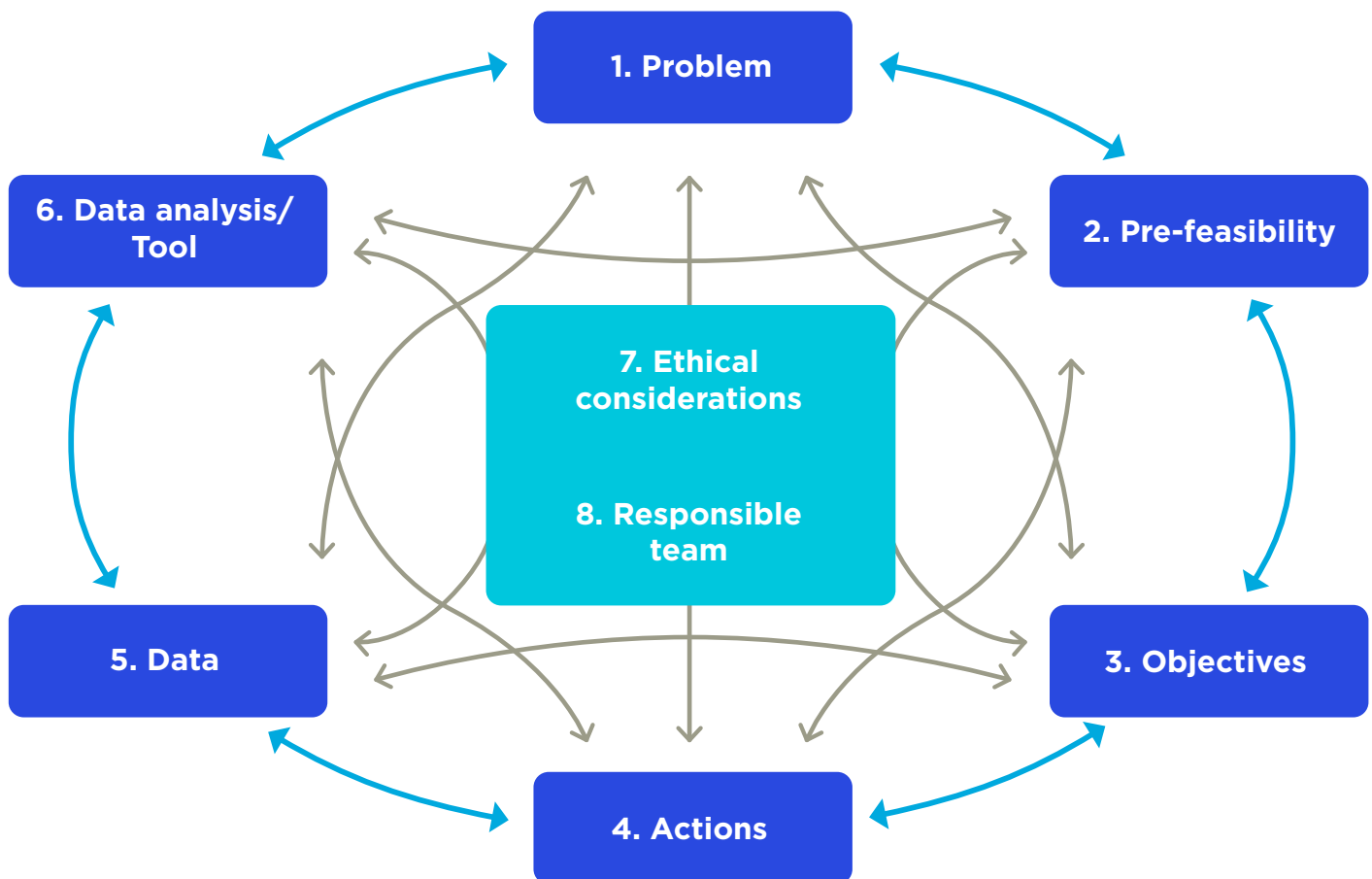
Source: Our own elaboration.

Once the project has been formulated, it should be possible to visualize a linear relationship among the steps as follows (Figure 4):

1. The analysis or tool will depend on the available data.

2. The result of the analysis or tool should improve the actions of the institution.

3. The improvement of the actions will allow achieving the defined objective and the desired state of the situation.

4. Reaching the objective will help solve the problem.

Ethical considerations accompany the entire process.

Figure 4. Relationship of the planning and design stages

Responsible team						
Data analysis/ Tool creation	Will improve the	Actions	Will enable the achievement of the	Objectives	Will solve the	Problem
Ethics						

Source: Our own elaboration.

The essential steps in the planning and design stage are detailed below. Annex 1 of this manual contains the design and feasibility sheet to be completed by the project manager on which the steps described here are based. Each of these phases will be illustrated with the example of the Chilean DART project on the use of AI in the prevention of blindness by completing the appropriate subsections of the Project Design and Feasibility Sheet.

Theory **1.1 Problem definition**

The first step in the planning and design of a project involving AI is to define the public policy problem that an institution seeks to solve; as such, it is also part of the life cycle of a public policy. The person in charge of managing the project and making the decisions must have expert knowledge on the subject to be addressed and will be the one who prioritizes its resolution and communicates the problem to the technical team.

Public institutions must deal everyday with multiple problems of a different nature, such as response times, resource allocation, task distribution, improve the allocation of social benefits, among others. At this stage, it is important to limit the problem to be solved with the implementation of the AI tool as much as possible. The prioritized problem can affect users or beneficiaries external to the organization as well as those who work within it.

The motivations that are often given to start an AI-supported project are listed below:

> **"I have a lot of data and I want to do something with it"**
>
> **"I want to make a predictive model"**
>
> **"We have this software and we must use it"**
>
> **"I would like to use X technology in the institution"**

However, these reasons do not justify undertaking an AI-based project. While it is true that AI tools can be highly beneficial in terms of saving time and increasing efficiency within the institution, they are also costly in terms of the resources and time required to develop them. In this sense, it is key to clearly identify a priority problem that the entity is bound to solve and for which an AI-based tool represents added value.

Once the problem has been defined, it is extremely important to be clear about the institution's current state of response, identify its limitations and areas of opportunity, and determine how the AI-based support or decision-making system could improve the status quo. A recommended good practice is to study comparative experiences of implementation of similar tools in other institutions or other countries. Thus, information will be available on the challenges that had to be faced, which will also contribute to determining the project feasibility.

Within the problem definition, it is important for the team to accurately quantify the number of people affected and the budget that the solution will require. This will help measure the magnitude and seriousness of the problem, and therefore the need to solve it.

Finally, there must be political will to act and solve the problem. Entities—especially public bodies—have to deal with a multiplicity of problems of different kind and therefore they must be the ones who prioritize the problem to be addressed so that the project has the required institutional support.

The topic to be addressed must be clearly and concisely defined, considering all the factors mentioned above. The project manager must bear in mind that it is not just about communicating the problem to the technical team, but also to the target population. For this reason, it is recommended to use a simple description that conveys its importance and keep the focus of the definition on the problem itself and how it affects people and organizations, as seen in the examples below.

Examples of problem definitions

Project name	Problem definition
Development of Predictive Risk Models to Support Allegheny County Child Maltreatment Hotline Decisions (Pennsylvania)[11]	In order to combat the problem of child maltreatment, Allegheny County has two emergency lines where these abuses can be reported: a county emergency line, and a specialized emergency line called "Childline". Complaints can be classified as CPS (Child Protective Services) and GPS (General Protective Services). A household enters the system from the moment a thorough investigation is ordered. Of all the complaints classified as GPS, 48% were entered into the system from April 2010 to May 2016. The problem is that 53% of the complaints that did not generate an intervention were repeated within two years. This indicator shows that the current decision-making system, as is, leaves out children who need county support. Given the high flow of calls, and the asymmetry of information between the people who enter the complaints and the operators of the emergency lines, the latter must make decisions in a short time and are unable to consider all the information available.
Suicide Risk Detection in Chat Applications	In Chile, suicide is the first cause of death among young people between 15 and 24 years old. According to data from 2015, every 2.8 days a minor commits suicide. The foundation Todo Mejora12 is an NGO whose objective is to promote the well-being of children and adolescents who suffer bullying and suicidal behavior as a result of discrimination based on their sexual orientation, gender identity and expression. Todo Mejora offers a non-face-to-face help service called "Hora Segura" ["Safe Hour"], which began through Facebook Messenger and whose objectives are to prevent suicide, guide and, in specific cases, refer cases to more advanced instances of resolution. The current problem is that the number of users who need help at times exceeds the number of volunteers available for this task and it is not possible to prioritize contact with those children and adolescents who are at high risk of suicide. Specifically, Todo Mejora has a team of between 60 and 80 professionals, who must respond to between 8,000 and 9,000 telephone assistance requests per year. Annually, around 60% of queries are received from adolescents from the LGBTIQ+ community and 40% from cisgender heterosexual adolescents. 68% of the people who contacted "Safe Hour" had presented suicidal behavior in the last two months.

11 Allegheny Family Screening Tool (USA): https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx

12 Foundation Todo Mejora (Chile): https://todomejora.org/

"There is the additional risk of considering AI projects from the technology point of view and not from the particular social problem point of view. Even a necessary and functional AI project can present risks if the correct public policy action is not considered" (Cabrol, et. al., 2020).

Common mistakes in problem definition

Although the problem identification phase seems to be very simple, it is perhaps the most critical of the planning process and where most mistakes are made, such as the following:

- Include excessively broad and unrestricted descriptions and concepts (such as "improve decision-making", "be more efficient", etc.), instead of focusing on a specific institution problem.
- Define the problem as the absence of a tool/model. For example, "the problem of my institution is that we do not have an automatic aid delivery targeting tool".
- Not specifying the current status and/or the gaps that need to be closed.
- Not quantifying the number of people affected and/or the budget required to solve the problem.

Don't forget to...

1. **Clearly** define the public policy problem to be solved, identifying and quantifying the groups of people who are affected, and determining their impact on the budget. The definition of the problem must be easily understood by someone outside the institution.

2. **Contact** the people in the institution who are in charge of addressing the problem in order to establish how it is currently addressed. What information can they offer on how to improve the response system?

3. **Investigate** how other agencies—either national or foreign—with a similar problem have implemented an AI-based solution. Ideally, contact them to learn about the challenges and difficulties they encountered along the way.

4. **Discuss** with the institution management staff the priority of solving the problem and achieving a commitment to the project at the highest level. If you are a senior manager yourself, document how this priority is reflected in the institution's strategic plans.

Actividad — After this section, complete the corresponding **Problem Definition** in the Project Design and Feasibility Sheet.

PROJECT DESIGN AND FEASIBILITY SHEET

1   Problem Definition

What is the problem to be solved?

Describe the population(s) affected by the problem (people, groups, entities, etc.)

How many people/organizations/locations/etc. are affected and to what extent?

Why is solving this problem a priority for your organization?

Have you heard of any similar use cases for AI that have been implemented before? Which one?

PROJECT DESIGN AND FEASIBILITY SHEET

DART EXAMPLE: Using Artificial Intelligence in Blindness Prevention

1   Problem Definition

What is the problem to be solved?
Diabetes is a chronic disease that occurs when the pancreas cannot produce enough insulin or when the body does not use all of the insulin it produces. There are two types of diabetes: type 1 diabetes and type 2 diabetes. The first one arises as a result of an attack by the immune system, it cannot be prevented and has no cure, while the second one results from lifestyles that lead to sedentary lifestyle and obesity. The pancreas continues to produce insulin, although in insufficient quantities, and in some cases it can be cured. Both types of diabetes can cause blindness, kidney failure, strokes, and even lead to leg amputation.

Diabetic retinopathy (DR) is the most common eye disease among people with diabetes in the world. Diagnosis of this disease requires a fundus examination, where a medical technologist obtains a detailed image of the eyeball that is then analyzed by a specialist.

According to estimates made, in Chile, there is an annual deficit of 39,168 hours of ophthalmologists, which indicates that the human capital necessary to analyze all fundus examinations does not exist. (Hojman, 2014).

Describe the population(s) affected by the problem (people, groups, entities, etc.)
This problem basically affects three groups:

1. People diagnosed with diabetes who cannot have an eye fundus examination once a year, according to international recommendations.

2. The Ministry of Health, that must spend additional resources on diabetic retinopathy as it cannot carry out adequate prevention.

3. Ophthalmologists, whose workload is excessive.

How many people/organizations/locations/etc. are affected and to what extent?
In 2014 there were 422 million people with diabetes in the world. According to estimates by the World Health Organization, by 2040 that number will rise to 600 million. In Chile, it is estimated that one in 10 people has diabetes and that between 15% and 20% of these patients have some degree of diabetic retinopathy. According to the 2010 National Health Survey, only 34.8% of patients were examined by an ophthalmologist.

Why is solving this problem a priority for your organization?
Teledx was created with the sole objective of technologically supporting the ophthalmological task of the health system and thus reducing the rates of vision loss and blindness in the adult population. The main focus is diabetic retinopathy, as there is an urgent need to increase diagnostic coverage.

La Estrategia Nacional de Salud, elaborada por el Ministerio de Salud para la década 2011-2020, establece distintos objetivos, siendo uno de ellos "Incrementar la proporción de personas con diabetes controlada". La estrategia específica para lograr este objetivo consiste en mejorar el control de los pacientes diabéticos a través de una cobertura más amplia de distintos exámenes clínicos, entre ellos el de fondo de ojo.

Have you heard of any similar use cases for AI that have been implemented before? Which one?
Yes, a literature review of other studies that seek to detect diabetic retinopathy was carried out (Arenas, 2012).

1.2 Pre-feasibility analysis

After defining the problem, it is necessary to ask how feasible it is for the project to be carried out, beyond its technical aspects in charge of the specialists team. This step is key before starting a project since it avoids wasting the scarce financial and human resources of the institution. Some of the questions to be answered in this step are the following:

1. **Is it within the powers of the entity to act on the problem? Will it be necessary to partner with other public institutions? Are there sufficient human and financial resources to carry out the project?**
   This point refers to both the legal and regulatory competence to respond to the problem, and the availability of human (internal) and financial resources that could be used to produce the tool. It is necessary for the institution to be able to respond to the problem within the legal and budgetary framework.

2. **Is there relevant data (enough to be able to change the way the problem has been addressed so far)? Is it possible to access this data?**
   The data must be enough to be able to create the AI tool to help solve the problem. For example, if the issue to be resolved is that the competent authority cannot predict environmental events of poor air quality, which results in an increase in medical consultations for respiratory diseases, it is necessary to at least have historical information on the presence of polluting particles in the atmosphere in recent years. If the data is aggregated by months, then it is unlikely that a predictive model can be developed.

3. **What are the project risks (ethical, social license, implementation risks, etc.)?**
   A project of a public body supported by AI will seek to generate value for the population. However, since AI tools and their execution impose some risks, an analysis should be carried out comparing the expected positive effect versus the potential risks of AI tools (more on these topics later). Some of these risks are:

   Ethical risks: Automated response systems can reproduce the biases present in the data, or they can be biased as a result of a wrong model. Since there is a probability that the tool presents some of these risks, it is up to the project manager to take the necessary measures to mitigate them.

   Social license: This refers to the approval given by the target population to the use of the tool. In other words, if the target population knows about the project, would they approve the use of the AI? To obtain the social license it is necessary, first, that the population is clear about the benefit they will obtain from the application of the tool, and second, that there is transparency about data protection, the way the tool is used and the measures taken to mitigate biases, in case they are detected.

   Implementation risks: These are the risks that may arise once execution begins by the institution: the first risk is that the model does not work well enough, and therefore there is no improvement in institutional work both internally and externally; the se-

cond risk is that it is never implemented due to failures in any of the key components (commitment of the authority, lack of resources, and/or resistance to change by the people who must use the tool); and the third risk is that it is implemented but never used.

