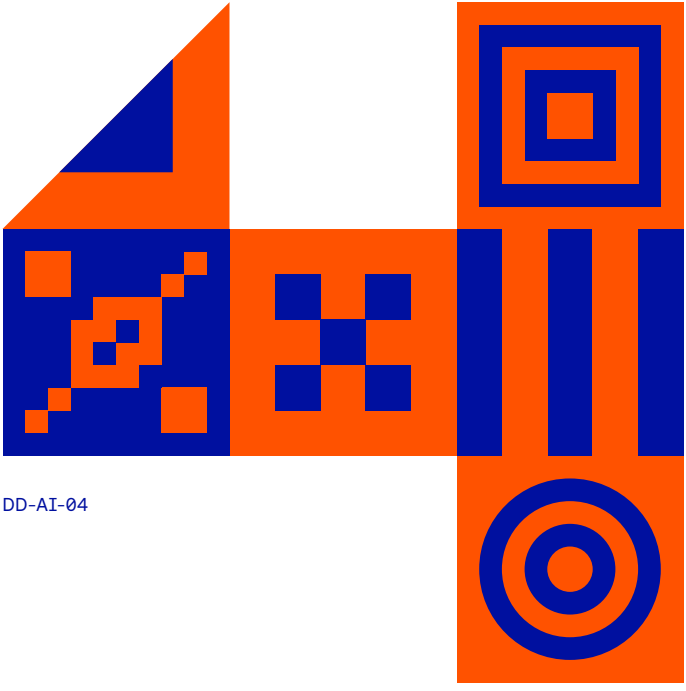
How to Implement an Automated Decision-making System

# Questions for people dealing with machines that deal with people



DD-AI-04

**How to Implement an Automated Decision-making System:**
**Questions for people dealing with machines that deal with people**

This publication was prepared by the Derechos Digitales team, under the "Artificial Intelligence and Inclusion" project, coordinated by Jamila Venturini, Juan Carlos Lara and Patricio Velasco. This edition has been reviewed, updated and translated.

For more information on this project, visit https://ia.derechosdigitales.org

Text by Jamila Venturini, J. Carlos Lara, Patricio Velasco and Vladimir Garay.
Edition and proofreading by Vladimir Garay.
Design and layout by Comunicas Unidas.
Translation and adaptation by Urgas Tradu.c.toras.

DERECHOS DIGITALES
América Latina

Inteligencia artificial e inclusión en América Latina

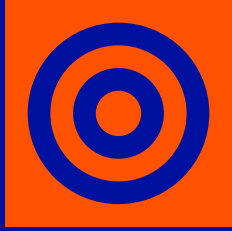
This is a set of questions and answers. The goal is to propose a series of key questions and considerations for deciding on the implementation of an artificial intelligence system to address social problems, under a public policy.

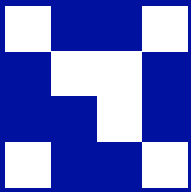
Each question is posed based on assumptions about the assessment, design, implementation or evaluation stage of public policies that incorporate artificial intelligence. It is not the goal of this tool to certify that the public policy is appropriate, necessary or proportional; rather, it simply attempts to aid in reflecting on these concepts and to support a critical, respectful approach to fundamental rights when integrating technology for public administration.

# DD-AI-04



# Is the problem you wish to solve defined?





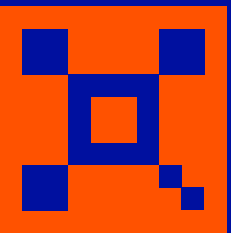
## If you answered “no”

Then it is still not time to seek technological solutions. Before proceeding, it is essential to create participatory and multidisciplinary spaces, as well as to reflect on the need to implement an automated system and on existing alternatives. Remember that implementing this kind of technology requires mobilizing significant resources, both monetary and in terms of infrastructure, personnel and management, but their implementation can involve less transparency in public administration and affect the exercise of basic rights.

