Sandbox IA

EXIT REPORT

SPANISH ARTIFICIAL INTELLIGENCE SANDBOX PILOT

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

ENIA ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL

España | digital 2026

AUTHORSHIP: This report was prepared as part of the first pilot of the Spanish Artificial Intelligence Sandbox, led by by the Secretariat of State for Digitalisation and Artificial Intelligence.

REPORT TITLE: Results Report. First Pilot of the Spanish Artificial Intelligence Sandbox

PUBLICATION DATE: September 2026

VERSION: 1

RESPONSIBLE BODY: Directorate-General for Artificial Intelligence (Secretariat of State for Digitalisation and Artificial Intelligence)

NIPO: 230260497

CONTACT: For enquiries regarding authorship, licensing, rights of use or reuse of the contents of this report:

Ministry for Digital Transformation and the Civil Service

Executive summary

The first pilot of the Spanish AI Sandbox has been a pioneering initiative at the European level, putting into practice the obligations established by the Artificial Intelligence Act (EU Regulation 2024/1689) for high-risk AI systems. Beyond serving as an exercise in regulatory validation, the project has demonstrated how real organisations, with different sizes, operating across a range of sectors and with systems with a high degree of technological maturity, address the challenges of adapting to a complex regulatory framework that is still under development.

The initiative, designed, organised and led by the Secretary of State for Digitalisation and Artificial Intelligence (SEDIA), was conceived with a dual purpose. On one hand, it aimed to support companies providing high-risk AI systems in understanding and implementing the requirements established by the AI Act, using the previously developed technical guides as a reference. On the other hand, it sought to generate practical knowledge that would improve these guides, support the work of future supervisory authorities and prepare the Spanish ecosystem for the effective implementation of the AI Act.

For this purpose, twelve high-risk artificial intelligence systems belonging to particularly relevant sectors such as medical devices, biometrics, employment, access to services, machinery and critical infrastructure were selected, with eleven of them ultimately completing the pilot. Over more than a year, the entities participated in a structured process of development, implementation and review, which has enabled the validation and improvement of the technical guides, the identification of difficulties in interpreting them and opportunities for improvement, which were incorporated into the final versions of the documents. As a result, Spain has developed and made available to the public a set of practical tools and methodologies that facilitate organisations' adaptation to future regulatory obligations.

The analysis of the work carried out by the participating organizations shows that the main challenge has not been so much understanding the obligations established by the Regulation as translating them into internal procedures, documentary evidence and effective governance mechanisms.

Among the most relevant aspects identified during the pilot, the following should be highlighted:

- The adaptation of systems to technical requirements, in accordance with technical guides, has been influenced more by the sector to which these systems belong than by the size of the providers.
- The effort required to achieve the expected levels of compliance has been more related to the complexity of the AI system and the organization's level of maturity than its size or operational capacity.
- Most of the participating entities have managed to implement and document a significant proportion of the measures set out in their adaptation plans.
- It is essential that post-market monitoring be effectively integrated into companies' operational and organisational processes, going beyond the mere conceptual design of monitoring mechanisms.
- It has become clear how important it is to have collaborative models in which the public administration, in its role as regulator, businesses, independent experts, specialised mentoring teams and market surveillance authorities all contribute to creating an environment of shared learning.

As a general conclusion, this pilot has shown that regulatory sandboxes are an effective tool for reducing regulatory uncertainty, accelerating organizations' preparation for compliance and generating capabilities, tools and applicable knowledge for both companies and public authorities. In addition, this experience, a pioneering initiative in the European Union, has served as a reference for other Member States and has contributed to consolidating Spain's position as one of the most advanced countries in the practical application and governance of ethical and trustworthy artificial intelligence.

TABLE OF CONTENTS

IMAGE TABLE

1. Introduction to the *Spanish AI* Sandbox Pilot

This document presents, in a structured and aggregated way, the main results, learnings and conclusions derived from the development of the first pilot of the Spanish AI *Sandbox*. This exit *report* has been prepared as part of the process of closing the initiative in order to consolidate the knowledge generated during the pilot and make it available to the artificial intelligence ecosystem, the competent authorities and future regulatory support initiatives.

The purpose of the report is not only to describe the actions carried out, but also to systematise the practical evidence obtained through the application of the technical guides developed to facilitate the understanding of the requirements of Regulation (EU) 2024/1689 for high-risk artificial intelligence systems. Based on the experience of the participating entities, the document identifies common challenges, good practices, needs for improvement and methodological elements that may be useful for the design and deployment of future controlled test environments.

In this sense, this *exit report* constitutes both a closing document and an instrument for knowledge transfer. On one hand, it presents the institutional and methodological configuration of the pilot, the actors involved and the activities carried out. On the other hand, it analyses the results achieved and formulates conclusions aimed at strengthening the preparation of the Spanish ecosystem for the progressive application of the European regulatory framework for artificial intelligence.

To address these objectives, the document is structured progressively. The first section, **Origin and ,** presents the context in which the initiative emerged, its motivation, its objectives and its institutional configuration, as well as the main elements that determined its implementation.

The second section, **Methodology and Key Milestones**, describes the framework within which the pilot has been implemented, including the technical guides developed for the AI Act, the governance structure and the ecosystem of participating actors. Likewise, the different phases of the *Sandbox*, the methodology used and the main milestones achieved during its development are described systematically.

On this basis, the third section, **Analysis of results,** analyses the results obtained throughout the pilot, the lessons learned from each of its phases, the operational aspects identified during its implementation and the estimation of the effort required by the participating entities to implement the indicative measures.

Finally, the document presents the **main conclusions and lessons learned,** in the section **Lessons learned and conclusions**, which constitute a practical knowledge base to guide future supervisory

support, support and development activities for controlled test environments in the field of artificial intelligence.

Origin and Context

Prior to the approval of **Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024** (hereinafter, the "AI Act ") establishing harmonised standards in the field of artificial intelligence, Spain, through the Secretary of State for Digitalisation and Artificial Intelligence (hereinafter, SEDIA), identified the need to prepare the Spanish artificial intelligence ecosystem to meet the future regulatory requirements established by European legislation.

For this purpose, SEDIA promoted a **pioneering initiative** aimed at testing, in a controlled and supervised environment, the practical application of the requirements established for high-risk artificial intelligence systems. Its implementation was set out in **Royal Decree 817/2023**, of 8 November, which established a controlled testing environment to test compliance with the proposed Regulation. This AI Act gave rise to the first regulatory *sandbox* for artificial intelligence in Europe.

The objective of the pilot was to **validate the practical application of the technical guides developed by SEDIA,** conceived as indicative and non-binding instruments. These guides provide recommendations, reference measures, interpretative criteria and support tools to enable providers, deployers and other relevant actors to analyse their degree of readiness, identify compliance gaps and plan adaptation measures in a structured way. At the same time, the pilot aims to generate empirical evidence and documented technical-regulatory knowledge based on the lessons learned, with the aim of improving the guides themselves, transferring this knowledge to the Spanish artificial intelligence ecosystem and facilitating its preparation for the future application of the obligations established in the AI Act.

The knowledge generated had to be useful not only for providers, deployers and other actors involved in the life cycle of artificial intelligence systems but also for the authorities and bodies called upon to perform conformity assessment, market surveillance and regulatory supervision functions. All this is part of the need to ensure that AI systems marketed and used in the European Union are safe, respect fundamental rights and promote responsible innovation.

With the approval and publication of the AI Act on 12 July 2024 and its entry into force on 1 August 2024, the European Union moved towards the creation of the **first comprehensive regulatory framework for artificial intelligence based on a risk management approach.** The AI Act establishes

a system of obligations of progressive application for providers, deployers and other actors involved in the life cycle of artificial intelligence systems.

In this framework, Article 57, on AI regulatory sandboxes, establishes that Member States must ensure the creation of at least one AI regulatory ***sandbox* at national level**, fully operational, by 2 August 2027 at the latest.

In general, a *sandbox* is an isolated, controlled environment that allows users to test, experiment with, or validate a solution without affecting the actual environment in which it could later be deployed. Its purpose is to reduce the uncertainty associated with the introduction of new technologies, products or processes, allowing their operation to be observed, risks to be identified and adjustments to be made before their final implementation.

In the regulatory context, the pilot of the *Spanish AI Sandbox* was configured as a practical experience aligned with the principles set out in Article 57 of the AI Act, under the supervision of the competent authority. In this environment, participating entities were able to analyse and implement the requirements applicable to their systems before the full implementation of the European framework.

The pilot was developed in close coordination with the **European Commission** and the **European Office for Artificial Intelligence (AI Office)** and paid special attention to small and medium-sized enterprises (hereinafter referred to as **SMEs**) and ***startups***, in order to reduce the barriers to adaptation to the new regulatory environment.

The project officially began with the *kick-off* session on 29 April 2025, at the Rio headquarters of the Ministry for Digital Transformation and Public Administration.

The Sandbox was operated using a methodology structured in five phases, from the initial training activities and understanding of the technical guides to the implementation of measures, the self-assessment of the systems and the post-market monitoring activities.

On 11 June 2026, the closing ceremony [1]of the first pilot of the *Spanish AI* Sandbox[2] was held, with the participation of institutional representatives, participating entities, competent authorities, members of the expert advisory group and other stakeholders in the artificial intelligence ecosystem.

As part of the closure process, each participating system received an ***individual* exit report**, conceived as a feedback report that sets out the work carried out during the pilot, the actions carried out in each of the phases, the degree of progress achieved with respect to the measures analysed, as well as the main strengths, challenges and recommendations identified during its participation.

These reports have provided each entity with a structured view of its experience in the *Sandbox* and the opportunities for improvement identified to support its adaptation to the European regulatory framework.

The delivery of the individual reports and the publication of this ***exit report* at a global level**, which brings together the results, learnings and conclusions obtained throughout the project, mark the completion of the first pilot of the Spanish AI *Sandbox* and consolidate a practical and methodological knowledge base for future initiatives supporting regulatory compliance and the supervision of artificial intelligence systems.

