6

Companies developing compliance with requirements

Guide 6. Human oversight

European Artificial Intelligence Act

Financiado por la Unión Europea
NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Plan de Recuperación, Transformación y Resiliencia

España | digital 2026

This guide has been developed within the framework of the development of the Spanish pilot for the regulatory AI Sandbox, through collaboration among participants, technical assistance providers, potential competent national authorities, and the sandbox's expert advisory group.

The aim of the guide is to serve as an introductory support to the European Regulation on Artificial Intelligence and its applicable obligations. Although **it is not legally binding and does not replace or develop the applicable legislation, it provides practical recommendations** aligned with regulatory requirements, pending the approval of the harmonised implementing standards for all Member States.

This document is subject to an **ongoing process of evaluation and review**, with periodic updates in line with the development of standards and the various guidelines published by the European Commission. This guide has been updated in accordance with the Digital Omnibus amendments to the AI Act.

Among the relevant applicable technical references, particular note should be made of the standard **"prEN 18229-1 Artificial Intelligence Trustworthiness Framework – Part 1: Logging, Transparency and Human Oversight,"** which is currently under development.

NIPO: 230260272

Revision date: 17, July 2026

- Document version control:

Version	Revision date	Major changes
1.0.	10/12/2025	First official release
2.0.	17/07/2026	Revised to reflect the changes introduced by the Digital Omnibus and minor editorial changes

General content

1. Preamble 5
2. Introduction 7
3. European Regulation on Artificial Intelligence 10
4. How to approach the requirements? 13
5. Technical documentation 30
6. Self-assessment questionnaire 33
7. Annexes 34
8. References, Standards & Norms 35

Detailed Index

1. Preamble 5
1.1 Purpose of this document 5
1.2 How to read this guide? 5
1.3 Who is it for? 5
1.4 Use Cases 5
2. Introduction 7
2.1 What is Human Oversight in Artificial Intelligence? 7
2.2 Why the Need for Human Oversight 8
3. European Regulation on Artificial Intelligence 10
3.1 Previous analysis and relationship of the articles 10
3.2 Content of the articles in the AI Act 10
3.3 Correspondence of the articles with the sections of the guide 12
4. How to approach the requirements? 13
4.1 Human Oversight Requirements in the AI Act 13
4.1.1 Paragraph 1. Design and development for effective oversight 13
4.1.2 Section 2. Risks on fundamental rights 14
4.1.3 Section 3. Types of measures 15
4.1.4 Section 4. Understanding and Autonomy 16
4.1.5 Section 5. Remote biometric identification 20
4.2 Applicable measures to achieve Human Oversight 21
4.2.1 Design and development measures for effective oversight 21
4.2.2 Enable a human-machine interface (HMI) 23
4.2.3 Governance model 23
4.2.4 Awareness. Forced error 25
4.2.5 Governance. Human in/on the loop 27
4.3 Executive summary. List Section-applicable measures 29
5. Technical documentation 30
6. Self-assessment questionnaire 33
7. Annexes 34
7.1 Glossary 34
8. References, Standards & Norms 35
8.1 Standards 35

1. Preamble

1.1 Purpose of this document

Article 14 of **the European Regulation on Artificial Intelligence (*AI Act*)** is dedicated to Human Oversight of high-risk AI systems. This article is part of the second chapter (*Requirements for high-risk systems*) of the aforementioned regulation.

The version of the European Regulation on Artificial Intelligence referenced in this document has been published by the Council of the European Commission on 13 June 2024.

This document provides implementation measures that facilitate compliance with the requirements expressed in the aforementioned article.

1.2 How to read this guide?

As mentioned above, **this document provides** implementation **measures** for entities providing and responsible for the deployment of AI systems **that facilitate, on an indicative basis, compliance with the** obligations expressed in Article 14 of the AI Act, dedicated to Human Oversight.

To this end, the document **goes through** all the sections of said article in order, answering the fundamental questions necessary to **facilitate the fulfilment** of the obligations expressed in these sections.

1.3 Who is it for?

This document is an **implementation guide** on Transparency in Artificial Intelligence to achieve the objectives set by the European Regulation on Artificial Intelligence (*AI Act*).

Therefore, this document is aimed at:

- Managers of the provider entity who conceptually design the AI system according to the requirements of the deployer, who may take into account the measures described in this document to create a Transparent AI system based on the requirements described in the AI Act.
- Deployers, who must be aware of the requirements on Transparency that they will have depending on the use case and process that the AI system will support.

Throughout the document, language is used that is understandable by all of them, minimizing the technicalities necessary for its understanding.

1.4 Use Cases

To contextualise each of the measures presented, which may help, on an indicative basis, to meet the requirements of the AI Act, examples will be used on two use cases:

- Aid granting automatic system

- Chronic Disease Management - Smart Insulin Pump

These use cases are developed in detail in the **Guide 2. Practical guide and examples to understand the AI Act**.

The examples of these use cases are presented at a high level, without going into detail or being exhaustive, in order to try to cover as many cases as possible. In addition, they do not respond to real experiences (but with the intention of being realistic from a didactic point of view), with the aim only of clarifying the measures a little more, therefore they cannot be taken as specifications in a real implementation.