Answering these questions will allow to identify some key aspects that will lead to the successful formulation and execution of the project, to the extent that this exercise facilitates a better understanding of its limitations and risks.

The idea is that from the very beginning the need for the project is fully justified, that there is the capacity to design and execute it effectively, and that both the risks that it may present and the measures to mitigate them are identified.

Don't forget to...

1. **Engage** the people who will execute the tool and incorporate them into the project team.

2. **Identify** the possible ethical and legal risks of the model. Talk to the legal team and incorporate them into the project team.

3. **Ensure** the financial and human resources that allow the tool to be sustained over time.

Activity After this section, complete the corresponding **Pre-feasibility Analysis** in the Project Design and Feasibility Sheet.

PROJECT DESIGN AND FEASIBILITY SHEET

2	Pre-feasibility analysis

Is it within the powers of the entity to act on the problem? Will it be necessary to partner with other public bodies? Are there sufficient human and financial resources to carry out the project?

Is there relevant data (enough to be able to change the way the problem has been addressed so far)? Is it possible to access this data?

What are the project risks (ethical, social license, implementation risks, etc.)?

PROJECT DESIGN AND FEASIBILITY SHEET

EJEMPLO DART: Using Artificial Intelligence in Blindness Prevention

2	Pre-feasibility analysis

Is it within the powers of the entity to act on the problem? Will it be necessary to partner with other public bodies? Are there sufficient human and financial resources to carry out the project?
It requires partnering with other agencies, especially with the Ministry of Health, so that the tool can be implemented in health centers. Human resources exist, but not financial ones, so it will be necessary to apply for some type of financing.

Is there relevant data (enough to be able to change the way the problem has been addressed so far)? Is it possible to access this data?
Yes, but a greater number of health centers should be contacted to request more exam images and thus increase the available sample size (N).

What are the project risks (ethical, social license, implementation risks, etc.)?
The results of eye fundus examinations are personal data, and therefore patients must give their consent for them to be performed. One of the main risks of the project originates from the implementation since it is necessary to obtain the commitment of the health centers, medical technologists, and ophthalmologists who will interact with the model so that they use the tool effectively. Given that the project seeks to reduce waiting times, which will also reduce the risk of blindness, it is highly likely that the general population will grant social license.

Theory   **1.3 Definition of objectives**

The objectives will be understood as the state desired to be reached in a given matter. For example, if at any given time a public body takes an average of five days to respond to citizen requests, one objective of the AI-supported project could be to reduce the response time to three days. Thus, the objectives express concrete progress from a situation of deficiency to one of satisfaction. Therefore, any indicators that are declared in the objectives must also be described in the problem definition.

A good methodology to define objectives is the one known by the acronym SMART, which means that the objectives must be **S**pecific, **M**easurable, **A**chievable, **R**elevant, and **T**ime-based.

As noted above, one of the most common mistakes is to define the objective as "create an automated response system" or "develop a predictive model", since neither really answers the central question at this stage: "Does achieving the goal help solve the problem?" The answer in these two cases would be "No", since the simple fact of having a predictive system does not imply that the problem in question will be solved. In other words, the effectiveness of the tool or model will depend not only on how accurate it is but also on how it is applied to its potential beneficiaries. Auxiliary activities of the project, such as "exploring the data" or "understanding the historical behavior of the program users", are also not project objectives.

When it comes to defining objectives, a good practice is to ask yourself "what for": Why do you need a predictive model? Why is it necessary to implement a chatbot on the website? Why do you need to know the behavior of users? By constantly iterating this "what for", the real objective of the project will finally be reached. In the synoptic table below, three examples of projects are presented with a brief description of the problem, an initial objective, and an improved objective according to the parameters discussed in this section.

Case	Problem	Initial objective	Weakness	Improved objective
Lead Poisoning Prevention[13]	X number of children have some level of lead in their blood. There are no acceptable levels of lead in blood. The main source of lead poisoning in children is in older homes whose walls are covered with lead-containing paint. Lead detection is done through a blood test that is performed on children when they enter the educational system. When lead is detected, an inspector is sent to the child's home and actions are initiated to remove this heavy metal, for which state subsidies are available. However, there are not enough resources to repair all houses covered with lead paint in cities.	Prevent lead poisoning in children.	The objective is very broad. While it does help solve the problem (child lead poisoning), it is not specific or related to the concrete action the city can take. There are also no success metrics offered, so it is difficult to know if they are achieved or not.	By 2019, the goal is for X% of total inspections to detect lead in homes where pregnant mothers live.
Diabetic retinopathy (DR) screening[14]	Currently, there is a delay of X months in the delivery of the results of eye fundus examinations (to establish whether or not a patient suffers from DR). Undiagnosed diabetic retinopathy, or a late diagnosis, can lead to total blindness. Currently, there is no human capacity (ophthalmologists and medical technologists) necessary to comply with a prudent delivery time for the results of fundus examinations.	Analyze the fundus examinations of all patients in the health system.	The objective does not refer to the problem, which is the delay in the analysis of fundus examinations. This would raise the rates of diabetic retinopathy which, without treatment, would increase cases of blindness.	Reduce, within a period of two years, the delivery time of results for those patients with a high probability of having diabetic retinopathy by X%.
Import control	It is estimated that currently X% of imports entering the country are fraudulent and are not subject to customs control.	Develop a predictive system for fraudulent imports.	The technical solution is not an objective.	Increase the percentage of fraudulent imports seizures over the total of controlled imports in the following fiscal year.

---

13   Lead poisoning prevention (USA): http://www.datasciencepublicpolicy.org/projects/public-health/poison-prevention/

14   DART (Chile): https://www.teledx.org/dart/?lang=es

Although the achievement of the objectives leads to solving the problems identified and allows reaching the desired state, it should not be forgotten that the projects suffer from limitations that can affect them and put the established goals at risk. Among the typical limitations are the lack of human and/or financial resources, and resistance to change, among others. Identifying potential constraints will allow the project manager to anticipate these obstacles at the implementation stage.

Don't forget to...

1. **Define** objectives about the public policy problem that needs to be solved.

2. **Set** SMART objectives: Specific, Measurable, Achievable, Relevant, and Time-based.

Activity After this section, complete the corresponding **Objectives** in the Project design and feasibility sheet.

PROJECT DESIGN AND FEASIBILITY SHEET

3	Objetives	
	Objetive	**Limitations**
1		
2		

PROJECT DESIGN AND FEASIBILITY SHEET

DART EXAMPLE:	Using Artificial Intelligence in Blindness Prevention	
3	**Objetives**	
	Objetive	**Limitations**
1	Decrease the delivery time of the results for those patients with a high probability of having diabetic retinopathy by X% within two years.	There are a limited number of ophthalmologists who can review fundus exams.
2	Reduce public spending associated with the treatment of diabetic retinopathy by 50% within 10 years.	Primary health care centers have to reach the coverage goals of the annual eye fundus examination for diabetic patients, in order to detect early retinopathy and prevent it.

After defining the project objectives, it is worth asking: What is the institution currently doing to solve the problem? How could your intervention be improved with the implementation of the AI tool? It is at this stage that the current attention or response processes of the institution regarding the detected matter must be clearly identified and that could be improved with the implementation of analyses or tools based on AI. Some actions of the institutions are the following:

- **Implementation of strategic products or programs:** Delivery of housing subsidies, allocation of food subsidies for students, planning of import inspections, etc.
- **Institution support processes:** Attention to users, hiring of personnel, planning of transport routes, etc.

At this stage, it is important to identify precisely who are the people currently carrying out these activities, as they will probably need to be included in the project team, and through which channels they are being implemented.

This had already been outlined in question 2 of the pre-feasibility analysis, Section 1.2: Is there the capacity to act on the problem? Is it within the powers of the organization? Are there human and financial resources required to act on the problem? At this stage, the project manager is expected to analyze in-depth the actions that the institution is currently carrying out to solve the matter in question.

The importance of correctly defining these actions derives from the fact that an institution executes multiple interventions and/or programs, but not all of them are related to the identified problem. At this stage, only those actions that could have an impact on the problem should be defined. **For example, if what needs to be solved is the excessive time elapsed between the moment a citizen's right of petition arrives and the response from the competent public entity**, a relevant action would be to determine what the bottlenecks are during the review of the request or to identify whether the notification system is experiencing delays.

In other instances, the same problem may concern different institutions, but the way it is addressed will depend on the mission of each one. For example, the fight against poverty is a very broad problem and therefore different ministries will be involved in various ways in its solution. The Ministry of Development, for example, can give vouchers to the most vulnerable families during the winter months to help pay for the higher expenses that this season of the year brings (increased visits to the doctor and higher food prices, for example), while the Ministry of Education can focus on universal access to education by appealing to social mobility that comes with more education. Therefore, response systems will differ depending on the capacities and powers of each institution.

Another case would be that of a project that seeks to increase the percentage of fraudulent imports seizures. Here the main action is to inspect imported goods, which requires a clear explanation of how the decision is currently made about which imports are subject to it. For example, each day it may be decided which cargoes to inspect based on the importer's past compliance history, or the inspectors in charge may decide on the spot, without further planning. It is important to consider the way the system currently operates, and how this change might affect the people in charge of

managing it. In this example, what if the inspectors receive a bonus for each load checked? There would then be an incentive to inspect a large number of cargoes, regardless of the outcome. Therefore, a predictive model reducing the number of loads to be inspected may arouse resistance in the team responsible for the task, since it would force its members to review fewer imports in search of greater efficiency, which in turn would reduce bonuses. Therefore, to ensure the viability of the project, it is essential to thoroughly understand the current modality with which the institution deals with the problem and incorporate someone responsible for these actions into the project team (head of inspections, for example) so that they can draw attention to possible weaknesses.

One of the most common mistakes made at this stage is to concentrate on ancillary project tasks such as collecting information, exploring data or linking different databases. Although these activities should be at the service of the project, they should not become its main activity. To avoid this mistake, it is necessary to go back to the issue to be solved and remember the chain of relationship between the different stages. Thus, a better response system (informed by the analysis or the tool) should help to achieve the objectives, which in turn will contribute to solving the challenge. The definition of the actions, together with the context in which they are performed and the people involved, is key to the success of the project at the implementation stage. However, unforeseen events sometimes arise that hinder the achievement of the objectives and that will have to be aborted so that they do not become obstacles.

Example: **Not everything is planned**

In Thailand, with the help of Google, an automatic system for detecting diabetic retinopathy by fundus analysis was implemented. The model worked very well in the laboratory, but when it was implemented it turned out that the facilities where the tests were taken were not sufficiently illuminated. Thus, the images were of poorer quality than those used to train the model and it was not possible to provide a reliable result on the presence of this disease. It was then necessary for patients to go to healthcare centers with better conditions, which rarely happened due to travel difficulties. On the other hand, the poor quality of internet connection made the wait for the result very long. The problem in this example is that Google did not incorporate the nurses—responsible for executing the action—in the planning stage of the project and therefore they were not aware of the problems of the care system to incorporate actions that could mitigate the effects of changes in lighting or poor internet connection before the project was implemented (Heaven, 2020).

If there is no capacity to act on the selected priority issue, and given the iterative process of project formulation, it may be necessary to return to the problem for complete reformulation. This is why pre-feasibility analysis is so important, as it prevents this situation from arising later on.

Don't forget to... (Box)

1. **Talk** to the people who are currently dealing with the issue to be resolved to determine how decisions are made and how they might be affected by the implementation of the AI tool.

2. Effectively **communicate** the implications and benefits of implementing an AI-based tool to gain buy-in from those who will ultimately use it.

3. **Incorporate** in the project team at least one of the people who was in charge of responding to the problem before implementing the AI system.

Activity   After this section, complete the corresponding **Actions** in the Project Design and Feasibility Sheet.

PROJECT DESIGN AND FEASIBILITY SHEET

4 Action description

Problem to solve	Action 1	Action 2
Actions		
Who executes the action?		
On whom or what is the action being performed?		
How often is the decision to take this action made?		
What channels are being used or can be used to perform this action?		
Other useful information about the action		

PROJECT DESIGN AND FEASIBILITY SHEET

DART EXAMPLE: Using Artificial Intelligence in Blindness Prevention

4   Descripción de acciones

Problem to solve	Action 1	Action 2
Actions	Fundus examination	Diagnosis of diabetic retinopathy by an ophthalmologist
Who executes the action?	Medical technologists in primary healthcare centers	Ophthalmologist
On whom or what is the action being performed?	People diagnosed with diabetes	People diagnosed with diabetes
How often is the decision to take this action made?	Annually per person	After fundus examination
What channels are being used or can be used to perform this action?	Healthcare centers that have the necessary supplies	Healthcare centers
Other useful information about the action		

The performance of the analysis or AI tool depends on the existence of the data that feeds them. Therefore, even before beginning, the project manager must explore the availability and quality of internal and external data that may be required. Regarding the latter, it should be consulted whether they are available and what the cost is. Some AI tools may only require open data. If necessary, at this stage the project manager should consult with the institution areas that collect or manage the data on some key aspects of the databases. It is not necessary to explore in-depth, as this task will correspond to the technical team in the execution stage.

The data must be sufficient for analysis and must have the required granularity. The data maturity matrix is a useful tool to know how mature (good) a database is. An adaptation of the Data Maturity Framework of the University of Chicago is presented in [Annex 2](#) of this manual[15]. This tool seeks to help the project manager to analyze, for each of the aspects on which it is intended to act, how advanced each of the databases that will be used in the AI tool is. It may be necessary to ask the data owners directly about their features and problems in this step.

An explanation of the features used in the matrix is given below.

Accessibility: Regardless of whether a database is internal or external, ease of access will be a major enabler of the AI project.

Storage: Institutions have massive amounts of data, but this can be stored in different ways. While data should ideally be in a digital database, many times the project manager will find records in physical format. This will require a pre-implementation step during which the data is digitized so that it is readable through the use of some software and usable in the relevant analysis. Since this can delay the execution of the project, it is of utmost importance to determine the status of the data in advance of this step.

Integration: Although institutions usually have a large amount of data, many times each of the databases is treated separately and independently of the others. The ideal scenario is that there is interoperability between the databases (both internal and external) so that it is easier to access the required information and unify the variables or key labels.

Relevance: The relevance of the data refers to its pertinence to the problem to be addressed.

Quality: It refers to the possibility that the database is incomplete or has errors. A database that does not include the data of a subject or observations about it (region, neighborhood or age, among others) can have important consequences on the resulting model since it could be biased towards a part of the population. A database with missing variables entails significant challenges since the technical team, led by the project manager, will have to decide to either eliminate the observation or impute data to it. Databases often also have typing errors, which occur at the time of collecting or transcribing the information. Here again, the technical team must decide, together with the project manager, what to do with those errors so as not to compromise the quality of the final tool. The ideal scenario would be for the databases to have none of the problems described above, which is rare.

---

15   http://dsapp.uchicago.edu/resources/datamaturity/

Frequency: It refers to the regularity with which data is collected, not the data itself (annual, monthly or daily observations). Data may be collected only once in the case of a one-time policy, for example, or it may be collected monthly, daily or in real time, i.e., as it arises. The important thing here is to determine what the data collection policy is, rather than to know exactly how many times the data has been collected.

Granularity: It refers to the level of detail of the data. One can have observations at the country, region, neighborhood, event (each time it occurs) or individual level. A more mature database will have a higher level of granularity, i.e., it will have more and more specific observations. Knowledge of the level of granularity of the databases is key to the success of the project. For example, a predictive model of school dropout will need data from each of the students; it is not enough to have information by geographical area, since this would not allow a correct targeting of prevention actions, which could be at the individual level.

History: The data from an observation may change over time. For example, a person may be single at age 25 but married 10 years later, so in the new data collection, the variable "marital status" will be different. There are several ways to store these histories: the most deficient one is to clean the history of the old data to replace it with the new information, while the most advanced one is to save the new data by recording its date and relating it to the previous ones.