As a main result, the pilot has made it possible to develop a set of **updated technical guides, methodologies and practical tools** aimed at facilitating the early adaptation of companies, especially SMEs and *startups*, to the new regulatory framework.

Likewise, the experience has contributed to positioning **Spain as one of the most advanced Member States in preparing for the application of the AI Act,** both through the design of the controlled testing environment and through the generation of useful knowledge for other Member States. The supporting documents developed as part of the pilot, including the technical guides, have been shared with other Member States interested in the experience and lessons learned from the *Spanish AI* Sandbox[3].

[1] Closing ceremony of the AI Sandbox. https://www.youtube.com/watch?v=oPdmIpuY86c

[2] The Government successfully concludes Europe's first artificial intelligence Sandbox. https://digital.gob.es/comunicacion/notas-prensa/secretaria-digitalizacion-e-inteligencia-artificial/2026/06/el-gobierno-concluye-con-exito-el-primer-sandbox-de-inteligencia

[3] First regulatory sandbox on Artificial Intelligence presented. https://digital-strategy.ec.europa.eu/en/news/first-regulatory-sandbox-artificial-intelligence-presented

In parallel, Spain continues to develop the institutional framework necessary for the effective implementation of the AI Act. Among the most relevant initiatives is **the Bill for the proper use and governance of artificial intelligence**, aimed at complementing the European framework by defining the national regime of supervision, administrative coordination and sanctioning power.

The following section briefly describes the actors involved in the development of the Spanish AI Sandbox.

Ecosystem

The development of this first pilot has been based on an ecosystem of **public and private actors** who, through their different roles and responsibilities, have contributed to the improvement and practical validation of the indicative measures set out in the technical guides. This collaborative model has made it possible to integrate technical, regulatory and supervisory expertise in a controlled testing environment, fostering a shared learning process between the different parties involved.

Together with the governance structure, described in more detail in the **Roles and governance**, the pilot involved different key actors:

- Entities providing high-risk artificial intelligence systems,
- The Advisory Group of Experts,
- and Market Surveillance Authorities (MSAs)

The selection of the participating entities was carried out through the **public call** approved by the **SEDIA Resolution, of December 20, 2024**, which opened applications for access to the controlled testing environment for reliable artificial intelligence, in accordance with the provisions of Royal Decree 817/2023.

Through this procedure, a representative sample of high-risk artificial intelligence system providers from the Spanish ecosystem was selected, based on criteria such as the maturity of the solutions, their innovative nature and the diversity of application sectors. The results of this selection process and the entities that ultimately participated are presented in the subsection **Participating Entities .**

On the other hand, the **Advisory Group of Experts**, comprising of 40 multidisciplinary professionals, was constituted by the **SEDIA Resolution, of 8 June 2025**, in compliance with the provisions of Article 27 of Royal Decree 817/2023.

This advisory body has provided independent technical, legal and scientific expertise during the development of the pilot, contributing to the interpretation of the requirements of the AI Act, the analysis of use cases and the improvement of the technical guides. Its composition and functions are described in the subsection **Advisory Group of Experts.**

Finally, this ecosystem also included the participation of the Market Surveillance Authorities (MSAs), whose intervention throughout several of the phases of the AI *Sandbox* was intended to incorporate the perspective of the future competent authorities in charge of supervising compliance with the obligations established in the AI Act. From a regulatory and supervisory perspective, the MSAs were able to analyse the evaluation mechanisms and documentation generated by the participating entities, providing a practical perspective on the future surveillance of high-risk artificial intelligence systems and contributing to the learning process within the controlled test environment and the improvement of technical guides. Their specific role is described in the **Roles and Governance** section.

Overall, the European collaboration fostered throughout **the Sandbox has helped position Spain as one of the leading countries in the practical implementation of the AI Act**. It has also enabled the lessons learned from the pilot to inform European discussions on the design, implementation, and future development of regulatory sandboxes for artificial intelligence.

2. Methodology and Key Milestones

The execution of the pilot of the *Spanish AI Sandbox* was articulated through a structured methodology aimed at both **improving the technical guides** prepared by SEDIA and generating **practical knowledge on the application of the requirements of the AI Act** by the different actors involved in its compliance, evaluation and supervision.

For this purpose, a progressive process of support for the participating entities was established, combining training sessions, self-diagnosis exercises, the preparation and implementation of adaptation plans, self-assessment of progress and analysis of post-market monitoring mechanisms.

This approach made it possible to observe in a structured way how **organizations of different sizes, sectors and levels of maturity approached the interpretation and application of the technical guides**. It also facilitated the identification of practical difficulties, areas requiring clarification, technical and organisational barriers, as well as good practices that could be incorporated into future monitoring and supervisory tools.

The experience gained during the development of the pilot was equally valuable in understanding the needs of the various actors who will be involved in the effective implementation of the AI Act. The lessons learned made it possible to draw conclusions not only about the adaptation processes of AI system providers, but also about the evaluation, supervision and institutional coordination mechanisms that will need to be deployed by the competent authorities and other bodies involved in the future AI governance ecosystem.

From this perspective, the *Sandbox* was configured as an iterative and incremental process in which each phase was supported by the results of the previous one, making it possible to progressively adjust the monitoring instruments and enhance the knowledge generated throughout the project.

The following sections describe the main methodological elements and key milestones that made the development of the pilot possible, including its governance model, the different phases of implementation, the mechanisms for managing regulatory and technical consultations, the publication of the technical guides and the activities aimed at their review, validation and improvement in light of the lessons learned during the *Sandbox*, the contribution of the advisory group of experts, the conduct of satisfaction surveys on the most relevant aspects of the pilot among the participating entities, and the national and European collaboration initiatives developed during the project.

Technical Guides

One of the main activities of the AI *Sandbox* has been the validation, updating and publication of a set of indicative and non-binding technical guides aimed at facilitating the interpretation and application of the requirements established in the AI Act.

These guides were conceived as **practical support tools** for providers and other actors involved in the life cycle of artificial intelligence systems, especially during the transitional period **prior to the adoption of the European harmonised standards** that will develop certain technical aspects of the AI Act.

Throughout the pilot, the practical application of the guides by the participating entities made it possible to assess the feasibility of the proposed measures, identify difficulties in interpretation and detect needs for clarification or further guidance. As a result, the guides were subject to a continuous **process of revision and improvement** that incorporated the knowledge generated during all phases of the *Sandbox*.

In addition, the updates to the guides drew on the contributions made by the participating entities, the queries resolved during the pilot, the advisory group of experts and the Market Surveillance Authorities. As a result, the published versions incorporate the main lessons learned during the project and present a greater degree of clarity, practical applicability and alignment with the technical and regulatory reality of the systems analyzed.

In order to ensure the quality and alignment of these documents with the future European standardisation framework, the guides were also subject to an extensive revision process, including their **adaptation to the most recent versions of the draft of the European standards for artificial intelligence**.

As a result of this process, in December 2025 the Spanish Agency for the Supervision of Artificial Intelligence (AESIA)[4] published a set **of sixteen supporting documents**, accompanied by a compendium of ***checklists*** aimed at facilitating the self-diagnosis of the degree of compliance with the proposed measures. Although these documents are not binding and do not replace the applicable rules, they provide practical recommendations aligned with the requirements of the AI Act and serve as a useful reference for the adoption of future harmonised standards.

[4] AESIA Portal. Technical guides. https://aesia.digital.gob.es/es/guias

The documentary set is structured in three complementary blocks:

- **Two introductory guides**, aimed at providing an overview of the AI Act and facilitating its understanding through explanations, examples and informative material.
- **Thirteen technical guides**, focusing on the main obligations associated with the development, placing on the market, deployment and supervision of high-risk AI systems, aimed mainly at providers and, where appropriate, deployers and other actors involved in their system life cycle. These guides address, among other areas, conformity assessment, quality, risk management, data governance, human oversight, transparency, accuracy, robustness, cybersecurity, automatic records, post-market monitoring, serious incident reporting, and technical documentation.
- **A manual for the use of *checklists***, designed to facilitate the systematic self-assessment of the degree of compliance and support the implementation of governance mechanisms and quality procedures in organisations.

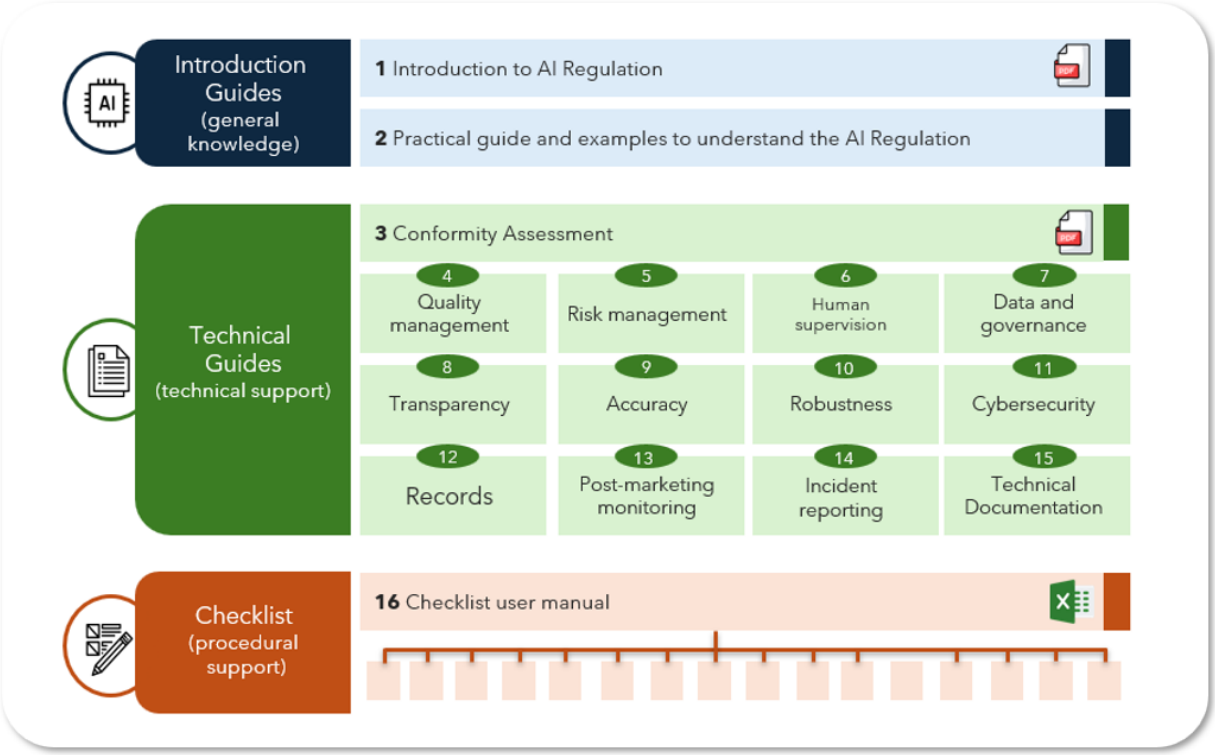

Image 1 - Technical Guides and Checklists prepared as part of the Spanish AI Sandbox pilot