2. Introduction

2.1 What is Human Oversight in Artificial Intelligence?

People must be able to make autonomous informed decisions in relation to AI systems. Because of this, they should be provided with the knowledge and tools necessary to understand and interact with AI systems in a satisfactory manner and, whenever possible, allow them to evaluate or question the system. AI systems should help people make better, more informed decisions in line with their goals. The general principle of human oversight **must be at the heart of the functionality of the system.**

Human oversight helps ensure that an AI system does not undermine human autonomy or lead to other adverse effects. Oversight can be carried out through **governance mechanisms**, such as human participation, control, or human command approaches. Human participation refers to the ability for **human beings to intervene in all the decision-making cycles of the system**, something that in many cases is not possible or desirable. Human control refers to the ability for human beings to intervene **during the life cycle of the system and in the monitoring of its operation**. Finally, human command is the ability to monitor the overall activity of the AI system, as well as the ability to **decide how and when to use the system in a given situation.** This may include deciding not to use an AI system in a particular situation, establishing levels of human discretion during the use of the system, or ensuring the possibility of ignoring a decision made by a system. It may be necessary to introduce monitoring mechanisms to varying degrees to support other security and control measures, depending on the scope and potential risk of the AI system. If all other circumstances do not change**, the lower the level of oversight a person can exercise over an AI system, the greater and more demanding the verifications and governance required.**

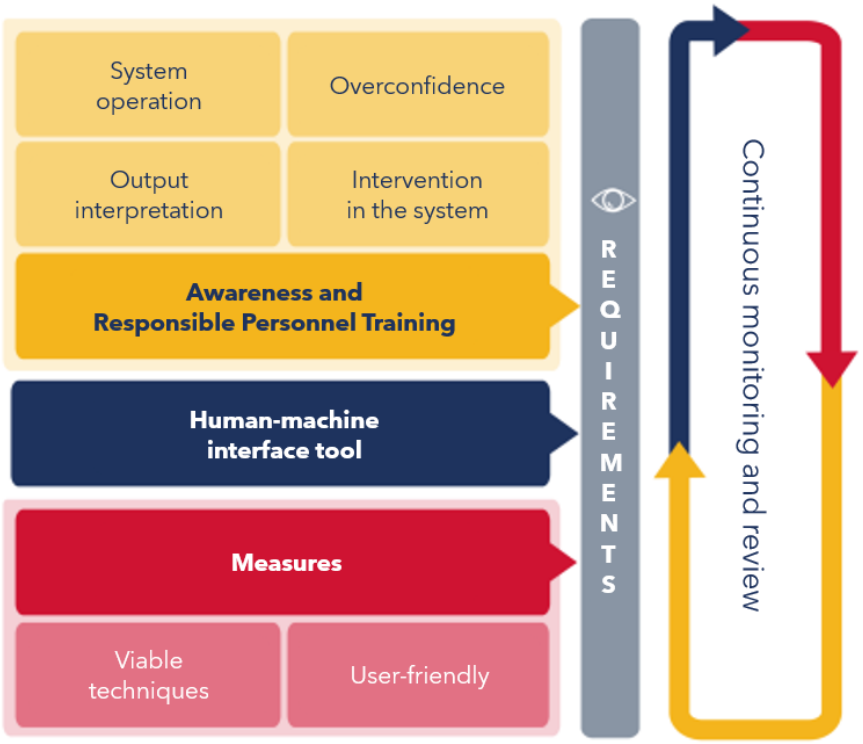
The final responsibility for the actions carried out by an AI system is the responsibility of the people of the provider and user entity responsible for it. It is therefore necessary that these persons be able **to monitor it** (first paragraph of Article 14). For such oversight to be effective, people must have **control over the system** and be able to manage **risks** that may arise from its use (second paragraph of Article 14). To do this:

- There must be **human presence** at some point in the process in which the AI system intervenes.
- The system must be designed and built in such a way as to allow **its reasoning mechanisms and results** to be interpreted.

In the first instance, we can identify an important relationship between the concept of human oversight and two others developed in the second chapter of the European Regulation on Artificial Intelligence (*Requirements for high-risk systems*):

- **Transparency**. For a system to be monitored, it is necessary to understand how it works, and for such understanding to exist, it is essential that the system be transparent.
- **Risks management**, since for oversight to have effective and full guarantee, it is necessary to have **control over the system** and to be able to manage **risks** that may arise from its use (second paragraph of Article 14).

As a visual introductory summary, an infographic is provided that tries to give an overview of the design of high-risk AI systems that allow effective oversight to be carried out on them:



2.2 Why the Need for Human Oversight

Artificial Intelligence is one of the most complex software technologies of all those we have assimilated. This is because their capacities (prediction making, decision-making, and even emulation of cognitive capacities of human beings to make such predictions and decision-making) are like those of human beings. **The mechanisms necessary to massively automate these capabilities** through AI systems entail a **complexity that needs to be monitored** to generate trust due to the criticality of the actions it can perform.

On the other hand, for efficiency reasons, people may decide to use AI systems. However, **this transfer of control** cannot be total, always maintaining a human component to also facilitate the **management of responsibility and accountability** due to the actions of the AI system.

In this document, for each paragraph of Article 14 of the European Regulation on Artificial Intelligence (dedicated to Human Oversight), possible measures are presented that may, on an indicative basis, provide human oversight of an AI system in accordance with the requirement set out in that section. Each of these measures is exemplified in detail by the use case described at the beginning of the document.

To introduce the concept of Human Oversight and the importance of its necessity, it is enough to ask ourselves a question through one of the use cases that will serve as a common thread throughout the following guide:

What would happen if, as a result of an error in the collection of blood parameters due to a physical failure of the device, the AI system proposed to the patient a dose of insulin that was

not appropriate for their ailment, a doctor did not supervise the prescription made by the system, and the patient proceeded to inoculate them?

The answer is simple: medical control would have been completely delegated to an AI system subject to errors, and the consequences of this action would have a negative impact on a fundamental right of people: health.

And although in the cases of biometric systems, a failure in human oversight has very obvious and striking consequences, this requirement is not dictated by the AIA only for one of them, but for all high-risk AI systems, although in the former with particular conditions.

Throughout this document we will see how to avoid this type of situation.

3. European Regulation on Artificial Intelligence

The putting into service or use of high-risk AI systems should be subject to compliance with certain mandatory requirements, including human oversight. Those requirements aim to ensure that high-risk AI systems available in the Union or whose output outputs are used in the Union do not pose unacceptable risks to important public interests recognised and protected by Union law.

This section includes the articles referring to the generation of records of Regulation 2024/1689 of the European Parliament and of the Council, of 13 June 2024 (European Regulation on Artificial Intelligence) and details in which sections of this guide the different elements of these articles are addressed.