Privacy: Regardless of whether personal or sensitive data is managed, the use, access and manipulation of this type of information must be regulated by a protection policy. Establishing a privacy policy implies regulating who (functions) have access to what type of data and for what type of projects. This can be achieved through an ad hoc process or software (usually a database management system) that handles the access policy.

Documentation: Databases usually use codes in the variables; in these cases, it is imperative to have a codebook or data dictionary that explains the meaning of these variables. Likewise, adequate documentation must include information on the metadata of the data set: its description, how it was extracted, and whether it is related to other data sets, among other aspects.

It should be emphasized that just because a database is rated as basic in the data maturity matrix does not mean that it is insufficient to train the tool. However, it is necessary to take into account that the technical team will have to invest more time and resources in handling it. This could mean that the project manager may have to put the development of the AI tool on hold momentarily to first implement more advanced data collection, storage and management policies, and then continue with the implementation.

If the data needed to create the tool does not exist, or if the granularity is insufficient, then it will be impossible to solve the selected problem through the use of artificial intelligence. This will require either reorienting the problem solution or going back to the definition stage to review in detail the availability of data in other similar cases that have used AI.

Don't forget to...

1. **Clearly** identify all the databases needed to carry out the project.

2. **Discuss** with the responsible parties the key features of the databases to assess their level of maturity and whether they are sufficient to carry out the project.

3. In case of requiring data from other institutions, **process** as soon as possible the necessary agreements to ensure timely access.

4. **Incorporate** data owners into the project team.

Activity After this section, complete the corresponding **Data Mapping** in the Project Design and Feasibility Sheet.

PROJECT DESIGN AND FEASIBILITY SHEET

5   Data Mapping

What data is available internally? No data available.
What data can be obtained from external, private or public sources?

	Fundus examination	Patient's data
What's the content?		
What level of granularity?		
How often is it collected and/ or updated once captured?		

Does it have unique and trusted identifiers that can be linked to other data sources?		
Who is the internal data controller?		
How is it stored?		
Additional comments		

PROJECT DESIGN AND FEASIBILITY SHEET

DART EXAMPLE:	Using Artificial Intelligence in Blindness Prevention

5	Data Mapping

	Fundus examination	Patient's data
What's the content?	Fundus examination images	Patient's medical record, sex, age, date of diagnosis, etc.
What level of granularity?	Individual	Individual
How often is it collected and/ or updated once captured?	In real-time	At the time of the examination
Does it have unique and trusted identifiers that can be linked to other data sources?	Yes, name and RUT ([Rol Único Tributario] Single Tributary Role)	Yes, name and RUT ([Rol Único Tributario] Single Tributary Role)
Who is the internal data controller?	Healthcare centers (Ministry of Health)	Healthcare centers (Ministry of Health)
How is it stored?	Images	Structured database
Additional comments		

Once the available data has been identified and depending on the nature of the problem, it is time to think about the data analysis or AI tools that will be needed to solve it. The role of the project manager is to identify the most appropriate ones, taking into account that these should inform and improve the current response process, which in turn should help achieve the objective, and, materialize the proposed solution.

AI enables various types of analytics—descriptive, predictive, and detection—and that is precisely its value: creating the ability to produce analytics that previously couldn't be conducted with a simple data processor.

Below are some examples of the use of AI in Latin America and the Caribbean according to the task performed by the AI system, based on the OECD categories (OECD, forthcoming):[16]

Recognition: It is mainly based on a categorization of images, texts and videos, through the identification of their key features, among other things. In Latin America, there are some examples of projects underway: DART (Chile), which, as has been seen, detects diabetic retinopathy by analyzing fundus examinations; Dymaxion Lab (Argentina), which monitors human settlements, floods and land use through satellite images; and MIDIS-Early Detection of Anemia (Peru), which detects cases of anemia by analyzing a photograph of the ocular conjunctiva.

Event Detection: The goal is to detect patterns and anomalies. An example is LAURA (Brazil), an AI platform that is used to identify those patients with clinical deterioration who are more likely to suffer from sepsis. This provides the clinical team with the necessary information about the patients to enable them to focus their attention on the highest risk cases.

Prediction: The goal is to predict a future state based on historical data. Some examples in the region are NotCo (Chile), a company that uses an algorithm to predict the ideal ingredients to create plant-based products that are generally of animal origin (mayonnaise, milk, etc.); Traive (Brazil), an alternative credit system that predicts the applicant's future performance; and Carabineros de Chile, which uses crime prediction models to forecast where it is most likely to occur, information that is in turn used in staff shift scheduling.

Personalization: The objective is to develop user profiles that, based on the data generated by their own actions, improve over time. Social media and streaming services use this tool to make personalized suggestions. An example is Livox (Brazil), a program that allows people with verbal disabilities to communicate through a selection of images. Over time the app begins to suggest likely images to users based on their past interactions.

Interaction support: This is a task that is related to the support of interaction between humans and machines (chatbots, virtual assistants, and others). There are some examples in the region such as Amanda Care (Argentina), a virtual assistant that accompanies patients during their medical treatments, and SpeakLiz (Ecuador), a software that transforms sign language into voice.

16 The cases listed here are a selection of those recorded in the document "La inteligencia artificial al servicio del bien social en América Latina y el Caribe: panorámica regional e instantáneas de doce países" [*Artificial Intelligence at the service of social good in Latin America and the Caribbean: regional overview and snapshots of twelve countries*" of the fAIr LAC Initiative of the Inter-American Development Bank] (Gómez et al., 2020).

Goal-focused optimization: It is about optimizing processes within an institution through scenario simulations. For example, Kilimo, in Argentina, aims to help agricultural producers optimize the use of water in their activities. To do this, it collects data from the field and combines it with historical information and satellite images in a patented software through which they provide: 1) seven-day crop water consumption estimates; 2) periodic advice summarized in plain language for producers on the amount of irrigation required to achieve their production goals, and 3) irrigation-related information for business intelligence purposes.

Reasoning with knowledge structures: Through this type of analysis, a causal relationship is established between the available data and a non-existent future event. Unlike simple prediction, this type of analysis focuses on causality between variables. An example of this is Portal Telemedicina (Brazil), which seeks to predict medical diagnoses based on patient records based on causal analysis. It is also a collaboration platform between diagnostic clinics and medical teams.

The types of analyses summarized here are only a selection of AI use cases in the region. The idea is that they serve as input for the project manager to communicate to the technical team which of them might be more appropriate to solve the public policy problem to be addressed. It is important to note that it is usual for the AI system used in a project to combine different tasks. Such is the case of the so-called composite systems that detect anomalous events and, from there, make predictions.

AI-based analyses also use a variety of tools, including regressions, decision trees and neural networks. The decision on the modeling technique to be used will be made by the technical team, whose members must explain to the project manager how the analysis was done and what were the errors and difficulties encountered, before deciding on the technique to be implemented with the tool. These aspects will be discussed in more detail in Section 2.2 on model building and validation.

AI system tasks	Possible analysis technique*
Recognition	• neural networks • support vector machines (SVM)
Event detection	• neural networks • support vector machines (SVM)
Prediction	• linear regression • regression trees • neural networks
Personalization	• collaborative filters
Interaction support	• neural networks
Goal-focused optimization	• linear, non-linear, dynamic programming, etc.
Reasoning with knowledge structures	• Bayesian networks

* Non-exhaustive list.

Don't forget to...

1. **Identify** the most appropriate type of tool and/or analysis to solve the selected public policy problem. It is not necessary to decide at the outset on the specific analysis technique to be used in the modeling.

2. **Meet** with the technical team (in charge of modeling) to discuss the public policy problem and agree on the type of analysis required.

3. **Include** the technical team within the project's multidisciplinary team.

Activity After this section, complete the corresponding **Analysis/Tool** in the Project Design and Feasibility Sheet

PROJECT DESIGN AND FEASIBILITY SHEET

6	Analysis/Tool

	Analysis 1
Analysis type (description, prediction, detection, behavior change)	
Purpose of the analysis (e.g., to understand the historical behavior of people, to estimate a patient's risk of disease)	
What action will support this analysis?	
How will this analysis be validated using existing data? (e.g., using historical data, running a randomized controlled trial)	

DART EXAMPLE:	Using Artificial Intelligence in Blindness Prevention

6	**Analysis/Tool**

	Analysis 1
Analysis type (description, prediction, detection, behavior change)	Detection
Purpose of the analysis (e.g., to understand the historical behavior of people, to estimate a patient's risk of disease)	Automatically obtain abnormal cases that can lead to diabetic retinopathy.
What action will support this analysis?	Action 2 (see Section 6 above "Action Description "). The ophthalmologist will now have filtered information when diagnosing the patient.
How will this analysis be validated using existing data? (e.g., using historical data, running a randomized controlled trial)	With trainings and evaluations with historical data from eye fundus examinations and with the corresponding diagnosis.

Theory — 1.7 Ethical, legal and governance considerations

At this stage it is necessary to make a diagnosis of the ethical, legal and governance considerations of the project. This will allow to measure its risks and plan appropriately according to its complexity. Again, it should be noted that algorithms and automated decision-making rules may reproduce biases and that, as a result, citizens may be reluctant to implement them. It is also possible that underrepresented population groups may be underrepresented in the models and thus be disadvantaged by the implementation of the tool. fAIr LAC has identified the following cross-cutting challenges for all AI projects:

Personal data protection: When implementing an AI-based decision-making and/or decision support system, it is likely to work with sensitive personal data[17] such as income level, medical diagnoses or other health information and unique identifiers such as the national identity document (DNI, for Argentina) or biometric data.

The project manager must ensure that the treatment given to said information complies with national data protection regulations and is guided by good international practices.

Personal data processing must always have a legal basis, that is, support that gives legal legitimacy to what is going to be done with them. Depending on the type of project, and the personal data regulations in each country, the legal basis may vary. The most common one is informed consent, as well as the organic laws of public agencies, legitimate interest, vital interest, or the existence of a contract. It is therefore important to be familiar with the data protection law of the country in question.[18]

A good practice is to have the consent of the people under study for their data to be used, whenever this is possible. Consent is understood as the "free, specific, unequivocal and informed manifestation of the owner's will through which he/she accepts and authorizes the processing of personal data that concerns him/her" (Red Iberoamericana de Protección de Datos, 2017).

For consent to be informed, individuals must at least know the identity of the person responsible for the project and the purposes for which the personal data will be used. They must also be aware of the security policy in the handling and storage of their data, the benefits associated with its use, and the potential risks and mitigation measures.

It is good practice to anonymize or pseudonymize personal observations; however, the project manager cannot forget that "personal data is not limited to names and surnames, but also includes any element that may lead to the identification of the specific subject" (Buenadicha et al., 2019). For this reason, care must not only be taken to eliminate the obvious personalization in the data, but also to code the possible data paths and crossings that would allow an individual to be identified.

Both the knowledge about the use of personal and sensitive data, as well as the need to have informed consent, must be taken into account in the planning and design process of the project. If the use of the data changes during its implementation, this must be duly communicated to the individuals so that they give their consent for the new use.

---

17   Personal data are those that allow a person to be identified directly or indirectly. (Red Iberoamericana de protección de datos, 2017).

18   For more information on personal data protection laws in different Latin American countries, see Bojalil and Vela-Treviño (2018).

Responsibility and transparency: The project manager will be ultimately responsible for the implementation of the AI tool and must communicate its benefits, risks and corresponding mitigation measures to the public. Responsibility is understood as the willingness to be accountable for one's own actions or activities, creations, or persons in charge, accepting the consequences of these acts (Oliver, 1994). In the case of AI, it implies responsibility for the decisions made by the system and for the effects that it may cause. To this end, the project manager must create simple mechanisms so that those affected can submit requests for review of decisions or complaints regarding the operation of the tool.

Regarding algorithmic transparency, it is an essential ethical principle in AI because it works as an enabler of other principles: it allows knowing what data is used, who uses it, how it is used, and how it affects public policy decisions (Sangüesa, 2018). When a system is transparent, it is possible to know, for example, if the data is protected and if the results are fair.

AI-based decision-making and/or decision support systems are often seen as black boxes: the general population does not know which of their data is being used, how it is used, what it means that decisions are now being made based on an automatic system, and what is the final impact that this will have on society. The role of the project manager here is to provide the public with the necessary and sufficient information so that they understand the general mechanisms of the tool, and so that they have security regarding the protection of their data and how decisions will be made.

Beyond the transparency of the model itself, it will be the responsibility of the project manager to inform the general public and stakeholders about the different steps and decisions taken during the implementation stage and even during the evaluation of the model in the field. A good practice is to document all decisions taken in order to keep a record and report appropriately when required.

Interpretability and explainability: This challenge arises from the need to understand AI-based decision support and decision-making models. Given that, as already indicated, algorithms are often seen as black boxes, the role of the project manager here is to explain to the population the general operation of the tool.

The explainability and algorithmic transparency of a model, as well as the amount of information that can be given about the operation of an algorithm, will depend on its level of opacity (Burrell, 2016; Buenadicha et al., 2019). There are three levels:

Intentional opacity: There will be situations where the algorithm cannot be explained to the public because it would put the effectiveness of the model itself at risk. For example, if a model is applied to predict tax evasion, the public should not know how it works specifically, as some might try to "cheat" the model. This intentional opacity can also occur for intellectual property reasons.

Illiterate opacity: It occurs when the people to whom the algorithm is explained do not understand how it works due to a lack of knowledge of technical issues. In this case, it is recommended to train the institution's front-line officials so that they can explain, in clear language, how the system works.

Intrinsic opacity: Sometimes the algorithms used are extremely complex, making it difficult to demonstrate a causal relationship between incoming and outgoing values. If this kind of opacity exists, it is impossible to explain the model.

Some tools are more explainable (easier for the general population to understand) than others. Therefore, the project manager must clearly inform the technical team about the requirements of explainability in light of national regulations and the problem definition, as this will impact the model that the team develops. Although it is not always necessary to provide all the information, citizens must be aware of how public policy decisions are being made.

In addition to the challenges described above, from the design phase of a project, it is important to ask about the possibility that the implementation of the tool may give rise to negative discrimination. These can originate in:

Bias: The system error is the difference between the predicted value resulting from the model and the actual value of the variable being estimated. If the error is systematic in one direction or a specific subset of data, it is called bias. For example, if a variable's value is consistently lower for one subgroup in the data, such as the salary of women with respect to equally qualified men for an equivalent job, the salary variable is biased. When the error is not systematic but random, it is called noise. ([González, Ortiz and Sánchez Ávalos, 2020](#)).

The bias of an AI system can have ethical implications when its results are used to make public policy decisions that lead to actions that may be considered unfair or prejudicial to some subgroups of the target population. For example, an AI model or tool could consistently favor or disadvantage the same sector of the population. There are different types of bias, some caused by problems intrinsic to the data. There are historical biases or undesired states, which are recorded when patterns appear in the world that do not want to be reproduced or propagated in the model; representation biases, which occur when information is incomplete, either due to missing attributes, sample design, or partial or total absence of subgroups; and measurement biases, which originate in the use or omission of variables that are going to be used in the models. (Harini Suresh, 2019).

Training an AI tool requires historical data from which it will "learn" and show results. It is up to the project manager, given his or her expert knowledge of the problem and the current situation, to anticipate which sectors of the population might be underrepresented in the data or which bias problems might affect them. Biases can be mitigated either by some remedial action or by calibration of the AI tool by the technical team at the implementation stage.

If the data is found to be biased, the tool could be biased too. A biased tool puts the viability of an AI project at risk. For example, Amazon developed an AI-based recruitment tool but had to withdraw it because it significantly favored the recruitment of men. This was because the tech industry has historically been dominated by men, making historical hiring data skewed. Given the arbitrary discrimination of the tool, it had to be scrapped (Dastin, 2018).