The modular organisation of the materials, illustrated in **Image 1 - Technical Guides and Checklists prepared as part of the Spanish AI Sandbox pilot**, allows their progressive use by organisations with different levels of maturity in relation to the implementation of the AI Act, combining informative content, technical guidance and practical self-diagnosis tools.

With their publication, the aim was to provide Spanish organisations with a greater degree of legal and technical certainty during the transitional period before the full application of the AI Act, facilitate early adaptation to its requirements and to contribute to improving the preparation of the national ecosystem for the development and deployment of trustworthy artificial intelligence systems aligned with the European regulatory framework.

Overall, the development and evolution of these guides constitute one of the main results of the Spanish AI *Sandbox*. The knowledge generated through their execution, the resolution of queries, the interaction with the competent authorities and the practical experience accumulated by the participating entities became one of the main inputs into their review and continuous improvement. As a result, **the guides have served as a support tool during the development of the pilot and represent a practical knowledge base to facilitate the future application of the AI Act and support new controlled test environments** in Spain and across Europe.

Roles and Governance

The pilot of the *Spanish AI Sandbox* was articulated through a **collaborative governance model** that integrated the Public Administration, the competent authorities, the participating entities and different expert profiles. This model constituted the organisational structure on which the pilot was developed, defining responsibilities, coordination channels and monitoring mechanisms that made it possible to guarantee a coherent and orderly execution of the different planned activities.

The participating entities were mainly made up of **providers** of high-risk artificial intelligence system. In many cases, these providers also had the support of collaborating companies and external specialists who provided specific technical, regulatory or sectoral knowledge throughout the project and who, on certain occasions, also acted as deployers of the systems analysed. This collaboration favoured the formation of **multidisciplinary teams** capable of jointly addressing the challenges associated with the development, implementation, evaluation and regulatory adaptation of the different use cases.

Without prejudice to the internal organisational structure of each participating entity, the *Sandbox* articulated a common scheme of operational roles aimed at facilitating the coordination of the

project and the traceability of the activities carried out. This scheme was not intended to replace the existing organisational models, but to identify the minimum functions necessary to ensure effective dialogue between the participating entities and the teams responsible for the implementation of the *Sandbox*. Within this framework, the following reference functions were defined:

- A **project manager (PM):** responsible for coordinating the activities carried out within the framework of the *Sandbox*, ensuring dialogue with the support teams and ensuring the alignment of the work carried out with the objectives of the project.
- A **technical team**: responsible for implementing the technical requirements of the Sandbox.

For its part, **SEDIA assumed the institutional direction and coordination** of the project, relying on a structure composed of various teams with complementary functions:

- **Mentoring team**, acting as **the main point of contact with the participating entities**, providing methodological support, technical guidance and support in the interpretation of the technical guides. It also played a relevant role in the identification of opportunities for improvement and in the systematization of the lessons learned during the development of the pilot.
- **An external technical team**, which collaborated with the mentoring team in the execution of the planned activities, the monitoring of the participating entities and the **resolution of first-level queries**. In addition, it played a fundamental role by providing expert knowledge of the technical guides, supporting their consistent interpretation, identifying opportunities for improvement and contributing to the incorporation of the learnings obtained during the pilot.
- **An advisory group of experts,** whose participation contributed both to the **analysis of complex issues** raised by the participating entities and to the **validation and improvement of the methodologies and tools** developed within the framework of the *Sandbox*.

In addition to these actors, the **Market Surveillance Authorities (MSAs)** represented in **Table 1 – List of Participating MSAs** also participated in the Sandbox as key institutional actors for assessing, from a learning perspective, the practical applicability of the technical guides and the compliance mechanisms established for high-risk artificial intelligence systems.

	Spanish Data Protection Agency (AEPD)

Market Surveillance Authorities (MSAs)	Spanish Agency for the Supervision of Artificial Intelligence (AESIA)
	State Aviation Safety Agency (AESA)
	Bank of Spain (BdE)

Table 1 - List of participating MSAs

Its intervention **made it possible to incorporate into the controlled test environment a perspective closely aligned with future market supervision**, particularly relevant in a context in which the AI Act attributes to the competent authorities functions of surveillance, control and application over AI systems placed on the market or put into service.

Specifically, its participation in phase IV (Self-assessment) and phase V (post-market monitoring) of the *Sandbox* was of particular relevance, as these phases focused on the self-assessment of the measures implemented and on post-market monitoring. The respective MSAs contributed their knowledge and had the opportunity to review first-hand the documentation prepared by the participating entities, as well as the approaches used to justify compliance with the guidance measures. They also analysed the practical difficulties related to the generation of the relevant documentation, the traceability of decisions, risk management and the continuous monitoring of the systems once deployed.

Their collaboration was particularly useful for identifying **sectoral particularities** among the participating entities and for assessing how the requirements of the AI Act can interact with other applicable regulatory frameworks. This approach made it possible to enhance the analysis of the Sandbox using supervisory criteria more closely aligned with those that will need to be applied in

real-market scenarios, helping to anticipate potential regulatory overlaps, the need for inter-agency coordination and consistent interpretation criteria.

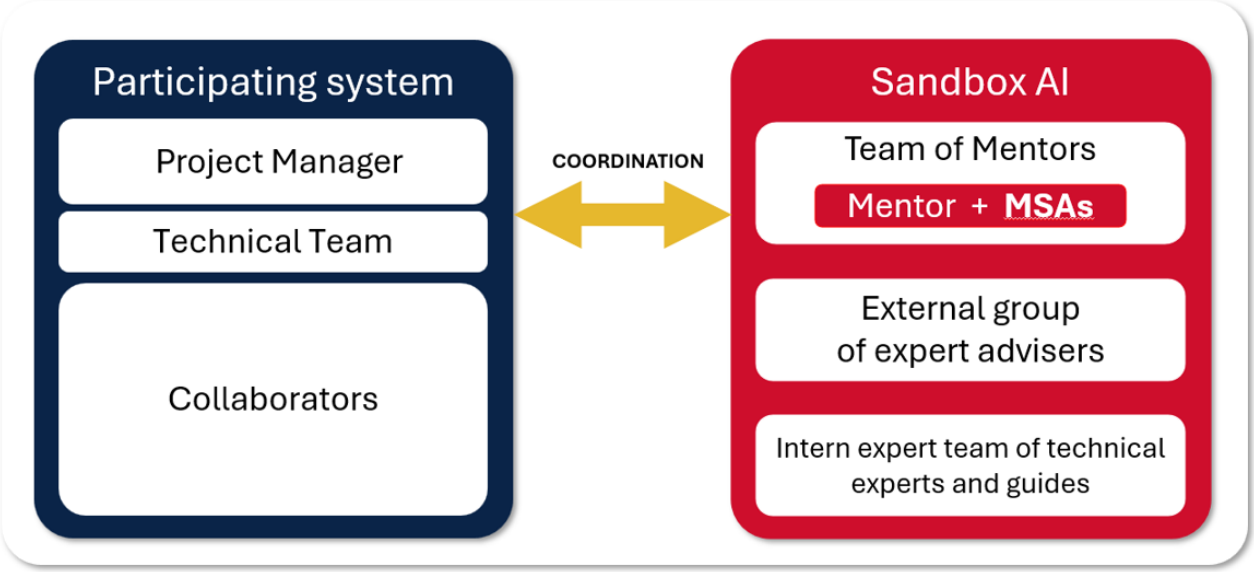

Image 2 - Roles and Governance of the first Spanish AI Sandbox pilot

Overall, the interaction between all these actors was articulated through a model of continuous collaboration, as mentioned above, in which **the mentoring team** played a key role as the link between the participating entities and the Public Administration. This mechanism favoured fluid and two-way communication, allowing technical queries, interpretative issues and difficulties identified during the implementation of the measures set out in the technical guides to be communicated effectively. It also contributed to capturing the lessons learned throughout the pilot, facilitating both the resolution of incidents and the progressive improvement of the guides, methodologies and tools developed within the framework of the *Sandbox*.

Image 2 - Roles and Governance of the first Spanish AI Sandbox pilot schematically illustrates the governance model adopted during the pilot, reflecting the roles and responsibilities of the different actors involved, as well as the collaborative relationships established between the participating entities, support teams and the authorities involved.

In short, this governance model lays a solid foundation for the design and future consolidation of the controlled testing environments provided for in the AI Act, by providing useful knowledge both for the organisations subject to its obligations and for the authorities and bodies that will have to carry out functions of assessment, supervision and regulatory support. In particular, for SMEs and start-

ups, this approach facilitated an early approach to future regulatory requirements, while providing the Public Administration with valuable practical experience for the development of future AI governance and oversight mechanisms.