3.1 Previous analysis and relationship of the articles

The obligations on the generation of records are mainly found in Article 14 "Human oversight"

- **Article, human oversight:** Establishes that HRAIS must incorporate the technical capacities necessary to enable human oversight. This article is divided into five sections:
    1. Effective oversight during the period they are in use.
    2. Establish mechanisms to minimize health, safety, fundamental rightsrisks.
    3. Establish responsibilities.
    4. Establish mechanisms of transparency, explainability and traceability.
    5. Ensure human oversight in systems with remote biometric identification.

3.2 Content of the articles in the AI Act

AI Act

Art.14 – Human oversight

1. High-risk AI systems shall be **designed and developed** in such a way, including with **appropriate human-machine interface tools**, that they can be **effectively overseen by natural persons** during the period in which they **are in use**.
2. Human oversight shall **aim** to **prevent or minimise the risks** to health, safety or fundamental rights that may emerge when a high-risk AI system is used **in accordance with its intended purpose** or under conditions of **reasonably foreseeable misuse**, particularly when such risks persist despite the application of other requirements set out in this Section.
3. The oversight measures shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system, and shall be ensured through either one or both of the following **types of measures**:

(a) **measures identified and built**, when **technically feasible**, into the high-risk AI system by the provider before it is placed on the market or put into service;

(b) measures **identified by the provider before** placing the high-risk AI system on the market or putting it into service and that **are appropriate to be implemented by the deployer**.

4. For the purpose of implementing paragraphs 1, 2 and 3, the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:

(a) to properly **understand** the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;

(b) to remain **aware** of the possible tendency of **automatically relying** or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;

(c) to **correctly interpret** the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available;

(d) **to decide**, in any particular situation, **not to use the high-risk AI system or to otherwise disregard, override or reverse** the output of the high-risk AI system;

(e) to intervene in the operation of the high-risk AI system or **interrupt the system** through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state

5. For high-risk AI systems referred to in point **1(a) of Annex III, the measures referred to in paragraph 3** of this Article shall be such as to ensure that, in addition, **no action or decision is taken** by the deployer on the basis of the identification resulting from the system unless that identification has been separately verified and confirmed by **at least two natural persons** with the necessary competence, training and authority.

The requirement for a separate verification by at least two natural persons shall not apply to high-risk AI systems used for the **purposes of law enforcement, migration, border control or asylum**, where Union or national law considers the **application of this requirement to be disproportionate**.

3.3 Correspondence of the articles with the sections of the guide

The following table lists the sections of the document in which the explanation and possible measures applicable to each section/subsection of the article dedicated to human oversight are found.

Article	AI Act requirement	Section
14.1	Design and development for effective oversight	Section 4.1.1
14.2	Risks on fundamental rights	Section 4.1.2
14.3.a	Types of measures defined and integrated, where technical feasible	Section 4.1.3
14.3.b	Types of measures before the introduction of the AI system	Section 4.1.3
14.4	Understanding and Autonomy	Section 4.1.4
14.4.a	Understanding Capabilities and Limitations	Section 4.1.4.1
14.4.b	Automation bias	Section 4.1.4.2
14.4.c	Interpreting the Output Information	Section 4.1.4.3
14.4.d	Autonomy to decide	Section 4.1.4.4
14.4.e	Interruption	Section 4.1.4.5
14.5	Remote biometric identification	Section 4.1.5

4. How to approach the requirements?

We remind you that this guide takes as a reference Regulation 2024/1689 of the European Parliament and of the Council, of 13 June 2024 (European Regulation on Artificial Intelligence).

The article on Human Oversight has **five sections.** Here's a **summary** of the **requirements** expressed about high-risk AI systems:

1. That they are **designed and developed so that, when they are in use**, effective oversight can be carried out on them.
2. Mechanisms are in place to prevent or **minimise risks** to health, safety or fundamental rights.
3. **Establish the responsibility** of the provider and the deployer.
4. That the **measures used to achieve the requirements** of the previous paragraphs **allow the persons** responsible for the oversight of the AI system:
   - Understand the system (sections 4a and 4c).
   - Be aware of their autonomy over the system (sections 4b, 4d and 4e).
5. Ensure human oversight in a particular type of AI system: those dedicated to **remote biometric identification.**

Each of the sections and subsections of Article 14 dedicated to human oversight is set out below, indicating, for each of them, the possible measures to address, on an indicative basis, the requirements set out in the AI Act.

4.1 Human Oversight Requirements in the AI Act

4.1.1 Paragraph 1. Design and development for effective oversight

AI Act

Art.14.1- Human oversight

High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use.

What we understand

The article begins with an overarching objective: that high-risk AI systems are designed and developed so that, when they are subsequently in use, they can be effectively monitored by the people responsible for the system.

Possible measures to carry it out

Therefore, first to meet this requirement, it is necessary to detail the design and development measures for effective monitoring of the AI system.

Once the system has been designed and developed under these measures, the system is in use. And then it is necessary to enable mechanisms that allow the result of these measures to be monitored. Because of this, the section anticipates the need to enable a "human-machine interface".

4.1.2 Section 2. Risks on fundamental rights

AI Act

Art.14.2 - Human oversight

Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular where such risks persist despite the application of other requirements set out in this Section.

What we understand

In this section, it emphasises how human oversight should take into account the **management of risks** that may **affect the health, safety and fundamental rights** of individuals, whether for the intended purpose of the AI system or for other reasons. reasonably foreseeable misuses made of it.

In fact, Article 9, dedicated to the Risks Management System, begins its requirements by also highlighting in its second paragraph these three areas: health, safety and fundamental rights.

As for **fundamental rights,** this concept is cited in the European Regulation on Artificial Intelligence more than fifty times since high-risk AI systems, the objective of the AI Act, must ensure the safeguarding of these rights. These rights are those enshrined in the Charter of Fundamental Rights of the European Union.