Inequalities in the process to intervene: It is likely that the current situation, without an AI project, will register inequalities between different subgroups of the population. For example, suicide rates are higher among LGBTQ+ youth than the general youth population. On the domestic violence side, the highest proportion of victims are female. The project manager should be aware of these baselines to incorporate case considerations and analyze whether or not an AI tool can help solve the problem at hand.

Subgroups of the population for which equality is to be ensured: While it is true that a database could be free of bias, it often happens that some minority groups in the population lose representation in the AI model. This is because, by using historical data, the model will give equal weight to all individuals or observations, which in itself leads to under-representation of those groups that constitute a smaller percentage of the population. As an expert on the public policy problem to be addressed, the project manager is the one who has clarity on how the implementation of the AI tool may affect minority groups differently, and who should therefore ensure equity in the intervention and implementation of the tool by defining protected groups or attributes. For example, if a disease detection system is to be implemented, it would not be acceptable for it to detect the pathology with different accuracy rates for men and women. Or if an automatic promotion model were to be applied in a certain institution, this could favor more experienced people over young people.

It will be the responsibility of the project manager to inform the technical team of any discrimination that may arise in the data, clearly identifying the population subgroups for which equality is to be guaranteed and taking measures to ensure algorithmic fairness (i.e., that the algorithm does not disadvantage some population subgroups and that the results are equitable for all). Metrics are now available to determine inequities that may arise in the modeling stage. The project manager should ask the technical team to conduct various tests to measure the level of algorithmic fairness in the tool (following definitions such as demographic parity, fairness of chances or counterfactual fairness, among others).[19] Given that there will be cases where it is not possible to intervene in the tool, measures to mitigate bias and discrimination can be adopted, not from AI but from public policy.

In addition to the above points, the project manager should ask at this stage whether the project has a social license. By social license is meant the acceptance, by the citizenry, of the implementation of the AI tool for a decision-making and/or decision support system (Data Futures Partnership, 2017). It is not enough just to comply with the countries' legal frameworks; an additional step is required, which is to obtain the acceptance of the population. Initially, AI was widely welcomed, given that it meant increased yields and decreased times for work, commuting, etc. However, in more recent times people have become increasingly aware of the potential adverse effects of AI on their lives, so it is key to obtain social license to implement such a project, regardless of whether it is done by a public or private institution (Hewitt, 2019). To find out whether a project has a social license, its manager may refer to opinion studies on the public value expected to be generated with the tool versus the risks it could entail, and hold meetings with stakeholders and communities where it will be implemented to learn about their perceptions in this regard.

Governance and security: Governance and security challenges are transversal to AI projects and are related to regulations and the security of the infrastructure in which the tool is developed. Here are three challenges:

Rules and regulations: The scope of an AI project will be framed within the rules and regulations of the countries where it is being implemented, and of the public policy sector in which the tool is used. It is expected that the project will be governed by the laws of personal data protection and access to information. It should be noted that there are specific sectors, such as health, that are regulated by special regulations and that several countries have anti-discrimination laws. Therefore, whenever an AI system is to be implemented, a compilation of all relevant standards is required. Having this information at the planning stage will contribute to the viability of implementation. It is recommended that the legal department of the institution be consulted and included in the project team.

19   For more details, see González, Ortiz and Sánchez Ávalos (2020).

Cybersecurity: AI systems are trained with data, so both personal data and algorithms results must be properly protected to prevent leaks of sensitive information belonging to citizens. That is why the project manager must establish security protocols for data storage and handling, and also for when information theft occurs.

Illegal penetration of AI and adversary attacks: Beyond the theft of sensitive information, the system may be subject to attacks that seek to confuse the algorithm, where the attacker will pose as a user of the tool. Here again, security systems must be installed to prevent these attacks, as well as response protocols in case they occur.

Don't forget to... (Box)

1. Make an inventory of the laws of the country applicable to the project to comply with legal requirements.

2. Have the necessary legal permissions to use information that may be personal and/or sensitive.

3. Determine all the ethical risks that may arise in the implementation of the tool.

4. Clearly define those population groups to which it seeks to guarantee equality in the intervention (algorithmic fairness).

5. Assess whether the project has a social license, i.e., whether the population accepts the proposed use of AI.

Activity

After this section, complete the **Ethical and Legal Considerations** section of the Project Design and Feasibility Sheet.

PROJECT DESIGN AND FEASIBILITY SHEET

7	Ethical and Legal Considerations
Proportionality	Do you think a data science/ AI system is the right way to solve the problem? Why? Have you evaluated other alternatives?
	What negative impacts could your project have? Study similar use cases identified in the "Problem Definition" section.
Social license	Do you think that the population will find the proposed use of data acceptable to solve the problem? Why?
	If the project population learns of the use of AI for the stated purposes, will they approve? Why?
Data Protection	Are you working with individually identifiable personal and/or sensitive data? Which ones?
	Have you identified the justification or legal basis for working with that data?
	Have you identified the regulations that could affect the project?
	Are mechanisms (for example, access, deletion or rectification mechanisms) required to ensure the quality of personal data?

Transparency	Which stakeholders should be aware of the project (policy makers, front-line workers, civil society organizations, public agencies, people who will be affected by the actions, etc.)? List specific organizations/individuals.
	Have you considered any mechanism for interested parties to contact the institution to obtain information about the project?
	Will it be necessary to explain the decision-making mechanisms or the analyses to be implemented? Why?
Discrimination/ equity	What structural inequalities are there in the process or environment where the project is inserted?
	Are there specific (vulnerable) groups for whom equity in outcomes or protection of their rights is sought (e.g., by gender, age, location, social class, educational level, urban or rural origin, ethnicity, etc.)?
	What biases do you think the data might have?
Accountability	In the event of a request for information regarding the project, who is responsible for responding?
	Who is responsible if the system is wrong?
	Do you have monitoring, control and evaluation mechanisms in place? How will they be documented and what frequency will they have?

PROJECT DESIGN AND FEASIBILITY SHEET

7	Ethical and Legal Considerations

Proportionality	**Do you think a data science/AI system is the right way to solve the problem? Why? Have you evaluated other alternatives?**	Yes, given that it is not feasible to substantially increase the number of ophthalmologists in the short term and that the project makes it possible to expand coverage. This will in turn contribute to increase the early detection of diabetic retinopathy and preventing it, eventually reducing the expenses of the Ministry of Health. Also, a tool is needed to significantly increase the number of fundus examinations that are reported. The option of training more ophthalmologists is a very long-term solution.
	What negative impacts could your project have? Study similar use cases identified in the "Problem Definition" section.	A possible risk is that the quality of the images is insufficient to perform the analysis. An adverse impact could be that the test results are false negative and therefore not analyzed by an ophthalmologist. This would cause harm to patients whose results are incorrect and who do suffer from that disease.
Social license	**Do you think that the population will find the proposed use of data acceptable to solve the problem? Why?**	Yes, because patients will obtain a clear benefit in terms of the opportunity to have their examinations reviewed, which increases coverage of the annual check-up and frees up the time of ophthalmologists, who will now be able to focus on the most extreme cases. As for possible resistance from ophthalmologists and medical technologists, this could originate from the fact that they will have to change the method of performing the examinations and diagnosing the cases. It is important to include them in the project to gain their commitment to the implementation.
	If the project population learns of the use of AI for the stated purposes, will they approve? Why?	Yes, since it has a clear benefit for everyone, given that it frees up human and financial resources of the health centers and focuses existing resources on the cases that need them. In any case, it is important to report on the long-term benefits of using the tool in health centers, given the initial investment to be made by the Ministry of Health. However, there may be some resistance to the result of an examination being produced by an algorithm. It is then necessary to incorporate the human factor and have a specialist confirm the cases of diabetic retinopathy found automatically, so that the healthcare team can then communicate the diagnosis to the patient. At the time of implementation it will be necessary to adjust patient expectations, since the tool has only been designed to detect diabetic retinopathy and not other eye diseases.

Data Protection	`Are you working with individually identifiable personal and/or sensitive data? Which ones?`	Yes, they are sensitive data because they refer to people's health status.
	`Have you identified the justification or legal basis for working with that data?`	Yes, personal data protection according to Law No. 19,628. Law of rights and duties of the patient (Law No. 20,584) and regulations thereof, CSAN, DFL 1 Salud de Chile.
	`Have you identified the regulations that could affect the project?`	Yes, personal data protection according to Law No. 19,628. Law of rights and duties of the patient (Law No. 20,584) and regulations thereof, CSAN, DFL 1 Salud de Chile.
	`Are mechanisms (for example, access, deletion or rectification mechanisms) required to ensure the quality of personal data?`	Patients must be able to access their fundus examination images and the results of the ophthalmologist's examination if they require them, for example, to attend a private medical consultation. Otherwise they may refuse to have the examination performed.
Transparency	`Which stakeholders should be aware of the project (policy makers, front-line workers, civil society organizations, public agencies, people who will be affected by the actions, etc.)? List specific organizations/ individuals.`	The heads of health services and clinics that implement the software, the Ministry of Health, the Chilean Society of Ophthalmology (SOCHIOF), the Chilean College of Medical Technologists, the National Confederation of Municipal Health Officials (Confusam), the Chilean Diabetic Association (Adich) and the Juvenile Diabetes Foundation of Chile.
	`Have you considered any mechanism for interested parties to contact the institution to obtain information about the project?`	Yes, through periodic meetings with those responsible in the Ministry of Health. The Ministry of Health, for its part, also collects perception data in health centers.
	`Will it be necessary to explain the decision-making mechanisms or the analyses to be implemented? Why?`	Yes, although not the algorithm itself because it is highly technical. It must be possible to explain its clinical validation.

Discrimination/ equity	**What structural inequalities are there in the process or environment where the project is inserted?**	There are inequalities in terms of diabetes risk factors (Sapunar, 2016). The risk of diabetes is higher in people who are older, female, and who live in urban settings and have a lower educational level. It has also been found that there is less risk in indigenous peoples.
	Are there specific (vulnerable) groups for whom equity in outcomes or protection of their rights is sought (e.g., by gender, age, location, social class, educational level, urban or rural origin, ethnicity, etc.)?	No; examinations of all diabetic patients will be analyzed. It is possible that the algorithm may perform differently for some subgroups of the diabetic population and therefore it would be advisable to carry out a disparity analysis.
	What biases do you think the data might have?	None, since the risk factors described above should not influence the quality of the fundus image.
Accountability	**In the event of a request for information regarding the project, who is responsible for responding?**	The Ministry of Health and Teledx
	Who is responsible if the system is wrong?	The Ministry of Health
	Do you have monitoring, control and evaluation mechanisms in place? How will they be documented and what frequency will they have?	Yes, an observational study is planned once the tool is implemented in enough centers in order to obtain a sufficiently large sample. Periodic monitoring of the tool will also be conducted in terms of prediction quality, algorithmic biases, user experience and exam access gap analyses.

Theory 1.8 Team composition

A critical part of the project planning and design process is to identify the people from within the institution—or outside—who should be part of the multidisciplinary team in charge and involve them from the beginning. Their active participation in the formulation will help ensure the sustainability of the tool during implementation.

Throughout this entire process, the project manager has been talking with people inside and outside the institution about the proposed initiative and incorporating their learning into the design. For example, if it is a project to optimize customs inspections, he or she will probably have met with the head of customs operators to gain a thorough understanding of how they are currently carried out, how they decide what to inspect, how long it takes and what problems are encountered in the field. It is also likely that he or she will have contacted the statistics department to request information on existing data and analyze its maturity. All of these people should be part of the project team, as the impact concerns them as well. In the case of the team in charge of audits, they will be the ones using the AI tool, so it is important that they are fully aware of the current and future implementation challenges. The statistics department will most likely be in charge of managing the data internally, so it will be their experts who will provide it to the technical team.

The following is a description of some of the possible positions of the possible members of an AI project and their functions. This list is only a suggestion and therefore does not address in detail the composition of the technical team, as this will depend on the nature of the project.

Project manager (responsible for decision-making)	They perform the main role in the planning and design stage and are in charge of supervising all the steps taken by the technical team in the execution stage.
Process referent	They participate in the process where the tool and AI will be implemented; for their experience and knowledge, they are respected by their peers. Their expert knowledge and recognition will contribute to the adoption of the technology.
Responsible for the data (internal and/or external)	They provide information on the quality of the data needed for the implementation of the tool. They determine the needs in terms of collection, storage and handling. They apply their expert knowledge to detect possible anomalies in the data.
Data analysts	They perform the exploratory analysis of the data and develop the tool. If these capacities do not exist within the institution, the project manager must contract this service with an external institution.
Legal team	They ensure compliance with the legal regulations present in each country regarding the personal data protection, transparency and sectoral regulations.

Institution's Directive Line	They support and promote the development of the tool. It is key for the project to be viable and to be implemented.
Team executing the action	They have expert knowledge about how the problem is currently being responded to and how it could be improved. They raise concerns about how officials will interact with the tool and possible consequences.
Communications	They communicate internally and externally the key aspects of the implementation of the tool.
Department of studies or evaluation	They contribute to the design of the methodological aspects of project evaluation.

Don't forget to...

1. **Consider** all relevant stakeholders when forming the team.

2. **Achieve** the commitment of the team members to the project.

3. **Ensure** that all team members participate in the planning and design of the project.

Activity   After this section, complete the corresponding **Team Composition** in the Project Design and Feasibility Sheet

PROJECT DESIGN AND FEASIBILITY SHEET

8	Team Composition	
Organization/ epartment	**Counterparty name/role**	**Description of desired participation**

PROJECT DESIGN AND FEASIBILITY SHEET

| 8	Team Composition	
Organization/ epartment	**Counterparty name/role**	**Description of desired participation**
Teledx	Data Analysts	In charge of creating the image analysis tool
Ministry of Health	Undersecretariat of Healthcare Centers	Implementation of the tool in health centers
Ministry of Health	Legal Division	Advice on project regulatory compliance
Teledx/ Ministry of Health	Training Team	Training for ophthalmologists and medical technologists on the use of the tool

PART 2.
EXECUTION
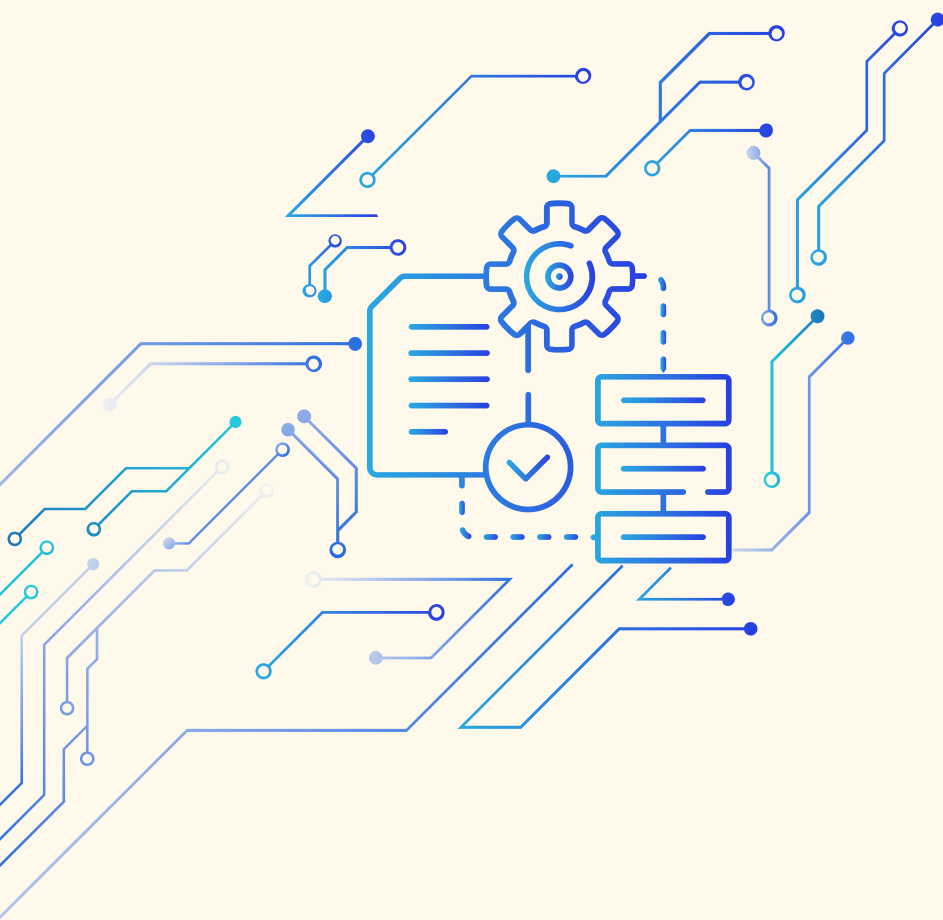
PART 2. EXECUTION

Once the project has been formulated, and the institution has decided to carry it out, it is time to execute it. How is it done? What aspects need to be focused on during this phase? How far should transparency and accountability for the specificities of the model go?