## If you answered “yes”

Make sure as well that the definition of the problem is well supported, evidence-based and responds as much as possible to specific needs of the affected population. The definition of the problem must not be linked to a specific technological solution proposal, but rather be open to different possible answers. The technological aspect, when considered, is only one component of the public policy. Technology is not the solution to everything, and its efficacy depends on a very careful diagnosis that makes it possible to anticipate its possibilities, limits and risks.

Is there technical, social and political consensus regarding the identified problem and the proposed solution?



## If you answered “no”

Then it is necessary to take a step back and review the definition of the problem. Remember that social problems are complex and have multiple implications; whether or not to adopt a technology cannot respond solely to technical criteria. The implementation of an automated system is only one part of a public policy response that must consider the potential causes and consequences of the problem to be addressed, as well as the positive and negative effects of the intervention to be implemented. This is essential for the legality and proportionality analyses that must be considered when deciding on the adoption of a given system.

## If you answered “yes”

Be sure as well that the technological solution proposed to respond to the problem meets basic human rights requirements. For this, analyze other cases that have implemented similar solutions. If the technology is provided by private suppliers, review their background to guarantee that the company is not involved in cases where the population's rights may have been affected. Processes for contracting technology providers, as in any area of government, must be transparent, and the service must meet very high security standards, especially if third-party processing of personal data is involved.

**Were there citizen engagement efforts in the identification, delimitation and prioritization of the problem?**

## If you answered “no”

Then you should implement some form of engagement confirming that the identified problem is perceived as such by the people to whom the solution is targeted. It is critical for affected groups (or their representatives) to have the opportunity to be heard and that their viewpoints are part of the problem assessment. There are different ways to do this, from open public consultations in collegiate bodies or with representative entities (like unions, confederations or associations of different types), to opinion surveys and qualitative investigations. Without citizens' opinions it is impossible to understand the problem in all its nuances, which compromises decision-making on the required response.

## If you answered “yes”

Also ensure that the engagement mechanisms for decision-making that affects basic rights are legally and institutionally guaranteed, and that there are mechanisms for transparency and accountability to show how the contributions obtained through the consultations were taken into account. Finally, citizen participation is also essential in decision-making and design processes for the proposed solution, as well as in evaluation.

Were other alternatives considered to solve the identified problem?

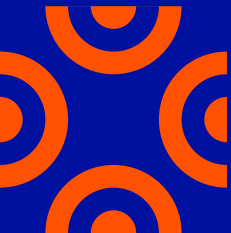
## If you answered “no”

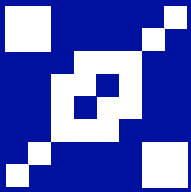
Then the intervention that you are thinking of developing does not meet the criteria of need and proportionality. The criterion of need prescribes confirming that the solution is indispensable and ideal among various options. The criterion of proportionality involves considering the balance between the different human rights that may be affected. It can be a difficult question to determine when privacy and personal data protection are involved, but it must be considered that the violation of these rights can have an impact on other fundamental rights, such as social and economic rights, potentially leading to an increase in inequality. For this reason, before implementing an automated system one must ask if this is the only alternative for responding to the identified problem and, where this is the case, whether the risks are justified from the perspective of public interest.

## If you answered “yes”

Make sure as well that the proposed intervention does not represent a disproportionate risk to human rights. Remember that automated decision-making systems and machine learning technologies can affect the right to privacy and other related rights. In addition, the risk of discrimination arising from the use of AI-based decisions is very high. In the case of remote biometric recognition technologies, international authorities have recommended that states refrain from implementing them until there are guarantees in place that they do not jeopardize basic rights.

Is the implementation of this technology the most effective and efficient solution to the problem posed?