As for the mention of the health domain, it is a specific specification of the aforementioned charter (Article 35). Risk management in this domain is especially important since, for example, the risk in an AI system in charge of administering a drug to a patient is high since such a dose has a direct impact on the health and life of that patient.

As for the mention of the domain of **security**, it is also a specific specification of the aforementioned charter (transversal, with specific mention in its sixth article). Risk management

in this domain is equally important since, for example, risk in an AI system tasked with safety in an autonomous vehicle can lead to accidents that affect people's lives.

Possible measures to carry it out

Human oversight must be able to manage risks in these three domains. To this end, risks management of the AI system must be carried out in accordance with the measures developed in the **guide round Article 9** (Risks management system), applying these **measures both to the intended purpose of the system, and to the reasonably foreseeable misuses** that may be made of it.

The measures defined in this guide include an **illustrative example of the development of a risk management system for various use cases**, supported by a documented *checklist* with a simple interface.

This risks management must be framed within the governance model, a specific measure to ensure Human Oversight, developed in the Governance **Model section** of this document.

4.1.3 Section 3. Types of measures

AI Act

Art.14.3 - Human oversight

The oversight measures shall be commensurate with the risks, level of autonomy and context of use of the high-risk AI system, and shall be ensured through either one or both of the following types of measures:

(a) measures identified and built, when technically feasible, into the high-risk AI system by the provider before it is placed on the market or put into service;

(b) measures identified by the provider before placing the high-risk AI system on the market or putting it into service and that are appropriate to be implemented by the deployer.

What we understand

Establish the responsibility of the provider and the person responsible for the deployment:

- The provider must provide the deployer with measures that facilitate human oversight of the system.
- The deployer must implement them.

Possible measures to carry it out

All those indicated in the section on applicable measures of this document.

4.1.4 Section 4. Understanding and Autonomy

AI Act

Art.14.4 - Human oversight

For the purpose of implementing paragraphs 1, 2 and 3, the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:

(a) to properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;

(b) to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;

(c) to correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available;

(d) to decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system;

(e) to intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state

What we understand

In short, that the measures used to achieve the requirements of the previous paragraphs allow the persons responsible for the oversight of the AI system to **understand** the system (sections 4a and 4c), and to be aware of their **autonomy** over the system (sections 4b, 4d and 4e).

Each of the subsections is developed below.

4.1.4.1 Section 4a. Understanding Capabilities and Limitations

AI Act

Art.14.4a - Human oversight

a) to properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;

What we understand

For human oversight to exist, it is essential that the people responsible for the AI system can understand in detail the capabilities and limitations of it.

This requirement is **also expressed in section 13.3.b** of Article 13 corresponding to Transparency. As mentioned above, this relationship makes sense: for a system to be monitored, it is necessary to understand how it works, and for such an understanding to exist, it is essential that the system be transparent.

Possible measures to carry it out

The measures reflected in the Transparency guide on paragraph **13.3.b**

4.1.4.2 Section 4b. Automation bias

AI Act

Art.14.4b - Human oversight

b) be aware of the potential tendency to automatically or over-rely on the output outputs generated by a high-risk AI system ('automation bias'), in particular with systems that are used to provide information or recommendations for natural persons to make a decision;

What we understand

People may become overconfident in the output of an AI system, even more so than in another person. This requirement aims to reinforce the concept of human oversight, making **it aware that the final responsibility,** whether for a prediction or a decision, is inherent to the people who interact with the high-risk system, who must be properly trained in the business process supported by the system, since they are the ones who must evaluate and make the final decision.

Possible measures to carry it out

- Include in the system a forced error mode to test the criteria and possible confidence of the people who will use the system
- The training indicated in the governance model.

4.1.4.3 Section 4c. Interpreting the Output Information

AI Act

> **Art.14.4c - Human oversight**
>
> c) correctly interpret the output results of the high-risk AI system, taking into account, for example, the available interpretation methods and tools

What we understand

Ensure the existence of methods and tools that allow the correct interpretation of the system's output.

Possible measures to carry it out

This is yet another example of the close relationship between human oversight of the high-risk system and its transparency, since in order to monitor a system it is necessary to understand how it works and, for such an understanding to exist, it is essential that the system be transparent. Therefore, in order to comply with this requirement, it is necessary to resort to **measures already available and reflected in the Transparency guide**, especially the following, as they are directly related to the interpretation of the system's output information:

- Detailing from the most global to the most particular.
- Adapt the language.
- Use counterfactuality.

4.1.4.4 Section 4d. Autonomy to decide

AI Act

> **Art.14.4d - Human oversight**
>
> d) To decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system;

What we understand

Reinforce the concept of **autonomy as part of human oversight** since, in general, AI systems must be designed in such a way that they increase, complement and enhance people's capacities, allowing them to decide when and how to use the system in each given situation. This may include deciding not to use an AI system in a particular situation, establishing levels of human decision during the use of the system, or ensuring the ability to prevail over a decision made by the system.

Both this requirement and that of the following section, related to the interruption of the system, are the ones that absolutely determine **the level of human oversight.**

Possible measures to carry it out

- Governance Human in/on the loop

4.1.4.5 Section 4e. Interruption

AI Act

> Art.14.4e - Human oversight
>
> e) To intervene in the operation of the high-risk AI system or interrupt the system through a 'stop' button or a similar procedure that allows the system to come to a halt in a safe state

What we understand

Like subparagraphs (b) and (d), **reinforce the concept of autonomy** as part of human oversight. This case requirement raised is an additional specification to the one set out in the previous subsection d (Autonomy to decide), going so far as to establish the requirement of having a mechanism for deactivating the system and/or an associated procedure.

Possible measures to carry it out

- Governance Human in/on the loop

4.1.5 Section 5. Remote biometric identification

AI Act

Art.14.5 - Human oversight

For high-risk AI systems referred to in **point 1(a) of Annex III, the measures referred to in paragraph 3** of this Article shall be such as to ensure that, in addition, **no action or decision is taken** by the deployer on the basis of the identification resulting from the system unless that identification has been separately verified and confirmed by **at least two natural persons** with the necessary competence, training and authority.