This manual describes the challenges involved in an AI project for both its manager and the technical team, depending on its life cycle. These challenges can also be understood through the different stages that the responsible team must develop, where the manager performs a series of key functions (Figure 5):

Figure 5. Connection between public policy and AI development

PUBLIC POLICY LIFECYCLE   AI LIFE CYCLE
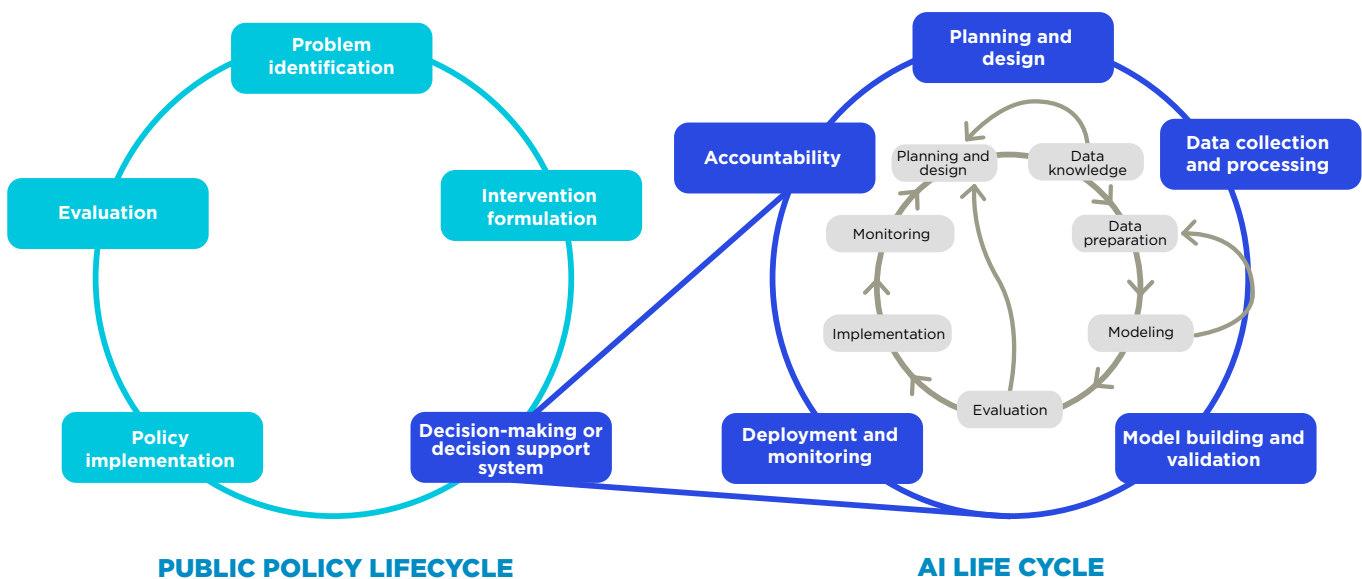
Source: González, Ortiz and Sánchez Ávalos, 2020.

1. Data collection and processing: In the planning and design stage, the project manager already mapped the different databases and determined their degree of maturity. At this stage, the technical team will start an exploratory analysis of the data to obtain more information and identify the possible challenges that may arise during its use.

2. Model building and validation: Once the variables to be used have been defined and the challenges presented by the data have been understood, it is time for the technical team to start developing the model using specific modeling techniques. At the planning and design stage, the project manager should have communicated to the technical team the objective of the analysis: "What type of analysis (or tool) is most appropriate to solve the problem"? At this point, the technical team must present different modeling options to the project manager, and together they will decide on the most appropriate technology according to technical and equity criteria. It is possible that during the

development and validation of the model, the technical team identifies new data needs, so iterations of the model must be carried out integrating the new sources of information. This process should be repeated until the models generated are satisfactory to the technical team and the project manager.

3. Deployment and monitoring: At the time of implementing the AI tool in the target population, or in a subset of it, its impact should be measured taking into account the project objectives. It will also be necessary to implement a periodic monitoring system to correct errors that may arise during its implementation and allocate the necessary resources to ensure the use of the tool over time.

4. Accountability: As the project manager is the person who is ultimately in charge of the implementation of the AI tool, it is his or her responsibility to document, together with the technical team, the entire development process in order to be able to justify to the public and other stakeholders the decisions made throughout the AI lifecycle.

In steps 1 and 2 described above, the technical team will be the main responsible for the model, since it will be in charge of modeling the data. However, the project manager will be responsible for supervising each of the steps and decisions of the technical team. At the beginning of each step, the role of the project manager is presented.

In the manual by [González, Ortiz y Sánchez Ávalos](#) (2020), to which reference has been made throughout this text, there are two materials aimed at the technical team that will be included below to promote fluid communication between the team and the project manager: the data profile and the AI model card. Both will support the documentation of project decisions for the sake of transparency and accountability.

1. Data profile:
This includes the main findings of the databases that will be used with the tool. It is based on the exploratory analyses of the data carried out by the technical team under the supervision of the project manager, with whom the findings will be discussed.

2. Model card:
Final description of an AI model. It includes critical aspects for the entire AI lifecycle.

For both tools, the roles of the project manager in each of the key aspects of the execution stages are explained.

2.1 Data collection and processing

Teoría

Project manager role

During the data collection and processing phase, the project manager will be in charge of:

- contextualizing the findings of the technical team in the exploratory data stage;

- providing background on the data collection process and the meaning of the variables, connecting the technical team with the people in the organization required for a better understanding of the data;

- validating the formulation and usefulness of the target variable;

- making decisions on data imputations or elimination of variables, weighting and recording what is gained versus what is lost (trade-offs);

- proposing improvements to the project's data governance processes that can contribute to the achievement of the objectives and facilitate model implementation; and

- making adjustments to the problem definition, objectives, and project implementation plan based on the results of this stage.

In this instance, the technical team must first carry out an exploratory analysis of the data. Although it is to be expected that the project manager has a basic knowledge of the databases and their maturity—according to what is detailed in Section 1.5 Data Mapping—, it is at this stage that the technical team proceeds to a detailed and complete analysis of the data using conventional statistical tools.

Here the role of the project manager is to act as a bridge between the technical team and the areas that generate or manage the data, providing information on the collection methodology, the update frequency, and even the dictionary of variables. The purpose of this is to make communication between both parties more fluid in case methodological or conceptual doubts arise.

Secondly, when the exploratory analysis is carried out, the technical team may come across situations considered anomalous. Here, given that the project manager is the person who has the greatest specialized knowledge about the public policy problem that is sought to be solved and who can connect with sectoral specialists who have the relevant expertise, he or she should help identify the potential findings in the data that may be the product of some sectoral context unknown to the technical team.

Several challenges related to the quality of the available data may arise at this stage of the AI project, some of which are listed below. These challenges are also listed in the manual by González, Ortiz and Sánchez Ávalos (2020) for the technical team[20], although here they are approached from the perspective of the project manager, who, as the person ultimately responsible for the tool, will have to solve them.

---

20 https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial

Database quality and relevance

Through the data mapping carried out in the planning and design stage, the project manager already has an approximation of the data that allows him/her to determine if it is complete, relevant and has the necessary granularity for analysis. However, in its exploration, the technical team may encounter additional challenges related to the quality of the database, such as:

Undesirable or suboptimal states in collected data. If the data is biased, the tool trained on it will also be biased, putting its implementation at risk. To identify these biases, the project manager will ask the technical team—with whom the adverse implications of having a biased tool will have been discussed in advance—not only to carry out an exploratory analysis of the data as a whole but also to take into account, based on the previously defined concept of algorithmic fairness, those subgroups of the population that may eventually be affected.

The variables for which data are available are not ideal. It may happen that the necessary data is not available for the target variable required by the project, i.e. the one that is to be explained or predicted (Greene, 2003). In such a case, the decision must be made to replace it with a proxy variable for which the necessary information is available.

For example, if it is necessary to measure work experience and this data is not available in the database, but the age and years of schooling of the individual are available, then a substitute variable could be a combination of the two previous ones. Although this seeks to approximate the target variable as closely as possible, it is not the same thing and consequently, its use can affect the quality of the predictions. In this example, the proxy variable does not identify periods of unemployment, prolonged travel abroad or even parental leave, which is especially complex for women (depending on national regulations on maternity leave). Therefore, in the absence of the ideal variable, the project manager should communicate to the technical team the need to develop a proxy variable, if possible.

For its part, the technical team should show the results of that variable, clearly explaining its advantages and disadvantages. It will be up to the project manager to decide which proxy variable to use or whether to rethink the problem based on the available data.

Incomplete information about the target population

As seen in Section 1.5 on Data Mapping in this document, there may be missing or incomplete attributes in the database. The main challenge for the project manager will be to approve the treatment to be given to said observations, based on the recommendations of the technical team. The following is a summary of other problematic situations of which the project manager should be aware:

Relevant probabilistic and natural samples. If the database is fed from surveys or interviews, the project manager must know that these data correspond to a subgroup of the population. There are two possible extremes of sampling that are related to the randomness of the sample selection[21]. A non-random sample may lead to participation bias, so the project manager should be clear that it may not be representative of the general population. By analyzing the sampling method, the project manager may decide to modify the way in which the data is captured, which would put the project on hold until the data meets the defined requirements.

---

21 For more information, see González, Ortiz and Sánchez Ávalos (2020, p. 18).

Missing or incomplete attributes. Through the exploratory analysis carried out, additional situations of missing or incomplete attributes can be detected. There may be several reasons for this: problems caused by the people responsible for entering the data; system failures; transcription problems and/or lack of information on that attribute for the subject in question. When attributes are missing or incomplete, the project manager should ask the technical team about the nature of these errors and whether they are random or not. Depending on their nature, and the number of errors, a decision must be made to impute data or to remove observations from the data, reducing the sample size (both have advantages and disadvantages.)

Causal comparison

The project manager should be very clear that traditional machine learning methods do not establish causal relationships between the variables used (Varian, 2014). However, he/she will also know that there are techniques to try to establish causality in an analysis, which should be communicated in a timely manner to the technical team so that they can decide which is the most appropriate tool for this purpose. If causality analysis is required, then the necessary data should be collected according to the chosen causality methodology.
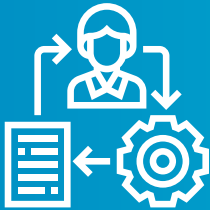
By carefully analyzing the available data and identifying its limitations, the project manager and the technical team can take future steps to improve the way information is collected and stored. For example, if there is no data on the desired target variable, can the way the information is collected be changed to capture it?

Activity After this section, complete with the technical team the **data collection and processing in the Model card, as well as in the Data Profile.**

The Data Profile and the data collection and processing section of the Model card are presented below, highlighting the role of the project manager in the development of these tools, which should be completed by the technical team[22].

MODEL PROFILE		
Data collection and processing		
3. Training data [23]	Data set used and labeled	
	Preprocessing steps or data preparation	
	Potential biases and deficiencies depending on the use case (2)	

Project manager role

Review, together with the technical team, the results of the exploratory data analysis. Explain, based on expert knowledge, the anomalies found (if any) and the results of the analysis, and contextualize the main findings.

Ask the technical team for exploratory analysis of the data by population subgroups. Adopt bias mitigation and algorithmic fairness measures.

DATA PROFILE
Overview of data and motivation
Name of the dataset used
Which institution created the database?
For what purpose did the institution create the database used?

Project manager role

For the data (both internal and external) to obtain detailed information on the databases, the method of collection and possible weaknesses in the information.

---

22 For more information on both profiles (data and model), see González, Ortiz and Sánchez Ávalos (2020).

23 This numbering corresponds to the model profile tool in the manual by González, Ortiz and Sánchez Ávalos (2020).

What mechanisms or procedures were used to collect the data (e.g. household survey, sensor, software, API)?	Discuss with the legal team the restrictions on the use of the data, given the way it was captured. For example: Is it necessary to obtain informed consent? Is it public or private information?
Number of individuals for whom data has been collected	Depending on the number, discuss together with the technical team if this sample can be representative. If not, how will this affect the training of the tool?
Capture frequency (weekly, monthly, daily) or an average number of observations per individual. Will the dataset be updated (for example, by adding new instances and/or removing others)?	
Essential controls	
Get documentation for each variable within the dataset. Brief description including its name and type, what it represents, how its value is measured, etc.	
Do an exploratory analysis of the data. Calculate descriptive statistics identifying the percentage of missing values and determine the distribution of each variable within the database.	Analyze the results of the exploratory data analysis. Ensure that the maturity of the database is adequate to solve the public policy problem.
Analyze the spatial and temporal coverage of the data.	
Analyze coverage of protected groups (sex, race, age, etc.)	Communicate to the technical team which groups will be considered protected.
Describe the important dimensions on which the data sample may differ from the population, particularly unmeasured selection biases. Use related literature and expert information.	Help determine the reasons that explain the differences found from expert knowledge. Share these inputs with the technical team.
Identify "undesirable states" in the data, such as biases and inequities harmful to subgroups, or any other pattern that is considered suboptimal or undesirable from a social policy point of view.	Explain, based on expert knowledge, those data biases that may arise from an undesired state of collection. Define what is considered a suboptimal state. Discuss with the technical team possible bias mitigation measures for protected groups.
Are values missing? If yes, please explain why that information is missing (This includes information intentionally removed). Determine if missing data is associated with the variable to be predicted.	Help determine the reasons for missing values (for example, an adverse circumstance that did not allow correct data collection) through expert knowledge.

Teoría   2.2 Model building and validation

Project manager role
During the model building and validation phase, the project manager will be in charge of:

- evaluating the model with metrics that consider not only efficiency and effectiveness, but also equity;

- deciding the model to be implemented taking into account the human and financial resources available in the institution; and

- understanding the limitations of the model and adjusting the project implementation plan based on the results of this stage.

This section will explain some basic concepts of the model training phase, a task that corresponds to the technical team, although it is the project manager who will ultimately decide which model to implement.[24]

How is a model developed?
In section 1.6 of this manual (Tool/Analysis), the project manager should have already informed the technical team about the type of AI-based tool or analysis that will be required to solve the problem. While some specific AI tools have already been introduced at this stage, the precise modeling technique will depend on the nature of the issue to be resolved and the available data. Here the technical team must test different options that will then be presented to the project manager and the multidisciplinary team.

In the case of a supervised learning model, when it is fed with information about the variables, the data should be divided into at least two large sections (ideally three):

- *Training data:* They are the ones used to train a model.

- *Validation data:* They are the ones used to validate the model according to the previous training.

- *Test data:* They are the ones that remain hidden until after selecting the model and are used to confirm the results.