## If you answered “no”

Or if you are not absolutely certain of the answer, then you may be a victim of "techno-solutionism". Decisions on the use of technologies, like any decision regarding public policies, must be based on specific, independent evidence and not on optimistic rhetoric or marketing tools. Any proposal for the implementation of a technological system must be subject to a comprehensive analysis by a multidisciplinary technical body that proves the system’s effectiveness for responding to the problem identified and its suitability as compared to other alternatives. This stage is also critical for addressing considerations on the proportionality of a solution.

## If you answered “yes”

Also ensure that appropriate indicators are used to monitor the implementation and to evaluate its effectiveness in light of the problem in question. Remember that technological systems and algorithmic models can degrade over time. Adequate monitoring and evaluation should provide warning of the need for interruption or review of the system. Make sure that the periodic analyses are available to all stakeholders, through either active or passive transparency actions and always respecting the privacy and protection of the data of the people affected.

# Did citizens participate in the design of the proposed system?

## If you answered “no”

Then you should think of ways to make possible some participation in the design of the proposed system. The incorporation of engagement mechanisms starting with the public policy design stage enables citizens to be part of the policy from the beginning, fostering their understanding and lending the initiative greater legitimacy.

## If you answered “yes”

Make sure as well that there really is broad social engagement, including citizen diversity. For example, ensure that there is participation from people who suffer from the digital gap, by using online and offline mechanisms for publishing the announcements and encouraging participation. On this topic, see our engagement recommendations in "Public Policies and Artificial Intelligence: Basic recommendations for the design of effective engagement efforts".

Are there mechanisms for system users and the target audience to provide feedback on its design and operation?

## If you answered “no”

You have a problem since it will not be possible to know if the system design and operation fully perform their functions. First, be sure that all the rules for public engagement are being respected. Second, consider that it is essential for people to have spaces for commenting, criticizing, asking questions or making suggestions about the system's operation. Remember that the people who use the system and those who are the system's target audience are those who are more directly affected by design or implementation problems. Addressing their feedback is critical to the success of the public policy.

## If you answered “yes”

Be sure that it is an accessible, easy-to-use option and that there are efforts to follow up on the feedback, for example, by contacting people who have left comments or complaints. Ensure the existence of a transparent procedure for intervening in the system based on that feedback. Be sure to generate active transparency mechanisms, reporting on the feedback received and on the measures adopted in response.

In addition to current legal obligations in the country, does the system's implementation follow recommendations and best practices provided by international bodies for the processing of personal data and sensitive data?

## If you answered “no”

Or if you are not sure, spend some time reviewing the available recommendations on personal data in terms of AI. There are specialist communities who have spent years formulating recommendations and directives that facilitate the development of digital technologies capable of protecting people's rights. These include the "Recomendaciones generales para el tratamiento de datos en la Inteligencia Artificial" [General recommendations for data processing in Artificial Intelligence] by the Red Iberoamericana de Protección de Datos and UNESCO's "Recommendation on the Ethics of Artificial Intelligence". Although often these recommendations can be more applicable to laws that are different from those of your country, they do serve as a guide for suitable processing of people's information.

## If you answered “yes”

You should pay attention to ensuring the recommendations used are known and reputable, as well as accessible to anyone. Be sure to be explicit in the information provided by the system that these recommendations are being followed, such that the people who interact with it have the possibility of observing the recommendation. Make sure as well that the information on how these recommendations are being followed is also available for the evaluations and external audits to which the system is subject.

**Are public officials adequately trained in implementing and operating the system?**

## If you answered “no”

It is likely that the technological system will not meet its objectives and, even worse, that this will have a negative impact on the target population. Consider that the technological system is only one part of a larger decision-making process, on which people's well-being may depend. In this case, it is necessary to stop and ask: What skills are needed for operating the system? What prior knowledge is it necessary to have about the purpose for which the technology is implemented? What tools do the people who will implement and operate the system have to correct errors or to handle unforeseen situations? Based on that information, the process for designing training of the people responsible for implementing or operating the system can begin.