Participating Entities

Participation was articulated through a public call, approved by the **SEDIA Resolution** of **20 December 2024**, for the purpose of selecting a representative sample of high-risk artificial intelligence systems from the Spanish AI ecosystem, and supporting them in validating, in a controlled environment, of the measures set out in the technical guides.

The application period remained open between 23 December 2024 and 23 January 2025, during which **41 proposals** were received, of which **twelve high-risk artificial intelligence systems** were finally selected.

The call was aimed at entities that had solutions already marketed or put into service, or that were sufficiently mature enough to reach that stage during the development of the Sandbox or immediately after its completion. Likewise, **the participation of SMEs and startups was expressly promoted**, with the purpose of facilitating their adaptation to the future regulatory framework and generating practical knowledge on the application of the technical guides in organizations with different levels of technological maturity and organizational capabilities.

During the evaluation process, various aspects were assessed, including the innovation or technological complexity of the system, its development maturity, its classification as high-risk artificial intelligence systems according to the AI Act or sectoral diversity were assessed, among other aspects. The selection aimed **to configure a heterogeneous sample of real use cases** that would allow the application of the applicable regulations in different areas of activity to be analysed and generate practical evidence to update the technical guides and draw conclusions on supervision methodologies and good regulatory compliance practices. The diversity of the selected entities has made it possible to analyse development and deployment scenarios with different degrees of technical and organisational complexity, as well as to support the development of recommendations applicable to the Spanish artificial intelligence ecosystem as a whole.

The selected entities, as indicated in the Resolution, are shown in the following **Table 2 - Entities participating in the AI Sandbox pilot**.

Medical	Bit&Brain technologies, S.L
	Tucuvi Care SL
	Made Of Genes, S.L.U
Access to Benefits and Services	Foundation Center for Research in Information and Communication Technologies of Galicia (FCITICG)
	Dedomena Artificial Intelligence, S.L
Machinery and	Airbus Operations SL
	Imotion Analytics, SL
	Loradix Red, S.L
Biometrics	Veridas Digital Authentication Solutions, S.L.
	Herta Security, S.L.
Employment	Shakers Global Solutions SL
	Adevinta Jobs, S.L.U.

Table 2 - Entities participating in the AI Sandbox pilot

It should be noted that the participating entity Loradix Red, S.L. voluntarily withdrew from the controlled testing environment during the course of the pilot, as communicated to SEDIA on 13 February 2026. The entity stated that it was unable to continue participating due to a lack of financial and human resources, circumstances beyond the control of the pilot itself. By **Resolution of 7 March 2026**, SEDIA accepted this voluntary withdrawal on the basis that the reasons given adequately justified the entity's withdrawal from the initiative before its completion. Consequently, the pilot concluded with the effective participation of eleven high-risk artificial intelligence systems.

Advisory Group of Experts

As part of the governance structure provided for in Article 27 of Royal Decree 817/2023**, the advisory panel of experts** was appointed by means of SEDIA Resolution of 8 June 2025.

The objective was to form a group of independent professionals of recognized prestige and extensive technical experience capable of contributing a multidisciplinary perspective to the development of the *AI Sandbox*. In accordance with the provisions of Royal Decree 817/2023, this group was to be made up of experts from academia, research, the business sector, professional colleges and associations and other **fields related to artificial intelligence and its technical, legal, ethical and**

social implications. Likewise, the AI Act established that the principle of balanced representation of women and men should be taken into account in its appointment.

Prior to the formal constitution of its members, an open consultation was **held for experts interested** in collaborating in the development of the AI *Sandbox*. The purpose of this consultation was to identify professionals with relevant experience who would be willing to contribute, independently, to the resolution of high-level consultations and to the development of practical knowledge on the application of the AI Act. In particular, profiles were sought capable of advising on the interpretation and implementation of the requirements applicable to high-risk AI systems, providing recommendations for the improvement of the technical guides and sharing their experience during the different activities planned within the framework of the *Sandbox*, in line with the functions attributed to the advisory group of experts by Royal Decree 817/2023.

This public consultation formed the basis for the **selection of the 40 experts** who made up the Advisory Group of Experts in the first *Spanish AI* Sandbox pilot.

Operational Phases

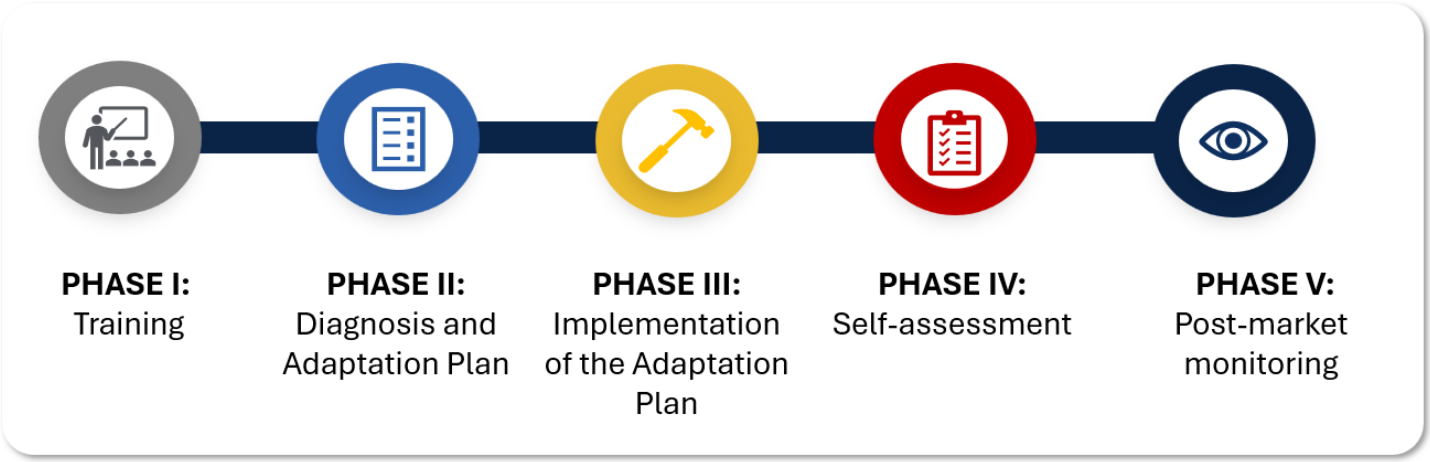

Image 3- Phases of the first Spanish AI Sandbox pilot

As noted above, the Spanish AI *Sandbox* began with the kick-off held on 29 April, 2025 and ran until its closure in June 2026. The initiative was structured in five successive phases, following an iterative and incremental methodology, as shown in **Image 3- Phases of the first Spanish AI Sandbox pilot .**

Phase I – Training

The first phase was developed during the month of May 2025 and was aimed at providing the participating entities with a common knowledge base on the technical guides prepared by SEDIA. Its objective was to ensure a homogeneous understanding of the main concepts, obligations and interpretative criteria, so that all organizations could face the subsequent phases with an adequate level of preparation. This stage was instrumental in ensuring a shared understanding of the regulatory obligations and the *Sandbox methodology*.

For this purpose, **specialised training sessions were delivered, and various support materials were made available to the organisations**, including a training plan, a package of guides in PDF format, presentations prepared for the training sessions and specific thematic documents. All this content was complemented by the recordings of the sessions, which were made available as internal reference material to reinforce learning.

The phase also incorporated monitoring and feedback mechanisms. Participants completed an assessment survey through the FORMA web form, which allowed the collection of useful information for the improvement of the sessions and the extraction of key indicators for the final report of conclusions. In addition, a Frequently Asked Questions (FAQs) document was continuously updated in order to resolve operational and content-related queries.

To ensure confidentiality and the proper organization of the documentation, all resources were centralized in dedicated SharePoint work folders.

Phase II – Self-Diagnosis and Adaptation Plan

The second phase was carried out during the months of June and July 2025 and was intended ro enable each entity to carry out a **self-diagnosis of the degree of alignment of its system** with the indicative measures included in the technical guides. To do this, the materials developed specifically for the Sandbox were used, including the ***verification* checklists** associated with each of the guides.

This exercise made it possible **to determine the initial situation of each system** with respect to the requirements and measures contemplated in the technical guides supporting the AI Act and

constituted the starting point for defining the necessary adaptation actions. As a result of this phase, **each entity drew up an adaptation plan**, identifying the measures to be implemented, the priorities for action, the resources required and the estimated effort to move towards an adequate level of compliance in accordance with the technical guides.

This process allowed each entity to have an individualized roadmap to address the next phase of the *Sandbox* and, at the same time, provided relevant information on the practical applicability, complexity and effort associated with the implementation of the measures contained in the technical guides.

The main steps developed during this phase are detailed below:

- **Kick-off meeting and materials:** meeting between members of the mentoring team and *project managers* to present the objectives, tools, roles, deliverables and provide an introduction to the verification checklists.
- **Initial self-diagnosis:** the entities completed the 12 checklists associated with the different technical guides for high-risk AI systems, identifying the areas in which there was a greater need for adaptation or effort.
- **Provision of additional measures:** possibility of incorporating complementary measures not initially contemplated in the guides. This activity made it possible to collect valuable information on the specific needs of the entities and to enrich the content of the technical guides themselves.
- **Design of the adaptation plan:** preparation of an individualised roadmap based on the results of the self-diagnosis, identifying the measures to be addressed, the necessary resources and the estimated timeframes for their implementation.
- **Periodic monitoring:** technical follow-up meetings to review progress, resolve queries and analyse the preliminary results obtained through the checklists.
- **Specialised technical support:** resolution of questions raised by the participants by the mentoring team, the technical leads responsible for the different guides or the advisory group of experts.

The approval of these adaptation plans by SEDIA marked the transition from an eminently diagnostic and planning phase to a stage focused on the effective implementation of the measures identified, which constitutes the practical core of the adaptation process developed in the *Sandbox*.