The technical team will be in charge of dividing the data according to the information available, but it will be up to the project manager to decide which data (not all) will be used in the training phase. Furthermore, the project manager must inform the technical team about possible anomalous situations in the data as a result of certain contingencies (extreme situations at the national level, change in the definition of a variable, etc.), so that the team can take the necessary measures when using the data. All the models will be trained with the data designated for it, which will then be refined using the validation data. Finally, its operation will be evaluated with the test data.

The project manager needs to keep in mind that models are simplifications of reality. Therefore, they will not be 100% accurate and are not expected to be. It is about the AI model having a sufficient level of adjustment and that it can, for example, predict future results, for which a certain percentage of error in the validation and testing process will be unavoidable.

---

24 For more information on modeling and validation, see González, Ortiz and Sánchez Ávalos (2020).

The confusion matrix is described below—a tool that helps validate the performance of an AI model during its development—as well as the role of the project manager and the technical team in its implementation.

- *Confusion matrix:* When the target variables are categorical (yes/no), adjustment metrics can be obtained in terms of false positives and false negatives, which are obtained according to the predicted value versus the real data. For example, for binary classification, the following confusion matrix is available:

		Real	
		Positive	**Negative**
Prediction	**Positive**	True positive	False positive
	Negative	False negative	True negative

Based on these errors and successes in the prediction, it is possible to elaborate on different metrics for adjusting a model,[25] a task that will be the responsibility of the technical team. For his/her part, the project manager must be aware of the levels of classification errors in the model to be implemented, according to the definition of the problem. False positives occur when the model establishes, for example, that a person is a beneficiary of a program but in reality, they should not be. On the other hand, false negatives occur when, according to the model, a person should not receive benefits from the program when in fact they do. Beyond the metrics developed by the technical team depending on the nature of the problem, the project manager must decide what type of error the project supports the most. For example, in the above case of benefit delivery, and taking into account that resources are scarce, the fact that a person does not receive a benefit to which he or she is entitled (false negative) may be a more serious error than delivering benefits to who does not need them (especially if a deeper analysis of the thresholds to be a beneficiary can be made and it is discovered that the person is actually very close to the threshold).

However, whether the errors produced by the model are acceptable will also depend on the resources of the institution. Thus, the project manager must not only be aware of its occurrence but must also carry out a cost-benefit analysis of it. Continuing with the previous example, if the institutional budget is very tight, limits could be set on the delivery of benefits when they do not correspond (false positive), which will have to be balanced against not delivering them when they do correspond (false negative). Performing this cost-benefit analysis is key before implementing the project.

Note that these errors are all made "in the laboratory"; therefore, it is necessary to keep in mind that other challenges may arise during implementation. The tool may be very difficult to execute or poorly accepted by users. If so, one could return to Section 1.6 to analyze again the need for the project, and also to the beginning of Section 2.2 to evaluate another AI-based tool whose implementation is more likely to be successful.

---

25 See González, Ortiz and Sánchez Ávalos (2020, p. 16.).

Equity and differential performance of predictors
Taking into account the above, the project manager must study in detail the biases and inequalities that may arise in the model to make the decision, either to correct them or to implement policies and mitigation measures.

As discussed in the previous section, biases can appear in the data and can be mitigated from there. However, with error metrics, a model may favor a subgroup of the population. For example, suppose the diabetic retinopathy detection model is more accurate for men than for women, a bias that does not necessarily come from the data but from the model itself. The project manager must be clear about this bias, make it known to the public, and mitigate it with public policy measures. One possible solution would be that, while the model is being recalibrated, all fundus examinations for women should be sent to an ophthalmologist for a thorough analysis. This would ensure greater accuracy in the results of such tests in women and feed the model with these analyses.

It may also happen that the objective of a project is to favor a certain subgroup of the population. Take, for example, the case of a subsidy that has historically been assigned to men, but now seeks to benefit at least 50% of women. Here the objective of the prediction tool (sensitivity, precision and accuracy, among others) does not correspond to the objective of the project, namely, to reach 50% of female beneficiaries. For this reason, the project manager must effectively inform the technical team of the project's objective—which may be different from the modeling itself—and clearly identify the subgroups of the population to be benefited.

Section 1.7 of this manual noted the importance of the project manager being aware of the ethical and legal implications of the project, but also of those of the AI model itself. Now, with more detailed information about the model to be implemented, the project manager should go back to that section and check that no other ethical risks, biases and discriminations have arisen as a result of the training. In Section 2.1 on data collection and processing, the biases that can occur have already been discussed, but as noted here, the model could give rise to other biases without necessarily biasing the data.

The technical team will inform the multidisciplinary team which model will be implemented, although it is up to the project manager to make the final decision taking into account the nature of the problem, the financial and human resources available in the institution, and the risks and biases that the implementation of the tool may cause.

Activity At the end of this phase, it is recommended that the Model building and validation section of the **Model card tool** be reviewed with the technical team and according to the roles identified.

The section on Model building and validation is presented below, highlighting the role of the project manager in the development of this tool, to be completed by the technical team. This profile is detailed in González, Ortiz and Sánchez Ávalos (2020).

MODEL CARD		Project manager role
Model building and validation		
4. Modeling	Algorithms used in training, assumed parameters or constraints	
5. Performance metrics	Technical metrics used to select and evaluate models	
	Cost-benefit analysis of the model for its use case	Carry out the cost-benefit analysis of the implementation of the tool, asking if the public policy problem can be solved without using AI.
	Definition of protected groups and selected equity measures	Define measures of equity in the results (subgroups of the population for which equity is to be ensured).
6. Validation data	Data sets used and their labeling	
	Preprocessing steps	
	Evaluation of adaptation of validation data according to the use case	
	Potential biases and shortcomings depending on the use case	Establish, together with the technical team, acceptable bias thresholds based on expert knowledge. Prepare mitigation measures based on public policy, if required.
7. Summary of quantitative analysis	Reported validation error	
	Summary of cost-benefit analysis	
	Report on equity measures for protected groups	Decide, together with the technical team, which model will be implemented.

Teoría 2.3 Deployment and monitoring

Project manager role

During the model deployment and monitoring phase, the project manager will be in charge of:

- ensuring the training of public servants who interact with the AI model so that the tool is sustainable over time;

- developing a user manual focused on the civil servants who will interact with the model;

- establishing feedback mechanisms for people interacting with the model;

- designing and implementing simple administrative processes to correct those errors present in the model that affect users;

- establishing automated, or at least periodic, model monitoring systems;

- maintaining a record of model results taking into account access and security restrictions;

- implementing necessary model and process improvements based on monitoring and evaluation findings, and

- allocating the necessary resources to maintain the tool over time.

Once the most appropriate model has been chosen in the previous stage, it is time to put it into practice. At this stage, it is important that the project manager designs a pilot of the tool to monitor its use under limited conditions, before deploying it among the entire population or the entities involved.

One of the critical aspects to consider is the training of public servants who will interact with the model. This should include information on its key aspects, its objective, the way it works, and the way its feedback systems will operate. If the right people were included at the design and planning stage, the related teams will already have information about the project, so their involvement and subsequent training will help mitigate the risk that the tool will never be used. It may be necessary to create a special training team within the institution. It is also suggested that an internal user's manual be developed for staff members who will interact with the model. This manual should be a summarized version of the tools presented here. It is suggested that, at a minimum, it should contain:

- the definition of the public policy problem;

- the definition of the objectives;

- the definition and justification of the protected population;

- the regulations on the use and storage of the data (together with the legal considerations relevant to each country);

- a simple explanation of how the tool works and the variables used in its training;

- a section of frequently asked questions, formulated in clear language, that may be asked by people affected by the implementation of the tool;

- contact information for expressing doubts and requesting additional information.

Effectiveness evaluation
This stage involve the model evaluation design, no longer in the laboratory—explained in Section 2.2 on its development and validation—, but in the context where it will be deployed, to measure its real effectiveness during the project implementation. The objective is to be able to attribute causality to the implementation of the AI tool, for which there are various impact assessment techniques, including natural experiments, randomized controlled experiments and quasi-experiments (Shadish, Cook and Campbell, 2002). If possible, it is recommended to favor randomized controlled experiment methods, given their high level of validity.

In order to carry out a randomized controlled experiment, it is necessary to have the informed consent of the participants and to take the necessary precautions against the ethical risks involved in issues such as experimenting with human beings, depriving potential beneficiaries of a measure because they are part of a control group, and making random versus necessity assignments, in addition to other problems of a legal nature (Shadish, Cook and Campbell, 2002). While randomized controlled experiments are one of the preferred methods to measure impact, the tool used in the project will depend on the data available and the ethical and legal restrictions in each country. The project manager will ultimately decide and communicate to the technical team how the effectiveness of the tool will be measured in the real world.

The success of the tool, and therefore of the project, will be determined by the achievement of the objectives. In this sense, the purpose of an impact evaluation **will not be to measure the level of adjustment of the model to reality**, but rather its fulfillment. An adequate measurement of compliance with the established goals will require having a clear baseline on the performance of the current process (without an AI tool), which will be defined in the exploratory data analysis stage. If the tool does not achieve the expected impact compared to the previous state, the implementation should be reviewed for flaws, or it should return to the model building and validation phase to produce a new model, or the project should be reformulated from the planning and design stage.

Performance degradation
The project manager should be aware that the results of a model can degrade over time due to a variety of reasons:

- the relationship of input and output variables may change;

- the way data is collected and stored may change;

- the user-tool feedback systems may be closed, which means that the interactions do not lead to the incorporation of new data originating from emerging situations that enrich the model.

To mitigate these risks, it will be necessary for the project manager to monitor the behavior of the input and output variables and, together with the technical team, update the assumptions of the model. Automatic systems can be set up, or with a certain periodicity, to monitor the results of the tool

and the behavior of the previously established error and fairness metrics. The frequency will depend on the nature of the problem.

There must also be a **system for feedback on the experience of using the tool among the users, the final beneficiaries, the technical team and the project manager**, who will be exclusively responsible for it. The objective of implementing a feedback system is to have timely information about the difficulties and errors detected by those interacting with the system, in order to correct them as quickly as possible, either by returning to the previous stage or by means of corrective public policies.

At this stage, it is key that all the necessary resources—human and financial—are allocated to retraining the model and training the people in the institution who will interact with it. It is important that, when implementing an AI-based solution, it is sustainable over time and that it is adjusted according to new data, emerging technologies and/or new user interfaces.

Activity

At the end of this phase, proceed to review, together with the technical team, the **Deployment and monitoring section of the tool in the Model card**, according to the functions identified.

The Deployment and monitoring section of the Model card is presented below, highlighting the role of the project manager in the development of this tool, to be completed by the technical team. This profile is detailed in González, Ortiz and Sánchez Ávalos (2020).

MODEL CARD		
Deployment and monitoring		
8. Monitoring recommendations	Monitoring and improvement strategy in production	Define, together with the technical team, a pilot program or randomized trial before deploying the model for the entire population.
	Prediction human monitoring strategies (if applicable)	Periodically monitor the tool together with the technical team and with the officials who interact with it.

Project manager role

Teoría

2.4 Accountability

Project manager role

In the accountability phase, the project manager will be in charge of:

- creating permanent and up-to-date communication mechanisms on the functioning of the AI system, for which it must use clear language;

- rendering periodic accounts on the monitoring and impacts of the system; and

- establishing an appropriate response system to manage individual requests for results from the application of the AI system.

It should be emphasized that the basic principles of public administration include transparency and the right of access to information by citizens, which also cover governmental artificial intelligence projects.

Interpretability and explanation of forecasts

Interpretability or explainability refers to the ability of the general citizenry to understand how an AI tool works (Miller, 2019). The benefits of this include the following:

- knowing how the problem is going to be solved;

- obtaining social license;

- detecting biases in the algorithms, and

- model debugging and improvement.

It is the responsibility of the project manager to create permanent and regular communication mechanisms on the operation of the AI tool (problems, impacts and adjustments over time) for all stakeholders.

The use of automatic decision-making tools can be strange and confusing for citizens, which can generate resistance to their implementation, especially since algorithms are often thought of as "black boxes". It is worth reiterating here that the algorithmic explainability and transparency of a model, as well as the information that can be provided about its operation, will depend on its level of opacity (Burrell, 2016 and Buenadicha et al., 2020), as discussed in Section 1.7 on ethical, legal and governance considerations.

Potential biases and debugging
It will be the responsibility of the project manager to provide stakeholders with information on the limitations and biases of the tool.

It has already been pointed out how the model and/or the data may be biased, which requires the project manager to implement measures to ensure fairness, in case biases have not been addressed at pre-implementation stages.

Likewise, in order to obtain and maintain the social license, it is important that the project manager pays particular attention to the risks and ethical considerations arising from the implementation of the model, discloses them, and clearly indicates how he/she intends to mitigate them.

Explanation of individual predictions
Here the project manager must develop response protocols for those particular cases in which people request additional information on the operation and application of an AI system. For example, if a person does not benefit from a government bonus, they should be able to request information about the reason for the automatic decision. This will require that there are people in charge of reviewing particular cases and giving a satisfactory response to the concerns of those affected.

For its part, the technical team will have to evaluate the best way to explain these individual predictions.

Traceability
It is considered of utmost importance that the project manager oversees the detailed documentation of all decisions made throughout the different stages of tool execution and that such information is available to all interested parties, regardless of the level of opacity of the information.
The objective of decision traceability is to mitigate the risks identified here.

The accountability section of the Model card is presented below, highlighting the role of the project manager in the development of this tool, to be completed by the technical team. This profile is detailed in González, Ortiz and Sánchez Ávalos (2020).

MODEL CARD	
Accountability	
9. (Optional) Explainability of predictions	Strategy to explain particular predictions (if necessary)
	Strategy to understand the importance of different attributes
10. Other ethical considerations, recommendations and warnings	

Project manager role

Design a response or review policy to address possible requests from citizens regarding the result of the tool.

Discuss with the legal team the risks that may arise in the implementation of the model (protection of personal data, biases, or others) and develop mitigation and explainability measures, if required.

WHAT HAPPENED TO DART AT THE EXECUTION STAGE?

Data collection and processing

Initially, the data used for the model included 550 fundus images from the Cordillera Oriente Health Reference Center (CRS – Centro de Referencia de Salud) of the Peñalolén municipality, in Santiago de Chile.

The technical team had to preprocess the images obtained, compressing them to standardize their size in light of previous international work in the same field.

At the time of developing the model, it was found that, in those images where t he contrast was insufficient, the prediction of the tool was affected, since it was more difficult to perform the segmentation of blood vessels in the retina, a key first step in the detection of diabetic retinopathy (Arenas, 2012).

The model continued to be retrained as more fundus examinations became available and the latest AI techniques were incorporated. In an observational study published in 2021, 1,123 fundus examination images were used (Arenas et al., 2021).

Model building and validation

In the initial stage of development, data was divided into two equal parts (275 images each) to create the training and test sets.

Expert knowledge of ophthalmologists from the University of Chile was used to label the ocular lesions present in the images. Detection of diabetic retinopathy has four stages: segmentation of blood vessels in the retina, localization of the optic disc, detection of bright lesions, and detection of red lesions.

When analyzing the model costs, it was established that the initial investment implied an extra cost for the Ministry of Health due to the acquisition of the license to use the AI model developed by DART. However, the benefit began to materialize when moving from a corrective action to a preventive one, with the effect this has on the health of people suffering from diabetes. In a 10-year horizon, it is estimated that the monetary resources allocated to the treatment of diabetic retinopathy would be reduced by approximately 50% (Shokiche, 2013).

Deployment and monitoring

Through Laboratorio de Gobierno de Chile26 a collaboration was achieved with the municipality of Recoleta, the Institute of Public Health and the Ministry of Health to carry out a pilot in that town during 2016.