## If you answered “yes”

Confirm it by conducting system operating tests. Be sure that when you perform these tests, people receive enough feedback on the way it operates that they can overcome difficulties and knowledge gaps. Ensure as well that there are mechanisms for this knowledge to be regularly reviewed and updated, and that it is passed on to any person who begins to participate in system implementation or operation.

Is the conduct of an independent evaluation of the impact that the system's implementation will have on human rights included?

## If you answered “no”

Before continuing with system implementation, it is necessary to perform this evaluation, to get ahead of foreseeable problems, prevent risks and guarantee that the system does not affect people's basic rights, either directly or indirectly. One of the essential elements for ensuring that AI systems do not cause harm through discrimination, poor decision-making and, in general, the failure of implementation, is the existence of ways to evaluate the system's probable impact on people's fundamental rights. Ensure the availability of resources to find external people to conduct the evaluation, who have the experience and knowledge to perform the analysis. Be sure you have all relevant information available for conducting the independent evaluation.

## If you answered “yes”

Ensure that the evaluation process is planned ahead of time; that it includes enough process time, financial resources and materials to take a comprehensive look at the system; and that there is information available for those conducting the evaluation. Ensure that the system does not begin to operate without the results of this evaluation and, where applicable, following the recommendations resulting from the process. Take care to maintain conditions of transparency to guarantee the independence and autonomy of the people performing the evaluation.

**Are periodic, independent audits of the system projected?**

## If you answered “no”

These audits must be incorporated into the implementation plan. The system cannot be implemented if no audits are contemplated. There is no perfect technological system, and failures can affect human rights. Without audits, the system could incorporate and perpetuate problems that are not detected in time, including those that have not been foreseen by an impact evaluation. It is necessary to plan and organize the execution of independent, periodic, transparent audits whose results are available to the public. Ensure the availability of resources for hiring expert auditors. The audit must contemplate not only a technical examination, but also variables related to the real impact on people's interests and rights. Make sure all relevant information is available for conducting the independent evaluation.

## If you answered “yes”

Be sure that the process is incorporated into the implementation calendar in a manner that is transparent and accessible to any interested party. To conduct the audit, consider the time needed and sufficient financial and material resources to ensure a comprehensive look at the system, and that there are mechanisms for responding to findings and recommendations. In addition, make sure to uphold conditions for transparency, to guarantee the independence and autonomy of the people performing the audit.

# Are citizen engagement efforts included in the system evaluation?

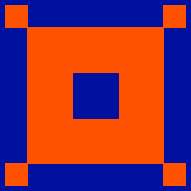
## If you answered “no”

Be sure to create these efforts. There are three important consequences to failing to contemplate citizen engagement: there will not be enough valuable information on the way the system operates in practice; the system will not generate sufficient trust among the people affected by it; and the system will not have enough information for introducing modifications and improvements. A system that does not include citizen engagement efforts will be a system lacking democratic legitimacy, with neither sufficient opportunities for improvement nor trust from the target population.

## If you answered “yes”

Ensure that these efforts reflect the citizenry as best as possible: the efforts should be inclusive, accessible, with understandable and transparent information, and with participation via channels that may be different but equivalent in the value assigned to them. Make sure that the evaluation includes an active attempt to engage with the people to whom the system is targeted. Be sure that there are follow-up, response, evaluation and accountability mechanisms.

Were procedures adopted for the mitigation, correction and reparation to victims of errors that may be produced by the system's implementation?



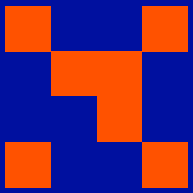
## If you answered “no”

The proposed system lacks the capacity to handle potential failures. The main consequence of this is that citizens are defenseless against the implementation of the technological solution, as passive subjects of its resolutions. This goes against the principles of public administration and jeopardizes the legitimacy of public policies. Remember that there is no infallible technological system, so it must have mechanisms that make it possible to deal with its errors. Otherwise, the recommendation is to not adopt or move forward with implementing the system.