The requirement for verification by at least two separate persons shall not apply to high-risk AI system **used for law enforcement, migration, border control or asylum purposes** where national or Union law considers **the application of this requirement to be disproportionate.**

What we understand

The systems referred to in *Annex III.1.a* are **remote biometric identification systems.** For such systems in particular, it should be ensured that the measures specified in paragraph 3 of this Article ensure the separate verification of two separate persons, except where European Union or Spanish national law considers that such a measure is disproportionate in the use of such systems to ensure compliance with the law, migration, border control or asylum.

It is important to **differentiate between biometric identification and biometric verification**, as they are frequently used as equivalent terms despite presenting relevant technical and operational differences.

- **Biometric identification:** a process intended to determine the identity of a person without that person having previously declared it, by comparing their biometric data against a database of biometric references. It is a one-to-many (1:N) search, aimed at finding a match within a broad population, which usually entails a higher processing load and greater resource requirements. Biometric identification is commonly used in governmental and security contexts, such as border control or certain public surveillance activities.
- **Biometric recognition:** an automated one-to-one (1:1) process, including authentication, whose purpose is to confirm the identity of a natural person by verifying that they are who they claim to be. It consists of comparing the biometric data captured at the time of verification with the biometric data previously provided by that same individual –the registered reference or template–. Remote biometric verification is used, for example, in digital customer onboarding in the banking sector, by comparing a selfie or video capture with the photograph on an ID card or passport, normally incorporating liveness detection mechanisms. It is also commonly used for secure account recovery or for authorising high-risk transactions.

Once both concepts have been differentiated, it should be noted that the European Regulation on Artificial Intelligence refers to **remote biometric identification** AI systems**.**

Possible measures to carry it out

The **governance model** measure on the AI system for effective oversight referred to in paragraph 3 shall take into account the casuistry indicated for these remote biometric identification systems.

4.2 Applicable measures to achieve Human Oversight

This chapter of the document includes the details of the possible measures to address, on an indicative basis, the requirements of human oversight set out in Article 14 of the AI Act.

4.2.1 Design and development measures for effective oversight

To be able to monitor any *software system,* it is necessary **to know how it is designed and built**. The next necessary step is to establish the **dimensions** of this design and construction on which to carry out human oversight. In the case of high-risk AI systems, the European Regulation on Artificial Intelligence defines these dimensions in its **third chapter second section** (*Requirements for high-risk systems*). This chapter is divided into the following articles:

- Article 9. Risks management.
- Article 10. Data and its governance
- Article 11. Technical documentation
- Article 12. Record-keeping
- Article 13. Transparency
- Article 15. Accuracy, robustness and cybersecurity

Given the references made in various sections of Article 14 to understanding and risks, among all the dimensions we can highlight:

- **Transparency**. For a system to be monitored, it is necessary to understand how it works, and for such understanding to exist, it is essential that the system be transparent.
- **Risks management**, since for oversight to have effective and full guarantee, it is necessary to have **control over the system** and to be able to manage the **risks** that may arise from its use (second paragraph of Article 14).

All the necessary measures are detailed in the implementation guides of each of the aforementioned articles.

Who it applies to

The measures reflected in each of the aforementioned guides reflect who applies, either the provider or the deployer of the AI system. As an overall summary:

- In the design and development measures of the system used by the deployer:
  - The deployer is responsible for identifying the requirements of the European Regulation on Artificial Intelligence that must be met by the business process that the system will support. In addition, it must guarantee through acceptance tests that these requirements are finally supported by the system.

  - The provider must implement the necessary technical measures aligned with these requirements.
- When the system is in use:
  - The deployer is responsible for ensuring that the behaviour of the system in production conforms to the requirements of the AI Act.
  - The provider's work does not end with deployment but also continues after the system is in production. This relationship, common in any *software* system, is especially important in AI systems, given their complexity, and particularly in high-risk ones given the criticality of the processes they support.

Example

We use as an example a measure used to achieve Transparency and which is detailed in the guide of that article: *Adapt the language*.

By way of a reminder in summary, this measure states that:

- The AI system must be designed so that it can provide information to all the profiles that interact with it and thus transparently ensure its complete understanding.
- There are many types of profiles that interact with the AI system throughout its life cycle. Therefore, technical mechanisms must be set up to allow this information to be displayed in a transparent and understandable way for all of them, adapting the type of language to their level of dialogue.

Below, we particularize it for the two use cases, using those reflected in the Transparency guide. In both cases, the system will provide a **user interface**, aligned with the level of dialogue of each of the profiles that interact with the system, and that provides the following information in an easily understandable way, visually and textually in natural language, especially to those non-technical profiles**, in order to facilitate their Human Oversight.**

Example – Aid granting automatic system

- **Those responsible for granting aid** must receive information from the system, **in natural language**, that allows them to ensure that they are complying with the policies for granting aid, the degree of accuracy they are having when predicting the social exclusions that finally occur and that give rise to the aid, how homogeneous with the predictions they are providing in families with similar characteristics, details about all predictions made and decisions made by the AI system.
- **The technicians** who implement the system and are responsible for monitoring it while it is in operation, must know the same information as above, but also **from a technical perspective.**
- The system must be able to explain **in natural language to the applicant families** the reasons why they have been granted a certain amount and not another, why their application has been denied and under what conditions applied to their reality they would be accepted. Etc.

Example - Insulin Pump

- The **physicians** responsible for administering the doses to their patients must receive information from the system, **in natural language** with the necessary technical detail from a medical point of view, which allows them to ensure that the doses administered are having the correct effect from a medical point of view, according to the desired levels of accuracy.
- The **software technicians** who implement the system and are responsible for monitoring it while it is in operation must know the same information as above, but **from a technical perspective**.
- The system must be able to explain **to patients in natural language** the doses they are receiving, their effect, etc., in the same way that the doctor would explain it to them in one of their check-ups.