To use the screening tool, a web interface was created that allows ophthalmologists to interact with their results. On this platform, the patient's personal data is entered, along with the fundus

26 Laboratorio de Gobierno is the Chilean State agency that since 2015 has been dedicated to co-creating solutions for priority public problems and building capacities to innovate in public institutions, in order to improve the services provided to citizens.

examination. The exams are then sent for analysis. The automatic system studies the images and returns them, along with the results, to the platform. Finally, ophthalmologists enter the platform and obtain information on probable and discarded cases, so that they can focus their efforts and time on the probable cases to give a definitive diagnosis (Arenas et al., 2015).

Use of the tool required training of health center staff, medical technologists who take fundus examinations and who must upload the images to the cloud-based tool, and ophthalmologists at the health centers who receive the screening results. Since 2018, synchronous video conferences have been held for user groups throughout Chile. Interactive tutorials are currently being worked on.

The Chilean Ministry of Health adopted the technology as a standard in its health centers; DART is currently used in more than 130 of them.

There is an ongoing relationship between the health centers and the Ministry of Health authorities, which as of 2021 was conducting surveys to establish the level of beneficiary satisfaction. With respect to the functioning of the tool itself, an observational study was conducted to evaluate the fit metrics. This was done based on a sample of 1,123 images from five primary care centers in Chile. Thirteen cases of false negatives were detected, while the area under the ROC curve (AUC) was 0.915. The performance measures exceeded what was required by the Chilean Ministry of Health (Arenas et al., 2021).

Accountability

DART started as the thesis project of José Tomás Arenas, current CEO of Teledx, so all the information about the data used in the first stage (before its implementation) and initial modeling are in the repository of the University of Chile.

After its implementation in the Chilean health system, multiple papers have been published in international scientific journals and it has participated in international talks, where significant portions of information on data, processes and even the algorithm's operation have been shared. Although the model underlying the tool is extremely complex, the value of its application in different health centers has been demonstrated by reducing the waiting time for test results and the monetary costs associated with the treatment of the disease. However, the possible discriminations of the model should be studied in detail, which could originate, for example, in the conditions in which the fundus examination is performed, since they can affect its reading. This may be more prevalent in some locations than in others.

Since the system is currently implemented under the auspices of the Chilean Ministry of Health, its operation is subject to the transparency and personal data protection laws in force in that country.

CONCLUSIONS

This manual describes the phases necessary for the correct implementation of an AI project from the perspective of the person in charge of making public policy decisions and leading the project, considering the challenges that arise in the AI life cycle.

An AI-based decision-making and/or decision support project is closely linked to the public policy cycle, insofar as it must solve real and significant problems affecting society.

Projects often start with an interest in using data available in the institution or a new technology. However, it cannot be overemphasized that for a project to have social value, it must arise from the need to solve a real public policy problem.

While the AI tool will be developed by the technical team, it will be up to the project manager to make the decisions and lead its planning and design, to communicate with the technical staff, conduct the feasibility analysis of the tool and its final implementation.

It is important to reiterate that the creation of a support and/or decision-making system should be the product of an iterative process. It is common for teams to have to revisit the stages mentioned here, even redefining the planning and design of the project as new challenges arise. This stage is critical because it forces the project manager to ask key questions that maximize the project's chances of success. Once feasibility is defined, the execution stage begins with data processing. Figure 6 contains the tools provided in this manual for the project manager, according to the AI life cycle.

Figure 6. AI life cycle and manual tools
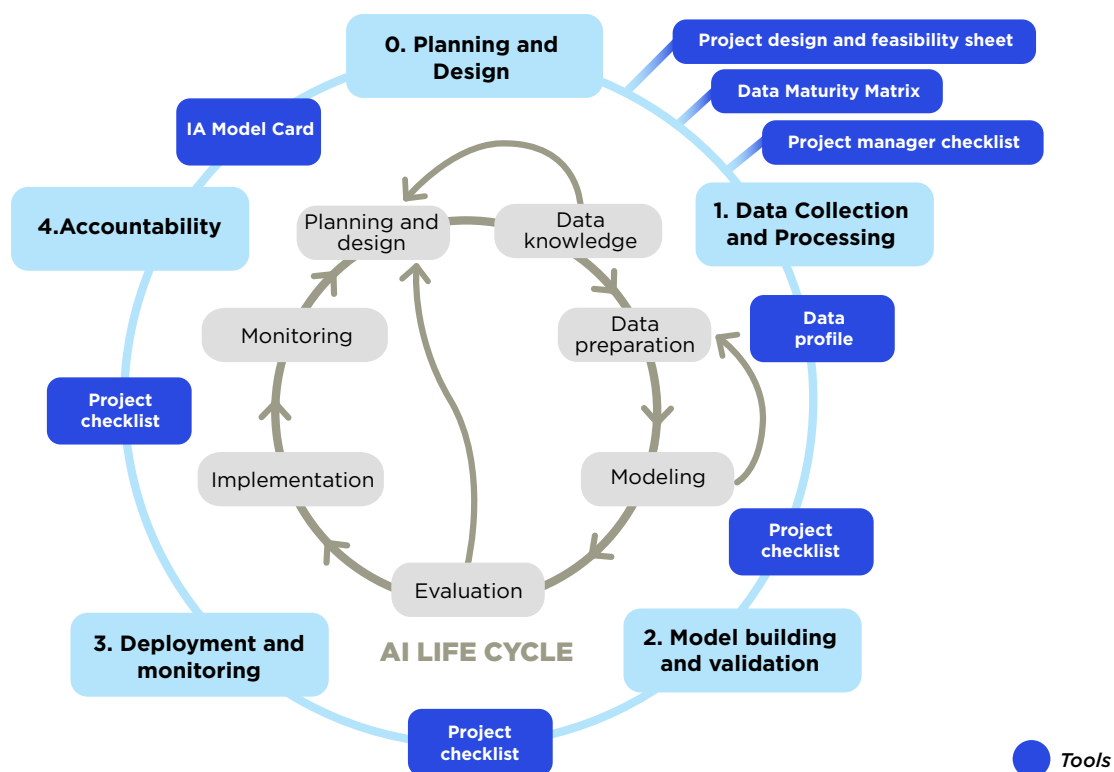
From the outset, the project manager should be clear about the potential risks of the AI tool in terms of social license implementation and ethical issues. He/she should also ensure that he/she has the human and/or financial resources to maintain and update it on a regular basis. The tool will be constantly being adjusted as the context changes or as new data becomes available. It is also necessary that staff members understand its benefits, adopt it and incorporate it into their work for the proposed purposes.

This manual presents the AI life cycle within the life cycle of public policy, which does not end at the end of the policy cycle. Therefore, it is necessary to evaluate the implementation of the policy, not only from the point of view of the tool but also of its effectiveness in terms of solving the problem identified in society.

We hope that this manual will be useful to decision-makers who are thinking of implementing an AI-based system. This manual should be used in conjunction with the responsible AI technical manual (González, Ortiz and Sánchez Ávalos, 2020), where the tasks and challenges of the technical team are specified.

REFERENCES

Arenas Cavalli, J. T. A. (2012). Diseño y desarrollo de un sistema para la detección automática de retinopatía diabética en imágenes digitales. Available at http://repositorio.uchile.cl/handle/2250/104406

Arenas-Cavalli, J. T., Abarca, I., Rojas-Contreras, M., Bernuy, F. and Donoso, R. (2021). Clinical validation of an artificial intelligence-based diabetic retinopathy screening tool for a national health system. *Eye*, 1-8.

Arenas-Cavalli, J. T., Ríos, S. A., Pola, M. and Donoso, R. (2015). A web-based platform for automated diabetic retinopathy screening. *Procedia Computer Science*, 60, 557-563.

Bojalil, P. Vela-Treviño, C. (2019). Despuntan las reformas en materia de protección de datos en América Latina. Conocimiento abierto. Available at https://blogs.iadb.org/conocimiento-abierto/es/proteccion-de-datos-gdpr-america-latina/. Retrieved March 29, 2021.

Buenadicha, C., Galdon, G., Hermosilla, M. P., Loewe, D. and Pombo, C. (2019). *La gestión ética de los datos. Por qué importa y cómo hacer un uso justo de los datos en un mundo digital*. Available at https://publications.iadb.org/publications/spanish/document/La_Gesti%C3%B3n_%C3%89tica_de_los_Datos.pdf.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 2053951715622512.

Cabrol, M., González, N., Pombo, C. and Sánchez, R. (2020). *Adopción ética y responsable de la Inteligencia Artificial en América Latina y el Caribe.* Available at https://publications.iadb.org/publications/spanish/document/fAIr_LAC_Adopci%C3%B3n_%C3%A9tica_y_responsable_de_la_inteligencia_artificial_en_Am%C3%A9rica_Latina_y_el_Caribe_es.pdf

Dastin, J. (2018). Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women. Available at https://www.reuters.com/article/us-amazon-com-jobs-automationinsight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-womenidUSKCN1MK08G Retrieved January 19, 2021.

Data Futures Partnership (2017). *A Path to Social Licence: Guidelines for Trusted Data Use.*

Gebru, T., Morgenstern, J., Vecchione, B., Wortman, J., Wallach, H., Daumé, H., & Crawford, K. (2018). *Datasheets for Datasets*. Obtained from https://arxiv.org/pdf/1803.09010.pdf

Giorgi, S. (2017). How to improve the evaluation of complex systems to better inform policymaking Learning from evaluating Defra's Reward & Recognition Fund. Fellowship Report, July. Available at https://www.cecan.ac.uk/sites/default/files/2018-01/Guidance%20Report%20 %20RRF%20Fellowship%20Final.pdf

Gómez, C., del Pozo, C., Martínez, C. and Martín del Campo, A. (2020). *La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países*. Available at https://publications.iadb.org/publications/spanish/document/La-inteligencia-artificial-al-servicio-del-bien-social-en-America-Latina-y-el-Caribe-Panor%C3%A1mica-regional-e-instant%C3%A1neas-de-doce-paises.pdf

González, F., Ortiz, T. and Sánchez Ávalos, R. (2020). *IA Responsable: Manual técnico: Ciclo de vida de la inteligencia artificial*. Available at https://publications.iadb.org/publications/spanish/document/IA-Responsable-Manual-tecnico-Ciclo-de-vida-de-la-inteligencia-artificial.pdf

Greene, W. H. (2003). *Econometric Analysis*. India: Pearson Education.

Harini Suresh, J. V. (2019). *A Framework for Understanding Unintended Consequences of Machine Learning*. MIT. Obtained from https://arxiv.org/pdf/1901.10002.pdf

Heaven, W.D. (2020). Google's medical AI was super accurate in a lab. Real life was a different story. Available at https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/ Retrieved January 29, 2021.

Hewitt, T. (2019). *Canada is a leader in AI research – but that's not enough*. Available at https://www.theglobeandmail.com/opinion/article-canada-is-a-leader-in-ai-and-we-must-lead-in-research-on-how-we-want/ Retrieved January 19, 2021.

Hojman Cano, I. (2014). El mercado de las especialidades médicas de anestesiología y oftalmología en Chile. Available at http://repositorio.uchile.cl/handle/2250/130678

Miller, T. (2019). Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artif. Intell*., 267, págs. 1-38.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., . . . Gebru, T. (2019). *Model Cards for Model Reporting*. Obtained from https://arxiv.org/abs/1810.03993

Morrison, J. (2014). *The Social License. How to keep your organization legitimate*. Palgrave Macmillan, ISBN: 9781137370716.

OECD. (2019). *Artificial Intelligence in Society*. Paris: OECD Publishing.

OECD (forthcoming). *OECD Framework for the Classification of AI Systems*. Paris: OECD Publishing.

Oliver, D. (1994). Law, politics and public accountability. The search for a new equilibrium. *Public Law*, 238-238.

Red Iberoamericana de Protección de datos personales (2017). Estándares de protección de datos personales. Available at https://www.redipd.org/es/documentos/estandares-iberoamericanos

Sangüesa, R. (2018). Inteligencia artificial y transparencia algorítmica: "It's complicated". *BiD: textos universitarios de biblioteconomía y documentación*, No. 41, December. <http://bid.ub.edu/es/41/sanguesa.htm>. Available at http://bid.ub.edu/es/41/sanguesa.htm#:~:text=La%20transparencia%20de%20datos%20y,vital%20de%20quien%20reclama%20esta. Accessed on January 29, 2021.

Sapunar, J. (2016). Epidemiología de la diabetes mellitus en Chile. *Revista Médica Clínica Las Condes*, 27(2), 146-151.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin. ISBN: 0395615569. Available at https://www.alnap.org/system/files/content/resource/files/main/147.pdf

Shokiche Vega, D. A. (2013). Estudio del problema de salud pública asociado a la patología oftalmológica retinopatía diabética en Chile y dimensionamiento del potencial impacto de detección basado en tecnología para abordar este problema.

Varian, Hal R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28 (2): 3-28.

ANNEXES

Annex 1. Project Design and Feasibility Sheet

Annex 2. Data Maturity Matrix

Annex 3. Project Manager Checklist

Annex 4. Data profile

Annex 5. Model card

Annex 1. Project Design and Feasibility Sheet

1	NProject name:

2	Organization name:

3	Problem definition

What is the problem to be solved?

Describe the population(s) affected. Who or what is affected by this problem? (Certain types of people, organizations, neighborhoods, environment).

How many people/organizations/locations/etc. are affected and to what extent? (For example, average waiting time for surgery, number of students dropping out of school, costs of tax evasion, etc.).

Why is solving this problem a priority for your institution?

Do you know of any similar AI use cases that have been implemented before? Which ones?

4	Pre-feasibility analysis

Is it within our power to act on the problem? Will we have to partner with other public agencies? Do we have the necessary human and financial resources to carry out the project?

Is the relevant data available (enough to be able to change the current way of responding to the problem)? Is it accessible?

What are the risks of the project (ethical, social license, implementation, etc.)?

5   Definition of objectives

Objectives are usually expressed in terms of improving, maximizing, increasing or decreasing, mitigating, and/or reducing an outcome. The objective must be measurable, which requires establishing a metric or indicator that reflects progress. Achieving the objective should help solve the problem. The technical solution (e.g., a predictive model) is not the objective.

Typical limitations relate to budget, lack of human capital, legal constraints, political will and social license.

Keep in mind that when there are competing objective, you may have to sacrifice something to gain a benefit.

	Objective	Limitations
1		
2		
3		

6   Description of actions

Actions are those activities that institutions carry out or can carry out in relation to a given problem, such as the specific programs they carry out according to their mission to society or their usual operating processes (hiring, user services, payment of salaries, etc.).

These actions can be improved when the institution has the information generated by the data science project. They should also have a connection with the results generated by the AI system and help achieve the objectives set (previous section).

	Complete		
	Action 1	**Action 2**	**Action 3**
Action E.g., responsible sexuality workshop for 13-year-old students/Contraceptive method delivery in the school infirmary. Note: Each action should be explained in a separate box.			
Who executes the action? E.g., psychology and psycho-pedagogy team of each establishment and/or school nursing team.			
On whom or what is the action being taken? E.g., students in schools currently in 7th grade and/or the general student body.			
How often is the decision made to perform this action? E.g., annually/monthly.			
What channels are being used or can be used to perform this action? E.g., face-to-face.			
Other useful information about the action			

7   Data mapping

If the institution is to achieve its objective, the data has to be connected to the actions it will support. Typical AI projects use administrative data as a primary source and enhance it with other data sources in the public domain (censuses, other open data, etc.). Partnering with the private sector or non-profit organizations can help obtain missing data internally.

What data is available internally?