## If you answered “yes”

In addition, consider developing complaint mechanisms that are accessible and inclusive for the whole population: that there are diverse channels and formats for reporting errors, recognizing the population's different capacities. Likewise, it is important that the correction and victim reparation measures be applied swiftly and transparently.

Is there human supervision of the decisions made by the system?



## If you answered “no”

Then the results issued by the technological solution will be directly integrated into the flow of decision-making. This could be critical considering potential biases existing in available data or in the programming of the algorithms that comprise the system. Without human supervision, these biases could be replicated in the system's implementation, which could result, for example, in cases of arbitrary discrimination. Remember that algorithmic systems represent additional layers of opacity in the decision-making process and thus it is critical to consider mechanisms for detecting and correcting biases and errors starting with the system design stage, along with adequate accountability for its operation.

## If you answered “yes”

You should pay particular attention to the point where the automated decision is located in the implementation design flow, and when the supervision is conducted, to avoid leaving an opening for carryforward errors in decision-making. Should the supervision be conducted on a sample of analyzed data, it is important that this sample be representative of the demographic parameters of the population targeted by the implemented system. In all events, the explainability and accountability mechanisms must be considered starting with the system's design, to guarantee an adequate diagnosis, identification and correction of potential biases and errors.

**Does the system include mechanisms that enable the traceability of decision-making, for both automated decisions and human decisions made based on information processed by the system?**

## If you answered “no”

It will be very difficult to adequately diagnose any kind of system failure or error. It is particularly important to identify when the coordination between automated and human decisions is produced, as well as any potential problems that could be generated. Traceability not only enables better system management, it is also required in order to have a policy implementation that satisfies the transparency requirements inherent to public administration. Therefore, any system that does not include this kind of mechanism should be reconsidered and its design reviewed before moving forward.

## If you answered “yes”

It is important to verify that the traceability in the decision-making process is exhaustive and that the information gathered is sufficiently complete for effectively identifying potential errors. Only with complete, exhaustive information will it be possible to account for the system's operation transparently. Furthermore, it is important that such information be communicated to citizens actively and understandably. Consider as well that explainability, human supervision and decision traceability are essential for preventing, mitigating and repairing harm to human rights, and these should be incorporated in the legal framework regulating AI use in the public sector.

# Is the implemented algorithm open source?

## If you answered “no”

The system will have a relevant aspect whose specific operation will be unknown to those who implement the system. This means that the decisions will be made taking the opacity of the algorithmic programming as a given element of the system. The above may turn out to be highly problematic, especially when conducting the necessary audits of the system and/or publishing them for public scrutiny. Remember that, most often, companies protect their code based on intellectual property, and this undermines the possibilities for state action and intervention, such that this type of clause should be carefully analyzed when determining which system to adopt.

## If you answered “yes”

Having an open source code algorithm enables a full scrutiny of its programming and operation, which is important for auditing its function in the decision-making process. The above does not exclude the need to be advised by experts who are able to understand the development of the programming and its effects on the results obtained. Be sure that there are sufficient resources and planning for system maintenance and to respond to analyses, whether internal or external, that may potentially identify errors in the code.

Are there restrictions on the state regarding the possibility of modifying the system, making information on its operation public or improving the way in which the collected data are protected?

## If you answered “no”

The state will have the minimum necessary tools for the implementation of technologies with openness to recognizing social problems. The above must be understood as a relevant step following the conduct of evaluations and audits of the system's operation, essential for adequately reporting on any improvement.

## If you answered “yes”

State institutionality will not have the tools required for fixing any error or failure of the system, whether in the data that feed it or in its operation, and it will not be able to adequately report on its performance evaluation to the public. All these are relevant questions for a suitable exercise of public administration, and, in light of the inability to comply, it would be prudent to critically evaluate the agreement under which the public policy may be developed. It is advisable as well to review alternative, less invasive options identified in the diagnosis statement of the problem before continuing.