Phase III – Implementation of the Adaptation Plan

The third phase took place between August and December 2025 and was dedicated to the implementation of the measures included in the corresponding adaptation plans. During this period, the entities worked on the **effective incorporation of the actions identified in the previous phase** with the aim of advancing in the alignment of their systems with the recommendations contained in the technical guides.

This was also the most demanding phase of the *Sandbox* in terms of dedication, workload and resources required by the participating entities. Although a shorter duration had initially been planned, various factors conditioned its implementation and made it necessary to extend the schedule. Among them, the technical complexity of certain measures, the limited availability of human and technical resources, especially during the summer period, and other organizational circumstances specific to some participating entities were particularly significant.

The development of this phase made it possible to assess in a real environment the feasibility of the measures proposed in the technical guides, identify the main implementation difficulties and obtain valuable information on the times, capacities and resources necessary to undertake regulatory adaptation processes in organizations of different nature, size and level of maturity.

For this purpose, the following activities were carried out:

- **Follow-up meetings:** periodic meetings, usually every one or two weeks, to monitor the evolution of the agreed measures and to refer, where appropriate, issues to the advisory group of experts.
- **Generation of a *follow-up* checklist**: an Excel based monitoring tool developed to monitor the measures selected by each entity between the different work sessions. This tool included:
  - The measures selected for implementation.
  - The effort perceived and effort actually devoted. The objective was to obtain *feedback* on the initial perception of the effort required and the outcome after implementation.
  - The degree of progress in the technical documentation associated with the measures developed.
- **Practical implementation of the measures:** work carried out by each entity with a variable effort and the participation of different technical, legal and organizational profiles.

The completion of this phase was a significant milestone in the pilot, as it allowed the entities to transform the results obtained during the self-diagnosis into specific adaptation actions and generate the documentation that would later be reviewed during the self-assessment phase.

Phase IV – Self-Assessment

The fourth phase took place between January and March 2026 and aimed **to assess the degree of implementation achieved** by the participating entities with respect to the measures included in their respective adaptation plans. This phase also made it possible to analyse the ability of organisations to demonstrate this level of implementation through verifiable documentation.

Each entity documented the degree of progress achieved in the implementation of the planned measures and compiled the documentation associated with its adaptation process. This documentation was subsequently reviewed by the technical teams responsible for the guides and by SEDIA in order to assess the adequacy of the measures implemented, the sufficiency of the documentation provided, and the level of maturity achieved by each system.

For their part, the MSAs were able to participate in the review of the results obtained, providing their perspective on the documentation generated, the mechanisms for justifying compliance and the practical challenges associated with the future application of the AI Act.

The phase was structured in three sequential stages:

- **Submission of documentation:** the entities collected and submitted the documentation associated with the measures implemented during the previous phase.
- **Review and evaluation process:** the evaluation teams analysed the documentation submitted in order to assess the degree of implementation of the measures, the quality of the documentation provided, and the level of maturity achieved by each system.
- **Issuance of self-assessment reports**: as a result of the review process, each entity received an individualized report of conclusions prepared by SEDIA in collaboration with the corresponding MSA, setting out the main conclusions of the self-assessment carried out, including the strengths identified, the areas for improvement detected and a series of recommendations aimed at reinforcing the level of adequacy and maturity of the system with respect to the measures analysed.

This organization made it possible to guarantee coordination between the participating entities, SEDIA, the evaluation teams and the MSAs, ensuring the obtaining of comparable, consistent and reproducible results across the different systems analysed.

The completion of this phase made it possible to have a structured view of the level of adequacy achieved by each system with respect to the technical guides, as well as to identify those measures that required additional actions before moving on to the post-market monitoring phase.

Phase V – Post-Market Monitoring

The fifth and final phase took place between April and May 2026 and was aimed at designing, **reviewing or improving its post-market monitoring mechanisms**. The objective was to transfer the continuous supervision approach provided for in the AI Act to the specific context of the participating systems, promoting the implementation of processes that would allow their behaviour to be monitored once they were put into service and enable the early detection of possible incidents, deviations or emerging risks at an early stage.

Unlike the previous phases, which focused mainly on the initial adaptation of the systems to the measures contained in the technical guides, this stage addressed a continuous dimension of regulatory compliance. In this sense, it allowed entities to analyse how to maintain over time the guarantees incorporated during the adaptation and implementation phases, reinforcing the capacity of the systems to operate in a safe and controlled manner throughout their life cycle.

In order to adapt the development of the phase to the different realities of the participating entities, the work methodology was structured according to the existing level of maturity in terms of post-market monitoring. To do this, the entities were segmented into two groups:

- **Group A:** entities that already had surveillance procedures or mechanisms in place and sought to optimise or adapt them to the requirements arising from the AI Act.
- **Group B:** entities that lacked a formalised surveillance system and had to design and implement a monitoring model from its initial phases.

The phase was articulated through the following subphases:

- **Subphase I – Initial diagnosis:** refresher sessions and completion of a self-assessment checklist aimed at determining the level of maturity of each entity and its allocation to the corresponding group.
- **Subphase II – Risk analysis:** identification of risks to establish the scenarios that the surveillance plan should monitor.
- **Subphase III – Design and implementation of surveillance measures:** development of differentiated activities according to the level of maturity of each entity. Group A entities reviewed and adapted their existing mechanisms to incorporate new risks and controls, while Group B entities defined indicators, developed their surveillance plans and initiated their implementation.
- **Subphase IV – Internal evaluation:** review of the functioning of the defined mechanisms through the analysis of the established indicators and verification of the operational viability of the surveillance procedures designed.

The completion of this phase allowed the entities to have a first structured approach to the post-market monitoring processes required by the AI Act, laying the foundations for continuous monitoring of the performance of their systems beyond the duration of the *Sandbox*. However, post-market monitoring is a permanent activity that should be maintained throughout the lifecycle of the AI system and does not end with the end of the controlled test environment.

Coordination and Resolution of Queries on Regulatory Matters and Technical Guides

The management of the knowledge generated during the development of the *Sandbox* was a cross-cutting element of the initiative. The practical application of the technical guides by the participating entities made it possible to identify interpretative questions, implementation difficulties and needs for clarification, which also provided valuable information for the continuous improvement of the guides themselves, and the methodologies used during the pilot.

In order to ensure the traceability of queries and the consistency of the responses provided, a structured procedure was established based on three main elements:

- **Periodic follow-up**, through meetings between the mentoring teams.
- **Centralised register of queries**, which made it possible to document the issues raised, ensure their traceability and identify recurring patterns and needs.
- **Expert query resolution and escalation model**, based on a progressive analysis process that combined review by the technical teams responsible for the guides with the participation of the advisory group of experts when the complexity or cross-cutting nature of the consultations required it.

Query Resolution and Escalation Model

The issues raised by the participating entities were initially analysed by the technical team and by the persons responsible for the corresponding guide. Those queries that required a more complex interpretation or that could have implications for a wide range of organisations were referred to the Advisory Group of Experts, made up of multidisciplinary profiles from the technical, legal, regulatory and standardisation fields.

This model made it possible to provide consistent and well-founded responses, promote the homogeneous application of the technical guides among the participating entities and consolidate a repository of practical knowledge generated from the Sandbox experience. Likewise, the information obtained through this process was especially valuable for identifying aspects requiring clarification or improvement in the guides themselves, supporting their progressive evolution as the pilot progressed

Annex I: Relevant Consultations within the Framework of the compiles the most relevant queries raised within the framework of the pilot and, where appropriate, a summary of the response given to the participating entities.

Dissemination and Knowledge Transfer Activities

In addition to its participation in the resolution of queries and technical analyses developed during the *Sandbox*, the Advisory Group of Experts played an important role in the dissemination of knowledge across the artificial intelligence ecosystem. With this objective, various initiatives were promoted as described below.

Together, these activities contributed to broadening the scope of the results obtained during the *Sandbox* beyond the participating entities, promoting the dissemination of practical knowledge on the application of the AI Act and creating opportunities for shared reflection between public administrations, companies, experts and reference organisations. They also made it possible to identify common needs across the ecosystem and strengthen the role of regulatory *sandboxes* as learning, support and knowledge transfer instruments for the implementation of artificial intelligence in Europe.

Meeting "Responsible AI in practice: Advances in Regulatory Sandboxes"

This participatory workshop, held on 23 October 2025, brought together members of the Advisory Group of Experts, representatives of the participating entities, representatives of the Market Surveillance Authorities involved in the *Sandbox* and members of the Public Administrations with the aim of **analysing the practical application of the AI Act and evaluating the role of regulatory** sandboxes as instruments supporting regulatory compliance.

The exchange of experiences made it possible to identify some of the main challenges and opportunities for the development and deployment of artificial intelligence systems in Europe. Participants highlighted the potential of AI to improve productivity, optimize processes, and develop more personalized services. However, they also pointed to relevant challenges for the European ecosystem, including technological dependence on external providers, interoperability difficulties and existing constraints in terms of investment, talent availability and technological leadership capacity.

In relation to the implementation of the obligations provided for in the AI Act, it was highlighted that adaptation efforts can be particularly demanding for SMEs and *startups* due to the technical complexity of certain requirements and the organisational burden associated with their implementation. In this context, the importance of promoting approaches proportionate to the level of risk of the systems and of strengthening coherence between the different regulatory frameworks that may be applicable was highlighted.

Likewise, the technical guides developed within the framework of the Sandbox were positively valued as support tools for interpreting the AI Act. However, participants identified opportunities for improvement related to the incorporation of more operational content, including templates, practical examples, use cases and roadmaps that facilitate the implementation of regulatory requirements in organizations with different levels of maturity.

As the main conclusion of the meeting, a broad consensus was reached on the need to strengthen the competitiveness of the European artificial intelligence ecosystem through an **appropriate**

balance between regulation and innovation. Among the priorities identified were the strengthening of support mechanisms for companies, the simplification of certain compliance processes, the improvement of coordination between competent authorities and the consolidation of regulatory *sandboxes* as permanent instruments of support and learning.