	Complete		
	Data source 1	Data source 2	Data source 3
Name E.g., hospital discharge system.			
What does it contain? Describe attributes in as much detail as possible (e.g., hospital admission and discharge records nationwide, with patient sociodemographic data, diagnosis, days in hospital, type of health insurance, doctor's information).			
What level of granularity? E.g., transaction, person, organization, location.			
How often is information collected and/or updated once it is captured? E.g., real-time, daily, weekly, monthly, yearly, occasionally.			
Do you have unique and reliable identifiers that can be linked to other data sources? E.g., RUN, SSN, DNI, depending on the country.			
Who is in charge of the data? E.g., the hospital records department			
How is it stored? E.g., in a database, PDF, Excel, SPSS.			
Additional comments			

What data can you obtain from external private or public sources?

	Complete		
	Data source 1	**Data source 2**	**Data source 3**
Name E.g., air quality record.			
What does it contain? Describe the attributes in as much detail as possible (e.g. concentration of pollutants—such as particulate matter of different sizes—in the air).			
What level of granularity? E.g., hourly geolocated monitoring station.			
How often is information collected and/or updated once it is captured? E.g., daily.			
Do you have unique and reliable identifiers that can be linked to other data sources? E.g., monitoring station code.			
Who is responsible for the data? Ministry of Environment			
Are legal agreements required for the exchange and/or access to the information?			
How is it stored? E.g., downloadable database via an API on an open data portal.			
Additional comments			

In an ideal world, is there additional data relevant to this problem that you would like to obtain (surveys, CCTV, phone records, DNA, range of frequency or granularity for currently available data, etc.)?

8   Analysis/Tool

Typical AI projects include a combination of several analyses, depending on the needs and particularities of each project. Analyses are tools, not the goal of the project.

Choose the right analyses for the right problem.

- The analyses or tools chosen should improve current actions in response to the problem.
- Analyses should be tested, and the validation process must match the objective.

	Complete		
	Analysis/Tool 1	**Analysis/Tool 2**	**Analysis/Tool 3**
Type of analysis/tool E.g., description, prediction, detection, behavioral change			
Purpose of analysis E.g., to understand the historical behavior of individuals; to estimate a patient's risk of disease; to identify actions that would reduce overfishing.			
For what type of actions will the information generated from this analysis be used? E.g., inspection of industrial and artisanal fishing vessels.			
How will you validate this analysis using existing data? E.g., using historical data, conducting a randomized controlled trial, etc.			

9   Ethical and legal considerations

Proportionality	**Do you think a data science/ AI system is the right way to solve the problem? Why? Have you evaluated other alternatives?**	
	What negative impacts might your project have? Review similar use cases identified in the section "Problem Definition".	
Social license	**Do you think the users or affected people will find the proposed use of data to solve the problem acceptable? Why?**	
	If the project's target population learns about it, will they approve? Why?	
	Have you identified the justification or legal basis for working with these data?	
	Have you identified the regulations that could affect the project?	
	Will it be necessary to have mechanisms to guarantee the quality of personal data, such as access, deletion or rectification mechanisms?	
Transparency	**Which stakeholders should be aware of the project?** Stakeholders usually include policy makers, front-line workers, civil society organizations, government agencies, people who will be affected by the actions, etc. Please list specific organizations and/or types of people.	
	Have you considered any mechanisms for stakeholders to communicate with the institution about the project?	
	Will it be necessary to explain the decision-making mechanisms or the analysis to be carried out? Why?	

Discrimination/ equity	**What structural inequalities are there in the process and/or in the environment where the project is inserted?**	
	Are there specific (vulnerable) groups for which equity of outcomes or protection of rights is to be ensured? E.g., groups by gender, age, location, social class, educational level, origin (urban or rural), ethnicity?	
	What biases do you think the data might have?	
Accountability	**Who is responsible for providing information about the project if requested?**	
	Who is responsible if the system is wrong?	
	Are monitoring, control and evaluation mechanisms in place? How will they be documented and how often?	
	Are training mechanisms in place to ensure that the team in charge understands the responsibilities, as well as the legal and ethical obligations of the project?	

10   Team composition

Generally, artificial intelligence projects require the involvement of various professionals from the same public entity, but also from other related organizations. This includes those responsible for the data, the IT infrastructure and the problem or process in question, as well as experts in data analytics, legal and communications. Add as many lines as required in the following table.

Organization/Department	Description of desired participation	Counterparty name/role
IT Department	Provide data infrastructure	Head of the IT Department
Statistical agency	Provide population data	Head of the Statistics Department

Annex 2. Data Maturity Matrix

Data Maturity Matrix[27]

Category	Areas	Deficient	Basic	Intermediate	Advanced
How information is stored	**Access**	Data is only accessed through the application that collects the data.	Data can be accessed, but with special and specific software.	Data is in accessible formats such as CSV, JSON, XML or a remotely accessible database.	Data is in accessible formats and can be accessed through an API.
	Storage	Paper	PDF or images	Text files	Databases
	Integration	Data is only in the system where the information is collected.	Data is occasionally exported and integrated on an ad-hoc basis.	Single data center with automatic updating.	Both internal and external data are integrated into the database.
What information is collected	**Relevance**	Data is irrelevant to the problem. For example, you need to know the probability of finishing college but you do not have the data about who graduates.	Some of the data is relevant, but it is insufficient because key components are missing. For example, going back to the previous case, only class attendance is available.	There is useful and relevant information, but it is not complete. For example, in the previous case, there is demographic data and annual grades, but no extracurricular support information.	All relevant information exists and is sufficient to solve the problem without making relevant transformations.
	Quality	Missing rows (remarks)	Missing columns (key variables of certain observations).	All the data is there; there are only minor errors, such as typing errors.	There are no problems of missing data or typing errors; the databases are clean.
	Frequency	Just once	Annual	Frequent	Real-time
	Granularity	Added at a city level	At the block or ZIP code level	At the individual level (person or address)	Event/milestone level detail
	History	There is no history; old data is deleted.	Historical data is saved but updates are overwritten.	Historical information is stored occupying a time stamp.	All data is maintained and related to previous data with an integrated model.
Others	**Privacy**	There is no privacy policy.	The policy does not allow the use of any data.	Ad-hoc approval is needed for data use.	Access to data is defined and data privacy is controlled to preserve the privacy of individuals.
	Documentation	There is no digital documentation or metadata. The variable codes are not documented.	A data dictionary explaining the variables and categories is defined.	There is a data dictionary and metadata availability.	There is a data dictionary, metadata, and additionally the assumptions, biases and data that are not being obtained.

27 Attribution ShareAlike (CC BY-SA). Adapted from the Data Maturity Framework of the University of Chicago http://dsapp.uchicago.edu/resources/datamaturity/

Category	Areas	Deficient	Basic	Intermediate	Advanced
How information is stored	**Access**				
	Storage				
	Integration				
What information is collected	**Relevance**				
	Quality				
	Frequency				
	Granularity				
	History				
Others	**Privacy**				
	Documentation				

Annex 3. Project Manager Checklist

The following checklist contains the main tasks and roles of the project manager (decision-maker) throughout the AI lifecycle.

Planning and design of public policies

- [ ] Clearly define the public policy problem to be solved, identifying and quantifying the groups of people who are affected, and determining its impact on the budget. The definition of the problem should be easily understood by an outsider.

- [ ] Contact the people in your institution who are in charge of addressing the problem to establish how it is currently done. What information can they provide on how to improve the response system?

- [ ] Investigate how other agencies—domestic or foreign—with a similar problem have implemented an AI-based solution. Ideally, contact them to learn about the challenges and difficulties they encountered along the way.

- [ ] Discuss with your institution's senior management the priority of solving the problem and gain commitment to the project at the highest level. If you are a senior manager yourself, document how this priority is embodied in the institution's strategic plans.

- [ ] Confirm that the technical team checklist for this phase has been completed.

Lifecycle

Data collection and processing

- [ ] Contextualize the findings of the technical team in the exploratory stage of the data.

- [ ] Provide background on the data collection process and the meaning of the variables, connecting the technical team with the people in the organization required for a better understanding of the data.

- [ ] Validate the formulation and usefulness of the target variable.

- [ ] Make decisions on data imputations or elimination of variables, weighting and recording what is gained versus what is lost (trade-offs).

- [ ] Propose improvements to project data governance processes that can contribute to the achievement of objectives and facilitate model implementation.

- [ ] Make adjustments to the problem definition, objectives and project implementation plan based on the results of this stage.

- [ ] Confirm that the technical team checklist for this phase has been completed

Model building and validation

☐ Evaluate the model with metrics that not only consider efficiency and effectiveness, but also equity.

☐ Decide on the model to implement taking into account the human and financial resources available in the institution.

☐ Understand the limitations of the model and adjust the project implementation plan according to the results of this stage.

☐ Confirm that the technical team checklist for this phase has been completed.

Deployment and monitoring

☐ Ensure the training of public servants who interact with the AI model so that the tool is sustainable over time.

☐ Develop a user manual focused on the public servants who will interact with the model.

☐ Establish feedback mechanisms for people interacting with the model.

☐ Design and implement simple administrative processes to correct those errors present in the model that affect users.

☐ Establish automated, or at least periodic, model monitoring systems.

☐ Maintain a record of model results, taking into account access and security restrictions.

☐ Implement necessary improvements to the model and process from monitoring and evaluation findings.

☐ Allocate the necessary resources to maintain the tool over time.

☐ Confirm that the technical team checklist for this phase has been completed.

Accountability

☐ Create permanent and updated communication mechanisms on the functioning of the AI system, using clear language.

☐ Provide periodic accountability for the monitoring and impacts of the system.

☐ Establish an appropriate response system to handle individual requests for results from the application of the AI system.

☐ Confirm that the technical team checklist for this phase has been completed.

Annex 4. Data profile

Data profile is an exploratory analysis that provides information to assess the quality, completeness, timeliness, consistency, and potential biases of the dataset that will be used to train a machine learning model (Gebru et al., 2018). In order for the technical team to generate an appropriate data profile, the project manager will need to ensure access to the datasets (when internal) or coordinate their procurement with external agencies, along with documentation on each variable within them. He/she may also assist in identifying important dimensions for which the data sample may differ from the general population, as well as in identifying undesirable states that may include biases, inequalities detrimental to certain subgroups and/or any other patterns that are considered suboptimal or undesirable from a social policy standpoint. Finally, it can help identify the reasons for missing information and decide, together with the technical team, how to fill this gap.

Overview of data and motivation
Name of the dataset used
Which institution created the database?
For what purpose?
What mechanisms or procedures were used to collect the data (e.g. household survey, sensor, software, API)?
Sample size
Frequency of capture (weekly, monthly, daily) or average number of observations per individual. Will the dataset be updated (e.g., adding new instances or deleting others)?

Rol del director del proyecto

Connect the technical team with the people in charge of the data (both internal and external) to obtain detailed information about the bases, the way of collection and possible weaknesses.

Discuss with the legal team the restrictions on the use of the data, given the way it was captured. For example: Is it necessary to obtain informed consent? Is it public or private information?

Discuss, together with the technical team, whether the sample is considered representative. If it is not, how will the training of the tool be affected?

Project manager role

Essential controls
Obtain documentation for each variable within the dataset. Brief description including its name and type, what it represents, how its value is measured, etc.
Perform an exploratory analysis of the data by calculating descriptive statistics, identifying the percentage of missing values, and the distribution of each variable within the database.
Analyze the spatial and temporal coverage of the data. ESTE NO TIENE
Analyze coverage of protected groups (sex, race, age, etc.).
Describe important dimensions in which the data sample may differ from the population, and in particular unmeasured selection biases. Rely on specialized literature and expert information.
Identify possible "undesirable states" in the data, whether biases or inequities detrimental to subgroups or any other patterns that are considered suboptimal or undesirable from a social policy standpoint.
Are there any missing values? If yes, explain the reasons for not having such information (This includes intentionally deleted information). Identify reasons why there is missing data and think about whether the miss-ing data is associated with the variable to be predicted.

Analyze the results of the exploratory data analysis. Validate that the maturity of the database is adequate to solve the selected public policy problem.

Communicate to the technical team which groups will be considered protected.

Explain to the technical team, based on expert knowledge, the differences found.

Explain, based on expert knowledge, the possible biases in the data that may arise from an undesired state in the collection. Define what is considered suboptimal status. Discuss bias mitigation measures in protected groups with the technical team.

Explain, based on expert knowledge, missing values (e.g., an adverse circumstance that did not allow correct data collection).

Annex 5. Model card

The model card is a tool that summarizes the main characteristics of a machine learning-based decision-making and/or decision support system, while highlighting its main assumptions and features, as well as the mitigation measures implemented (Mitchell et al., 2019). The project manager will be able to corroborate that the planning and design of the public policy, as well as the use cases, correspond to the problem being addressed, for which the background, the target population, the time horizon, and the actors and systems involved should be verified. The project manager should also ensure that the use cases conform to the situations where the decision support system is expected to be used, including taking into account non-considered uses and related warnings. Finally, the project manager should validate the protected groups and potential biases. The model card should be completed jointly by the project manager and the technical team.

Project manager role

Public Policy Planning and Design		He/she is in charge of the entire planning and design stage of the project.
1. Basic Information	People who developed the model, date, version, type	
2. Use Cases	Background	Correctly define the public policy problem that the institution seeks to solve.
	Target population and time horizon of forecasts	Correctly identify the affected population. Concisely and clearly define the objectives to be achieved to solve the described problem..
	Stakeholders and components that will interact with the results	Contact current responders to determine how they have proceeded and how improvements could be made with an AI tool. Take into account the particularities of the action that could jeopardize the implementation of the tool.
	Use cases considered during development	Together with the technical team, identify those project managers who have already solved—or are trying to solve—similar problems with AI and contact them to become familiar with their experiences.
	Non-considered uses and related warnings	Together with the technical team, identify AI tool misuse scenarios, if any, and design mitigation measures to prevent their occurrence.
	Definition of protected groups	Define, based on expert knowledge, which groups will be considered protected and which groups are to be guaranteed equity in the results.

Data collection and processing		
3. Training Data	Data set used and its labeling	
	Preprocessing steps or data preparation.	Review, together with the technical team, the results of the exploratory data analysis; explain the anomalies found (if any) and the results of the analysis based on expert knowledge, and contextualize the main findings.
	Potential biases and shortcomings depending on the use case (see point 2 here)	Request the technical team to conduct an exploratory analysis of the data by population subgroups. Adopt bias mitigation and algorithmic equity measures.
Model building and validation		
4. Modeling	Algorithms used for training, assumed parameters or constraints	
5. Performance Metrics	Technical metrics used to select and evaluate models	
	Cost-benefit analysis of the model for the use case (see point 2 here)	Perform cost-benefit analysis of the implementation of the tool and determine whether the public policy problem can be solved without using AI.
	Definition of protected groups and selected equity measures	Define equity measures in the results (population subgroups to which equity is to be guaranteed).
6. Validation Data	Data sets used and their labeling	
	Preprocessing steps	
	Evaluation of validation data adaptation according to the use case (see point 2 here)	
	Potential biases and shortcomings depending on the use case (see point 2 here)	Establish acceptable bias thresholds together with the technical team, based on expert knowledge. Prepare mitigation measures based on public policy, if required.

7. Quantitative Analysis Summary	Reported validation error	
	Cost-benefit analysis summary	
	Report on equity measures for protected groups	Decide, together with the technical team, which model will be implemented.
Deployment and monitoring		
8. Monitoring Recommendations	Production monitoring and improvement strategy	Define, together with the technical team, a pilot program or randomized trial before implementing the model in the entire population.
	Human prediction monitoring strategies (if applicable)	Schedule periodic monitoring of the tool (which will depend on the nature of the tool) with the technical team and also with the staff that interacts with it.
Accountability		
9. (Optional) Explainability of Predictions	Strategy to explain particular predictions (if necessary)	Formulate a policy for responding to or reviewing possible requests from the public regarding the results of the tool.
	Strategy to understand the importance of different attributes	
10. Other Ethical Considerations, Recommendations and Warnings		Discuss with the legal team the risks that may arise during the implementation of the model (personal data protection, biases or others), and develop mitigation and explainability measures, if required.