To which sections does this measure apply?

- Paragraph 1. Design and development for effective oversight

4.2.2 Enable a human-machine interface (HMI)

The measures implemented in the design and development phases of the system must be reflected when the system is in use in order to be able to supervise them and thus ensure that the operation is adequate. And this oversight must be able to be carried out by technical profiles as well as by non-technical profiles. Given that AI systems have been recently deployed, this human-machine interface (HMI) for monitoring and oversight (something common in traditional *software* systems) is usually in the embryonic phase. This is, together with the high-risk nature of AI systems, why the European Regulation on Artificial Intelligence emphasises the need for such an interface.

Who it applies to

The AI system provider will provide this interface to the deployer, who will make use of it while the system is in use.

Example

See example of the previous measure, where the need for an interface is already described.

To which sections does this measure apply?

- Paragraph 1. Design and development for effective oversight

4.2.3 Governance model

The concept of *IT governance* is key to achieving the goal of human oversight over high-risk AI systems. Such governance is necessary to be carried out during the design and development of the AI system to ensure that the creation of the AI system is aligned with the AI Act. But **when the system is already in use, monitoring the system through such a governance model is especially critical,** helping to ensuring that the behaviour of the high-risk system remains

aligned with the AI Act. This provides end-to-end monitoring of the AI system throughout its entire lifecycle.

The governance model should include:

- **A multidisciplinary organizational structure** including all the profiles that are part of the system's life cycle (business managers, technicians, lawyers and auditors, among others).
- **Procedures** that allow the solution to be supervised from the time it is conceptualized and designed **until it is in use** (responsibility from design). As regards the time at which it is in use, if the system is one of those referred to in *Annex III. 1a* (remote biometric identification systems), that governance model shall include a procedure ensuring the separate verification of two separate persons, except where the European Union considers that such a measure is disproportionate in using such a system to the guarantee of compliance with the law, migration, border control or asylum.
- A fundamental part of any IT governance model is **risk management.** It is desirable to have a risks management framework associated with the AI system and to carry out regular assessments of the risks and impact of the AI system throughout the life cycle of the, **especially when it is in use.**
- Artificial Intelligence and the measures necessary for its oversight are relatively new fields of knowledge for many people. As part of the governance model, it is also necessary to **enable training plans** that allow training all the people who interact with the system throughout its life cycle on:
    - The design and development measures that have been taken on said system, in accordance with each of the requirements defined in the AI Act.
    - How to use the user interfaces that allow monitoring the measures defined in the design and development.
    - The governance system itself, including the risks management described above.

Given the criticality of high-risk AI systems, such training should be accompanied by an evaluation model to ensure the assimilation of the concepts transmitted.

Who it applies to

- The **provider** shall have a governance model during the construction and use of the model and provide the governance guides that the deployer shall carry out when the system is in use.
- The **deployer** shall adapt that governance system to the performance of its organisational structure and procedures.

Example - Insulin Pump

When the system is in use, the governance model will include:

- A multidisciplinary organisational structure that includes profiles that can cover medical treatments, support the technical operation of the system to ensure its operation, and manage the risks that may affect the health of patients (a fundamental right on which the *AI Act* focuses).

- Procedures that involve these profiles and that, integrated into the business process supported by the system, allow the oversight of its functionalities, such as, for example:
  - The reception and monitoring of the indicators collected online by the system on the patient in order to monitor the treatment.
  - The management of alerts triggered by the system or directly by the patient himself.
  - The validation of the doses proposed by the system before the inoculation that will be carried out on the patient.
- A training and continuous evaluation plan that ensures:
  - That the medical and technical profiles described above are capable of understanding, using, supervising the system and managing the health risks that may arise from its use.
  - That patients also understand how to use it when inoculating themselves with the dose prescribed by the patient and validated by the doctor.

In both cases, paying special attention to the times when a new version or release is introduced into the system, either of data (for example, in the event of the incorporation of new pathologies or new compounds) or of the model (for example, in the event of the incorporation of a functionality that modifies the inoculation time patterns).

To which sections does this measure apply?

- Section 1. Design and development for effective oversight
- Section 5. Biometric identification
- Section 4b. Automation bias

4.2.4 Awareness. Forced error

The system may include a forced error mode that can be activated in the testing phase of the system, or at any time in an environment with data and cases equal to those that the system will have when it is in production. This mode will generate some erroneous outputs in a controlled and deliberate manner, with the aim of **testing the criteria and possible overconfidence** of the people who will use the system in production and thus **evaluate their ability to monitor** it. It will be possible to analyse the responses made by people to deliberate erroneous departures.

Who it applies to

- The system provider will provide the "forced error" functionality that can be activated only in test environments, since doing so in a production environment would be an induction factor to a real error.
- The deployer will use this mode in the tests of the system before its release to production, and also as a training activity for the people who are going to use said system, analysing their responses to be able to evaluate their ability to monitor the system.

Example - Insulin Pump

One of the mistakes in this forced error mode will consist, for example, in the system proposing to the doctor an inappropriate dose to one of his patients based on abnormal blood values also identified by the system on the patient. The doctor must identify this situation, mark it as erroneous, and indicate the correct dose so that the system communicates it to the patient, thus demonstrating that his vigilance over the decision proposals made by the system is correct.

Example – Aid granting automatic system

One of the mistakes in this forced error mode will consist, for example, in the system proposing to the person responsible for the concessions a range of aid amount not aligned with the economic income and needs of the family. The person in charge must identify this situation, mark it as erroneous, and indicate the correct range, thus demonstrating that his or her vigilance over the decision proposals taken by the system is correct.

To which sections does this measure apply?