Training Pills

As a complement to these technical exchange activities, SEDIA published a series of six training pills through its official LinkedIn and YouTube channels. These informational materials summarize some of the main aspects of the AI Act and capture part of the lessons learned during the development of the *Sandbox*, facilitating their dissemination among companies, professionals and other agents of the ecosystem.

The **six pills**[5] address the following thematic areas:

- **AI systems and risks**: Introduces the AI Act 's risk-based approach and classification of AI systems.
- **Harmonised standards**: Explains the role of harmonised standards in demonstrating compliance with the AI Act.
- **AI System Life Cycle**: Presents the main phases of AI system development and associated requirements.
- **Medical devices**: Discusses the application of the AI Act to medical devices that incorporate artificial intelligence.
- **Data governance**: Summarizes the requirements to ensure the quality and proper management of the data used by AI systems.
- **Technical documentation and conformity assessment**: Sets out the necessary documentation required and the process for demonstrating compliance with the requirements of the AI Act.

[5]Bringing AI Regulation closer: Pills with expert people from the AI Sandbox
https://www.youtube.com/playlist?list=PLPS6msFrRuV9Lam4tfwlhBRdIKtfNPBCf

Thematic Webinars

Two specialised webinars were also organised to analyse some of the main challenges arising from the practical application of the AI Act. These sessions brought together experts from different fields and made it possible to share experiences, discuss practical cases and familiarize companies and professionals with issues that are particularly relevant to the implementation of the artificial intelligence systems. The MSAs and companies participating in the Sandbox were also invited to both Webinars.

Webinar 1. AI Sandbox: No-Code AI and High-Risk Systems: The Challenge for Agents in the Face of the AI Act - *March 18, 2026*

This session was dedicated to the rise of **no-code *artificial intelligence tools***, which allow the development of advanced solutions without the need for programming knowledge, and to the challenges that this phenomenon poses from a regulatory perspective.

The participating experts analysed both the opportunities offered by these technologies to democratise the development of solutions based on artificial intelligence and the associated risks in areas such as governance, cybersecurity, privacy or legal liability.

The discussion highlighted the **need to develop appropriate control and oversight mechanisms that make it possible** to harness the potential of these tools without compromising security, transparency and compliance, especially when used in high-risk contexts. The role of initiatives such as the Sandbox in **facilitating the responsible adoption** of these technologies and promoting an appropriate balance between innovation and protection of fundamental rights was also highlighted.

Webinar 2. AI Act Sandboxes: Implementation, European Perspective and State of Play - 22 April 2026

The webinar brought together national and international experts who shared experiences on different regulatory *sandbox models* developed in Europe, including the presentation of the EUSAiR project and the experience of the *Luxembourg regulatory* sandbox.

During the session, the different approaches adopted for the implementation of these controlled testing environments were analysed, their contribution to the adaptation of organizations to the new regulatory framework and the need to move towards shared methodologies across Europe. The discussion highlighted the value of sandboxes in translating regulatory requirements into practical scenarios, facilitating the participation of companies - especially SMEs - and consolidating these environments as a structural tool for governance and support for responsible innovation.

The participation of professionals from different countries made it possible **to place the Spanish experience in a comparative context and to compare the lessons** learned during the development of the pilot with those of other European initiatives. Among the elements identified as key for future experiences were the need to develop common methodological criteria, strengthen coordination mechanisms between competent authorities and have practical tools adapted to different levels of organizational maturity.

European Collaboration and Participation in EUSAiR[6]

The European dimension was one of the differentiating elements of the Spanish AI *Sandbox*. Since its launch, the initiative has been developed in close coordination with the **European AI Office, the European Commission and other Member States,** through regular meetings and information exchange forums aimed at sharing experiences, identifying common challenges and promoting a consistent application of the AI Act across the European Union. This coordination also supported the exchange of experiences, the monitoring of the work on the development of harmonised standards complementary to the AI Act, in particular through the CEN-CENELEC standardisation groups, and alignment with other European initiatives.

As part of this collaboration, the Sandbox participated in one of the *workshops* organized by the project. The meeting brought together institutional representatives, experts and leaders of European initiatives with the aim of exchanging experiences on the development of these controlled test environments and promoting cooperation between the different national ecosystems.

In this context, participation **in the EUSAiR – *EU Regulatory Sandboxes for AI project is particularly noteworthy***. This initiative, funded by the European Commission's Digital Europe Programme whose objective is to support, coordinate and promote the development of AI regulatory *sandboxes* in Member States, in line with Article 57 of the AI Act.

During the workshop, the different approaches adopted by Member States for the implementation of regulatory *sandboxes* were analysed, identifying good practices, shared challenges and opportunities for future collaboration. The meeting also made it possible to showcase the experience of the *Spanish AI Sandbox*, share the main lessons learned during its development and compare the results obtained with other similar European initiatives.

Participation in EUSAiR contributed to strengthening collaboration between public authorities, innovation ecosystems and bodies involved in the implementation of the AI Act, promoting the

[6] EUSAiR website. https://eusair-project.eu/

adoption of common approaches, the exchange of methodologies and the generation of shared knowledge on the operation of *regulatory sandboxes*. It also made it possible to identify key elements for the development of future initiatives, including the need to strengthen coordination mechanisms between competent authorities, promote homogeneous methodological criteria and facilitate the access of innovators, especially SMEs and startups, to this type of regulatory experimentation environment.

Overall, the European collaboration carried out during **the *Sandbox* has contributed to positioning Spain as one of the benchmarks in the practical application of the AI Act** and has made it possible to bring the lessons learned from the pilot into the European debates on the design, implementation and evolution of future AI regulatory *sandboxes*.

3. Analysis of Results

This section presents an analysis of the **main results** obtained during the development of the first pilot of the *Spanish AI Sandbox*. Based on the information collected throughout the different phases of the project, it examines both the **aggregate results** derived from the participation of the entities and the **main lessons learned** during the implementation of the initiative.

The analysis is structured in two main blocks. First, the criteria used for the grouping of the participating systems are described. And secondly, the operational results and lessons learned identified in the different phases of the *Sandbox are presented*, including a specific analysis of the effort required for the implementation of the adaptation measures proposed in the technical guides, considering both the temporal aspects and the resources and capacities necessary for their execution.

Grouping Criteria

The first pilot of the *Spanish AI Sandbox* initially included twelve high-risk artificial intelligence systems, belonging to entities of different nature, size and sector of activity. This diversity made it possible to analyse the practical application of the technical guides in heterogeneous organizational and technological contexts, as well as to identify differences in terms of maturity, availability of resources, documentary burden, technical complexity and implementation capacity.

In order **to facilitate the analysis of the results**, the participating entities have been grouped according to two main criteria: The size of the organization and the sector of activity of the AI system evaluated.

The first grouping, shown in the **Table 3 - Sorting entities by size**, was carried out taking into account the size of the organisation, given that this factor can significantly influence aspects such as the availability of human and economic resources, the capacity for coordination between areas and the management of the internal processes necessary for adaptation to the AI Act.

The second grouping, shown in the **Table 4 - Classification of entities by sector**, was carried out according to the sector of activity of each system. This criterion is relevant because certain sectors are subject to specific regulatory frameworks or additional sectoral obligations that may condition the implementation processes of the measures analysed during the Sandbox.

Large Company	Foundation Center for Research in Information and Communication Technologies of Galicia (FCITICG)
	Airbus Operations SL
	Adevinta Jobs, S.L.U.
Medium-sized Company	Veridas Digital Authentication Solutions, S.L.
	Bit&Brain technologies, S.L
Small Company	Herta Security, S.L.
	Imotion Analytics, SL
	Loradix Red, S.L
Startup	Dedomena Artificial Intelligence, SL
	Made Of Genes, S.L. U
	Shakers Global Solutions SL
	Tucuvi Care SL

Table 3 - Sorting entities by size

Medical Devices	Bit&Brain technologies, S.L
	Tucuvi Care SL s
	Made Of Genes, S.L. U
Access to benefits and services	Foundation Center for Research in Information and Communication Technologies of Galicia (FCITICG)
	Dedomena Artificial Intelligence, SL
Machinery and critical infrastructure	Airbus Operations SL
	Imotion Analytics, SL
	Loradix Red, S.L
Biometrics	Veridas Digital Authentication Solutions, S.L.
	Herta Security, S.L.
Employment	Shakers Global Solutions SL
	Adevinta Jobs, S.L.U.

Table 4 - Classification of entities by sector

It should be noted that Loradix Red, S.L. communicated its voluntary withdrawal from the controlled test environment during the course of the pilot, for reasons related to a lack of financial and human resources. Consequently, although the entity forms part of the initially selected sample and remains included in the classifications presented, some of the analyses contained in this report refer only to the eleven systems that effectively completed the activities provided for in the *Sandbox*.

The two classifications described constitute the methodological basis used for the preparation of the aggregate analyses included in this report and for the interpretation of certain results observed during the pilot. However, the experience gained shows that the degree of progress achieved by each system does not depend exclusively on variables such as the size of the entity or the sector of activity. Factors such as the complexity of the use case, the level of previous maturity, the available documentation, the applicable sectoral regulatory framework, the composition of the teams

involved, and the scope of the prioritized measures also had a significant influence on the results obtained

For this reason, the aggregate analysis is complemented by an individualized evaluation of each participating system, incorporated in the **ANNEX II: Analysis of results by Entity,** which includes specific files organized by sector, use case and evolution observed during the development of the *Sandbox*.

In any case, the information included in these profiles has been treated in accordance with the **confidentiality framework agreed during the execution of the *Spanish AI* Sandbox pilot**. For the preparation of this document, the participating entities have expressly authorised the use and disclosure of the information contained herein, guaranteeing that it is processed in a limited, proportionate and appropriate manner and in accordance with the applicable confidentiality and privacy obligations.

Phased Results

This section presents the main monitoring indicators (KPIs) of the Spanish AI *Sandbox*, prepared from the information collected during the different phases of the pilot. Its analysis makes it possible to assess the degree of progress achieved by the participating entities, identify common patterns in the adaptation processes and extract relevant lessons on the practical application of the measures included in the technical guides.

The indicators included in this report were constructed from information drawn from different documentary and monitoring sources generated during the execution of the *Sandbox*. In the case of Phase I – Training, the data were mainly obtained from the project's coordination mechanisms, attendance records, satisfaction surveys and queries raised by the participating entities during the training sessions.

For the subsequent phases, the information analysed is drawn mainly from the different deliverables prepared by the participating entities, including the results of the self-diagnosis exercises, the Adaptation Plans, the monitoring tools used during the implementation of measures and the documentation generated throughout the self-assessment and post-market monitoring processes.

The analysis of this information makes it possible to assess both the **evolution** of the entities during the Sandbox and the main **challenges and benefits** identified throughout the adaptation process. It also provides an objective basis for validating the lessons learned during the pilot and supporting the conclusions contained in this report.

Once the available information had been compiled, reviewed and consolidated, the indicators and visualisations presented in the following sections were drawn up, which allow for a comparative analysis of the behaviour of the participating entities and the evolution observed throughout the different phases of the *Sandbox*.

Phase I: Training

This initial phase focused on delivering training sessions on the technical guides to the participants, which made it possible to establish a common understanding of the AI Act before the start of the *Sandbox*, provide structured support materials that were positively evaluated by the participants, and identify queries and needs at an early stage.

The analysis of this phase, from a quantitative perspective, allows us to understand the initial interest and the areas of greater perceived complexity.

Indicator 1: Attendance

- **Metric**: Number of attendees registered for each training session.
- **Analysis**: Assesses the degree of interest and participation in the training.

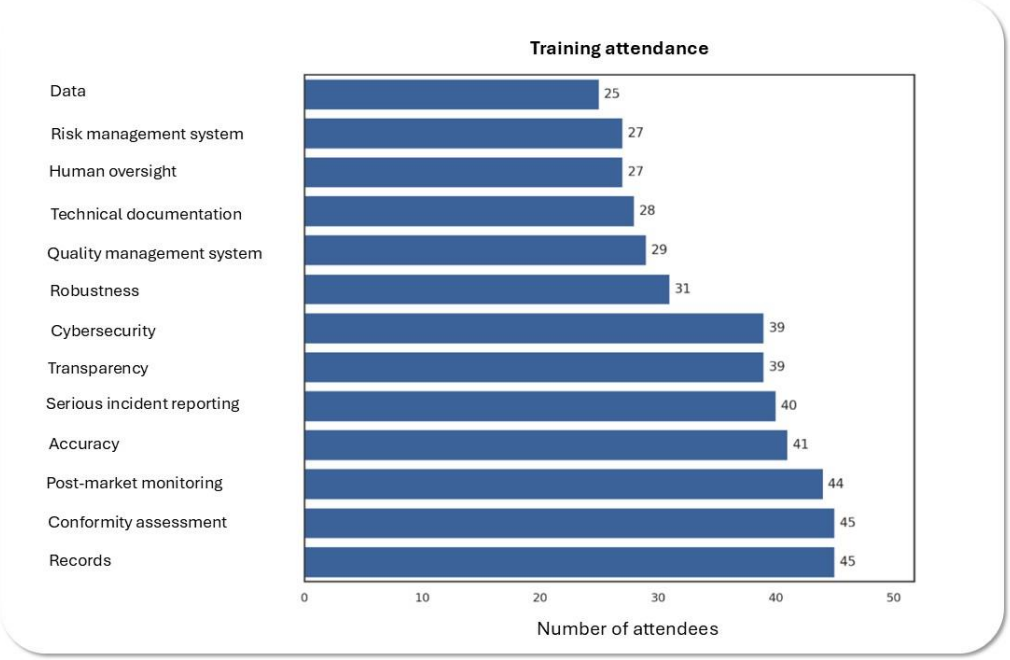

Image 4 –Number of attendees per training of the technical guides

After observing the **Image 4 –Number of attendees per training of the technical guides**, the following conclusions can be drawn:

- The sessions with the highest attendance were conformity assessment and records (45 attendees each), followed by post-market monitoring (44), accuracy (41) and serious incident reporting (40). This indicates a high interest in matters related to compliance demonstration, traceability and post-market obligations.

- The sessions with the lowest attendance were data (25 attendees), risk management system and human oversight (27 attendees each). However, lower attendance does not necessarily imply less relevance, since some of these topics generated a high volume of questions.

Indicator 2: Level of Interaction and Understanding

- **Metric**: Number of questions asked by attendees in each session.
- **Analysis**: Measures the degree of involvement and identifies the technical guides that present the greatest complexity for the participants.

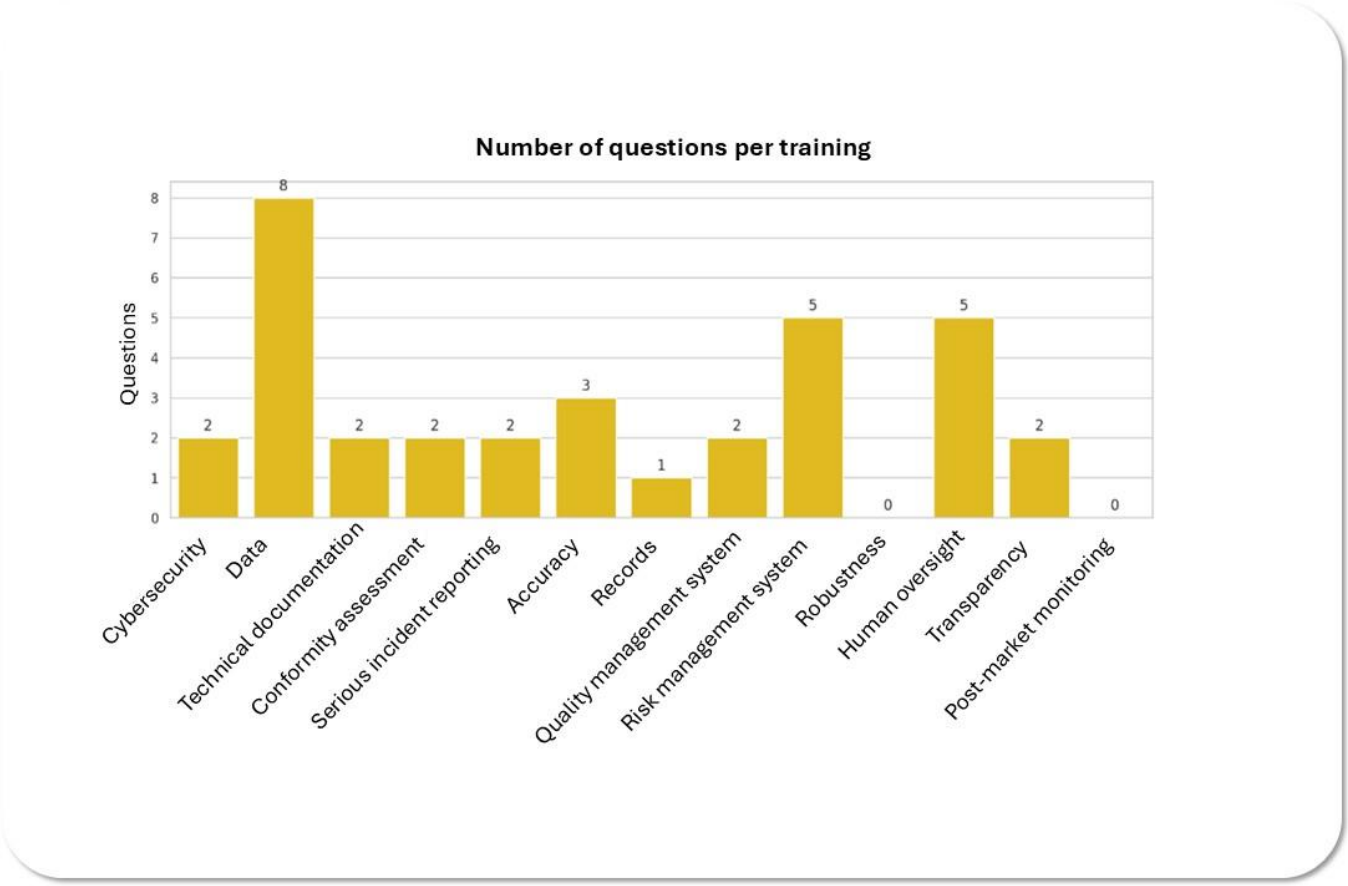

Image 5 - Questions by Training Session from the Technical Guides

After reviewing **Image 5 - Questions by Training Session from the Technical Guides** , we can reach the following conclusions:

- The topics that generated the most doubts were data (8 questions), and risk management system and human oversight (5 questions each), identifying them as areas of high complexity.
- The data guide is particularly noteworthy, as it had the lowest attendance but the highest number of questions, suggesting a high level of perceived complexity among a highly engaged group of participants.

- In contrast, robustness and post-market monitoring did not register questions. in the case of post-market monitoring, their high attendance could indicate a good understanding of the subject. In the case of robustness, this could indicate either lower engagement or greater difficulty in formulating questions. It is recommended that future editions actively explore whether this lack of interaction reflects a full understanding of the topic or a perception that it is of lesser relevance to participants.

Indicator 3: Attendance/Enrollment Ratio

- **Metric**: Percentage of registrants who finally attended each session.
- **Analysis**: Determine the level of effective participation, comparing attendees with those initially registered.