- Section 4b. Automation bias

4.2.5 Governance. Human in/on the loop

One of the decisions that must be made in the design of the AI system is the level of autonomy that is granted to it. This level of autonomy is part of the procedural aspects to be defined in the governance model of the AI system. There are **three levels:**

- The first, and most global that must always be taken into account as it reinforces the concept of Human Oversight, is that of **Human in Command** (HIC), and refers to how a human being **is ultimately responsible for** and makes critical decisions in the performance of an AI system. This concept delves into the idea of a position of ultimate responsibility of people in the operation of the AI system and the critical decisions that are made in its operation to ensure final decisions and the achievement of the desired objectives in a safe and reliable manner.
- **Human-in-the-loop** (HITL) refers to human intervention in every action in the system, which in many cases is not possible given the high volume of actions it manages. Human oversight in this case is called *ex ante,* and it implies that the action is partially automated. In **high-risk** use cases (those that concern us in the *AI Act*), where it is usually necessary for a person to validate each of the actions, even if the volume of actions is high, this is the level that is usually recommended.
- **Human-on-the-loop** (HOTL) refers to a human intervention of oversight over actions that is carried out a posteriori. Human oversight in this case is called *ex post,* and implies that the action is **fully automated,** although the possibility of human

intervention to reverse it must be enabled later. This level of autonomy is not particularly recommended in the high-risk cases that concern us, due to the risk of an impact that is difficult to reverse on the fundamental rights of individuals, one of the main objectives of the *AI Act*.

Regardless of the level of autonomy provided by both levels, the system could have a mechanism that allows its **immediate deactivation,** without prejudice to the existence of a *Business Continuity Plan* that allows the process supported by the system to recover and restore its functions.

The level used will determine the **level of human oversight** exercised over the AI system and will have an impact on the governance procedures associated with the system when it is in use.

In both cases, the system **will record the two actions:** the one proposed or even executed by the system, and the one finally executed by the person responsible for it, following the measures indicated in the guide in Article 12 (Records).

Who it applies to

- It is up to the deployer to determine the level of autonomy provided to the system based on the use case it supports within their business, as well as the governance procedure associated with that level when the system is in use.
- The provider has to implement the technical mechanisms that each of the aforementioned levels allows.

Example - Insulin Pump

This is a case where, because the immediate application of the decision could put at risk a fundamental right of someone (health), the system must be designed with a *Human-in-the-loop (HITL)* approach. Therefore, when the system proposes the next dose to a patient, the doctor in charge must validate or adjust it based on the patient's blood parameters collected by the system, at which time the patient can refill the insulin pump with the amount indicated by the system and administer it.

Example – Aid granting automatic system

This is also a use case that requires a *Human-in-the-loop* (HITL) level, both because of the "propositional" functional nature of the system that does not require immediacy and, mainly, because of the impact of the decision finally adopted. The system analyses the requests for aid made by families, and the public manager responsible for them will be able to analyse them one by one, establishing the amount finally assigned. Therefore, when the system, for example, proposes the amount assigned to a family, the public official must validate it based on the economic and social information collected from it.

To which sections does this measure apply?

- Section 4e. Autonomy to decide
- Section 4e. Interruption

4.3 Executive summary. List Section-applicable measures

	Measures						
Sections	MV1	MV2	MV3	MV4	MV5	MV6	MV7
1. Design and development for effective oversight	X	X	X				
2. Risks on fundamental right						X	
3. Types of measures	X	X	X	X	X	X	X
4. Understanding and Autonomy							
4a. Understanding capabilities and limitations							X
4b. Automation bias			X	X			
4c. Interpreting the output information							X
4d. Autonomy to decide					X		
4e. Inturruption					X		
5. Biometric identification			X				

MV1	Applicable measures to achieve Human Oversight
MV2	Design and development measures for effective overisght
MV3	Enable a human-machine interface (HMI)
MV4	Awareness. Forced error
MV5	Governance. Human in/on the loop
MV6	Article 09. Risk management system
MV7	Article 13. Transparency

5. Technical documentation

Article 11 (Technical Documentation) states that the system must be documented in such a way as to demonstrate that it meets the requirements set out in Section 2 (to which this article on Human oversight corresponds), providing the competent national authorities and notified bodies with the information necessary to assess the conformity of the AI system with those requirements in a clear and comprehensive manner.

The aforementioned article states that such documentation shall contain, **as a minimum**, the elements set out in **Annex IV.** [1]

Furthermore, this Human oversight guide sets out different measures to address, on an indicate basis, the requirements established by the European Regulation on Artificial Intelligence in the article dedicated to Human oversight in AI systems. **As a result of these measures, aspects of the system set out below can be documented**, which may help to generate the minimum documentation required.

Design and development for effective oversight

1. Deployer. Document with the requirements that the AI system must support in its case of use regarding its risks management, according to the requirements indicated in Article 9 of the AI Act.
2. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding its risks management, according to the requirements indicated by Article 9 of the AI Act.
3. Deployer. Document with the requirements that the system must support in its case of use regarding data and its governance, according to the requirements indicated by Article 10 of the AI Act.
4. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding data and data governance, according to the requirements indicated by Article 10 of the AI Act.
5. Deployer. Document of requirements to be supported by the system regarding its technical documentation, in accordance with the requirements indicated in Article 11 of the AI Act.
6. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding the technical documentation of the same, according to the requirements indicated by Article 11 of the AI Act.
7. Deployer. Document of requirements to be supported by the system regarding its Records, according to the requirements indicated by Article 12 of the AI Act.

[1] SMEs, including start-ups and small mid-cap enterprises, may provide the technical documentation specified in Annex IV in a simplified manner. To this end, the Commission shall establish a simplified technical documentation form tailored to the needs of small and micro-enterprises. Where an SME, including start-ups, chooses to provide the information required in Annex IV in a simplified manner, it shall use the form referred to in this paragraph. Notified bodies shall accept that form for the purposes of conformity assessment.

8. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding the Records of the same, according to the requirements indicated by Article 12 of the AI Act.
9. Deployer. Document of requirements that the system must support regarding its Transparency, according to the requirements indicated by Article 13 of the AI Act.
10. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding its Transparency, according to the requirements indicated by Article 13 of the AI Act.
11. Deployer. Document of requirements that the system must support regarding its Accuracy, according to the requirements indicated by Article 15 of the AI Act.
12. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding its Accuracy, according to the requirements indicated by Article 15 of the AI Act.
13. Deployer. Document of requirements that the system must support regarding its Robustness, according to the requirements indicated by Article 15 of the AI Act.
14. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding its Robustness, according to the requirements indicated by Article 15 of the AI Act.
15. Deployer. Document of requirements that the system must support regarding its Cybersecurity, according to the requirements indicated by Article 15 of the AI Act.
16. Provider. User and technical manuals with the mechanisms for the AI system to support the requirements regarding its Cybersecurity, according to the requirements indicated by Article 15 of the AI Act.

Enable a human-machine interface (HMI)

17. Provider. HMI user manual that allows the deployer to monitor the AI system regarding the requirements of the AI Act in its article 9 on the management of risks of the same.
18. Provider. HMI user manual that allows the deployer to monitor the AI system regarding the requirements of the AI Act in its article 10 on data and its governance.
19. Provider. HMI user manual that allows the deployer to monitor the AI system by accessing the technical documentation required by the AI Act in its article 11.
20. Provider. HMI user manual that allows the deployer to monitor the AI system regarding the requirements of the AI Act in its article 12 on Records.
21. Provider. HMI user manual that allows the deployer to monitor the AI system regarding the requirements of the AI Act in its article 13 on Transparency.
22. Provider. HMI user manual that allows the deployer to monitor the AI system regarding the requirements of the AI Act in its article 15 on Accuracy.
23. Provider. HMI user manual that allows the deployer to monitor the AI system regarding the requirements of the AI Act in its Article 15 on Robustness.
24. Provider. HMI user manual that allows the deployer to monitor the AI system regarding the requirements of the AI Act in its article 15 on Cybersecurity.

Governance model

25. Provider. Document with the Governance Model applicable on the AI system during its development, which must include, at least, the contents indicated in this document regarding said model.

26. Deployer. Document with the Governance Model on the AI system that you apply during the use of the same and that includes, at least, the contents indicated in this document about said model.
27. Provider and deployer. Document with the governance model that includes the specific requirement for cases of Remote Biometric Identification.

Awareness. Forced error

28. Provider. Technical and deployer manuals that describe at least the "forced error" functionality in the AI system to avoid automation bias.

Human in/on the loop

29. Deployer. Document that explains the level of autonomy (HITL/HOTL) that the AI system uses in its use case.
30. Deployer. Document with the governance procedure associated with that level of autonomy when the AI system is in use.
31. Provider. Technical and user manuals describing the mechanisms that allow the deployer to use the required level of autonomy.
32. Provider. Technical and user manuals with the description of the mechanism and/or procedure that allows the deployer to interrupt the operation of the system.

6. Self-assessment questionnaire

To carry out a self-assessment of compliance with the requirements of the European Regulation on Artificial Intelligence referred to in this guide, a global self-assessment questionnaire has been generated with a series of questions with the key points to be taken into account with respect to the obligations dictated by the articles of the European Regulation on Artificial Intelligence mentioned in this guide.

It will be necessary to refer to this document in order to carry out the section of the self-assessment questionnaire corresponding to this guide.

7. Annexes

7.1 Glossary

The content of this document aims to be didactic using understandable language and minimizing technicalities, but at the same time being precise from a technical and formal point of view. When technicalities are used, they are explained in the same text where they are exposed, but in others it is not done so since their explanation could divert the thread of argument of the document. These concepts are detailed in this section.

Life cycle

The life cycle of an AI system is the phases that the system goes through from its conception until it is retired.

The [ISO/IEC 22989] and [ISO/IEC 5338] standards define in depth what the phases of the life cycle of an AI-based system are, from a regulatory point of view. For example, [ISO/IEC 22989:2022, clause 6.1] fundamentally defines the following stages in the life of an AI-based system:

- Conception.
- Design and development.
- Verification and validation of the product or service.
- Deployment.
- Operation and oversight.
- Reassessment.
- Removal or dismantling.

Source: ISO.org

8. References, Standards & Norms

8.1 Standards

This document compiles some of the recommendations of a set of international standards in the field of artificial intelligence, following a standards-based approach. Some of these standards have been published, while others are in the process of being developed, as documented in the publication of the Joint Research Centre (JRC), the science and knowledge service of the European Commission, entitled "*AI Watch: AI Standardisation Landscape. State of play and link to the EC proposal for an AI regulatory framework*".

The Normative Standards, which include the contents of this document are:

- ISO/IEC 38507, Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations
- ISO/IEC DIS 42001, Information technology – Artificial intelligence – Management system.
- ISO/IEC AWI TS 8200, Information technology – Artificial intelligence – Controllability of automated artificial intelligence systems
- ISO/IEC CD TR 5469, Artificial intelligence – Functional safety and AI systems
- ISO/IEC AWI TS 6254, Information technology – Artificial intelligence – Objectives and approaches for explainability of ML models and AI systems
- ISO/IEC AWI 12792, Information technology – Artificial intelligence – Transparency taxonomy of AI systems
- ISO/IEC TR 24027:2021, Information technology – Artificial intelligence (AI) – Bias in AI systems and AI aided decision making
- ISO/IEC AWI TS 12791, Information technology – Artificial intelligence – Treatment of unwanted bias in classification and regression machine learning tasks
- ISO/IEC TR 24028:2020, Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence
- ISO/IEC DIS 5338, Information technology – Artificial intelligence – AI system life cycle processes
- ISO/IEC DIS 5339, Information technology – Artificial intelligence – Guidance for AI applications.
- prEN 18229-1 AI Trustworthiness Framework – Part 1: Logging, Transparency and Human Oversight (In progress).

Financiado por
la Unión Europea
NextGenerationEU