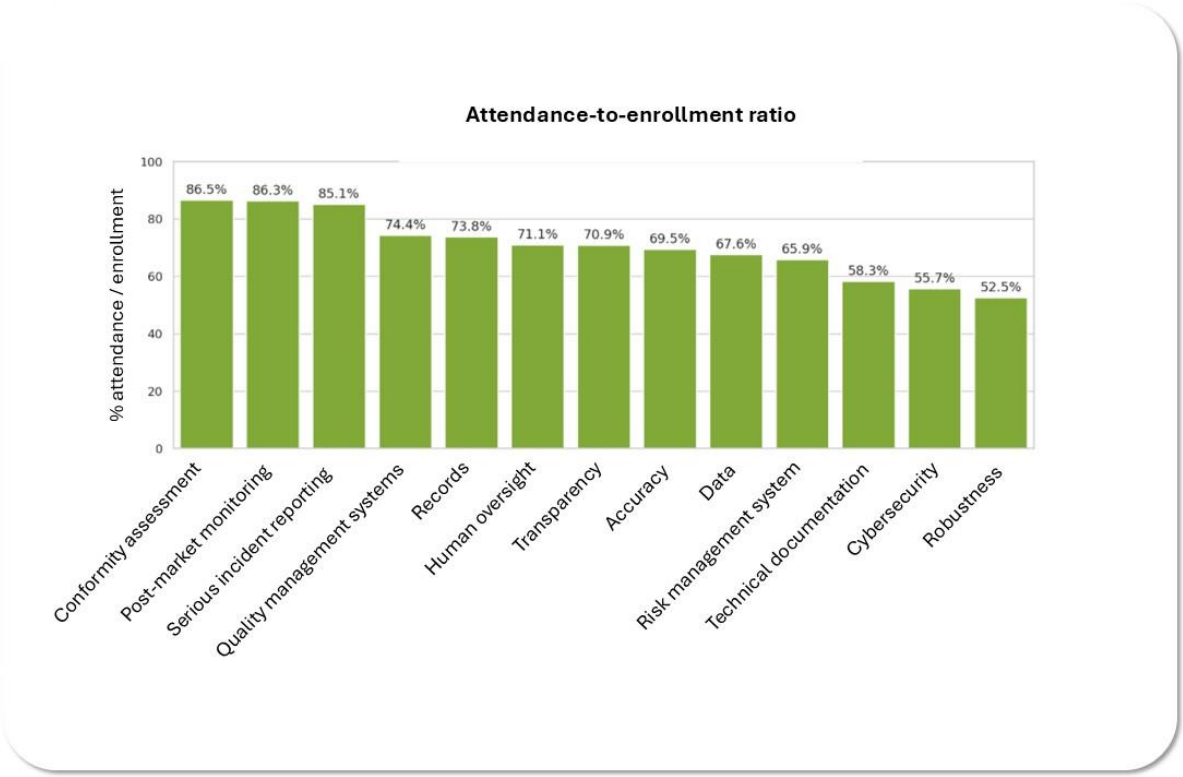

Image 6 – Attendance/registration ratio for training in the technical guides

After observing the **Image 6 – Attendance/registration ratio for training in the technical guides,** the following conclusions can be drawn:

- The sessions with the highest attendance-to-registration ratio were conformity assessment (86.5%), post-market monitoring (86.3%), and serious incident reporting (85.1%), reflecting a high level of engagement.
- The lowest ratios were recorded in robustness (52.5%), cybersecurity (55.7%) and technical documentation (58.3%), possibly due to the greater technical specialization of the content.

In summary, the conclusions drawn from this analysis of the training phase are:

- **Higher attendance didn't always translate into greater engagement**. Matters such as data, risk management system and human oversight generated a significant need for clarification despite attracting fewer attendees, sand should therefore be considered priority areas for additional support.
- The more technical guides - data, robustness, cybersecurity and technical documentation - could benefit from additional training, such as practical sessions or supplementary materials.
- The sessions on conformity assessment, post-market monitoring and serious incident reporting performed particularly well, **combining high attendance with a high participation rate**.

Phase II: Self-diagnosis and Adaptation Plan

This phase focused on the self-diagnosis of the initial alignment of the systems with the technical guides, by the participating entities, as well as the development of individualized adaptation plans. All this made it possible to have information on the applicability, complexity and initially perceived effort of the measures, as well as the detection of cross-cutting technical doubts.

The analysis of this phase provides insight into the initial situation, the perceived level of difficulty and the measures most frequently addressed.

Indicator 1: Proactivity and Continuous Improvement

- **Metric**: Ability of the entities to identify and incorporate value-added measures not explicitly required in the technical guides.
- **Analysis**: Assesses whether factors such as workload or technical knowledge influenced the proactivity of the entities.

After analysing the contributions made, the following conclusions can be drawn:

- Proactivity **was limited**. Of all the additional measures proposed by the participants, only one was ultimately incorporated. This focus on the minimum required measures could be due to the workload already involved in the adaptation process. To encourage greater proactivity in the future, the creation of incentives or specific spaces for the co-creation of improvements could be considered.

Indicator 2: Perceived Difficulty by Sector and Company Size

- **Metric**: Level of difficulty perceived by the entities in adopting the measures included in the self-diagnosis, with the results differentiated according to the sector of activity and the size of the company.
- **Analysis**: Identifies which technical guides are considered the most difficult to apply and assesses whether the perceived difficulty varies according to the sector or the size of the participating entities.

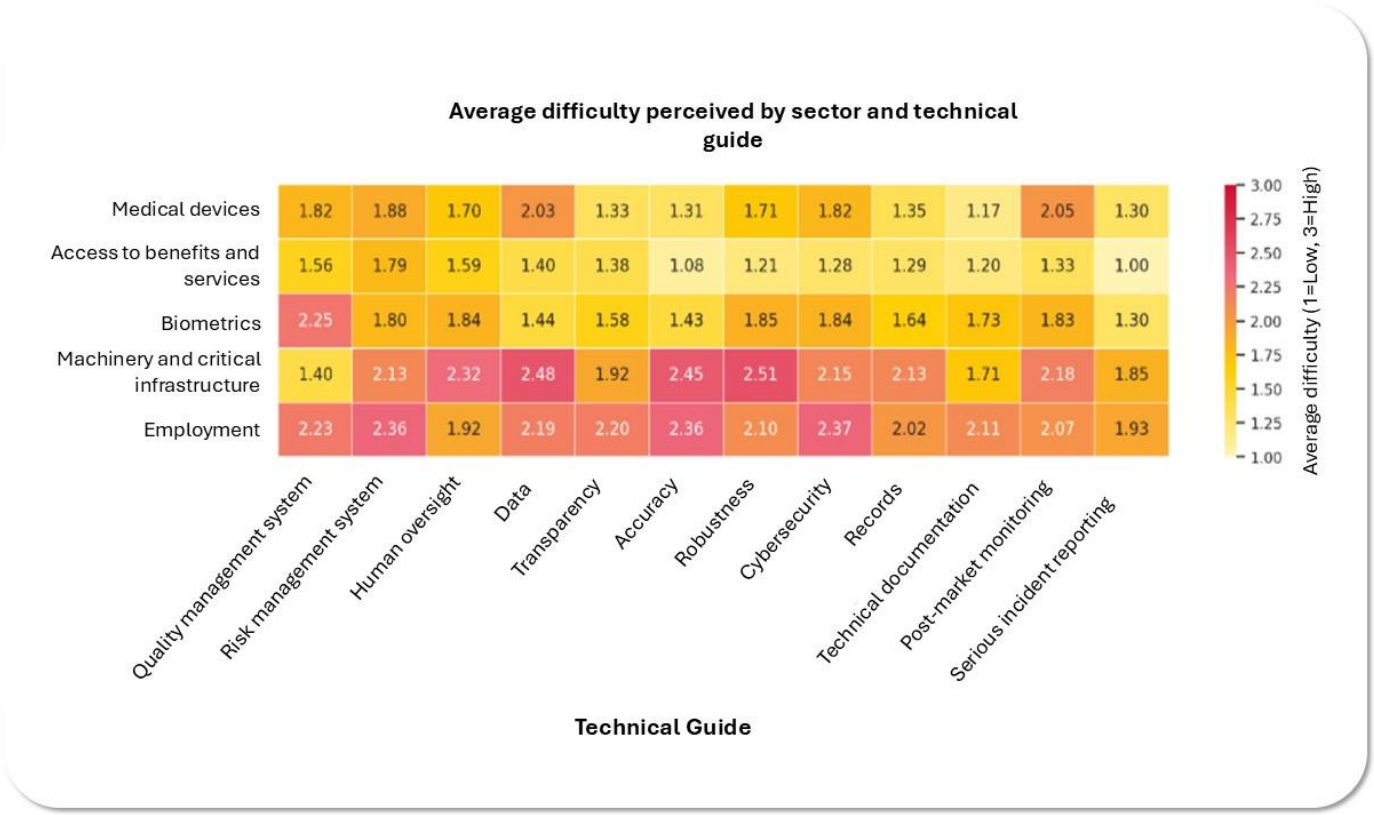

Image 7 - Heat Map: Average difficulty perceived by sector and technical guide

To determine whether the differences observed in the **Image 7 - Heat Map: Average difficulty perceived by sector and technical guide,** between sectors are statistically significant, the Kruskal-Wallis statistical test was applied:

- H = 51,312
- p-value < 0.001

The p-value is less than 0.05, therefore, there are **statistically significant differences between sectors**. This indicates that the observed differences in perceived difficulty do not appear to be due to chance.

Although differences between the sectors were already apparent at first glance, it was necessary to verify whether they reflected a real pattern or if they could have occurred by chance. To this end, a statistical test was applied that acts as a verification mechanism.

The test confirmed that the differences observed between sectors are not the result of random variations but reflect real differences in the perceived difficulty of undertaking the adaptation process.

Consequently, it can be stated with a high degree of confidence that the sector of activity influences the perception of the difficulty, so it is a relevant criterion for interpreting the results obtained.

To determine whether the perceived difficulty varies according to the size of the company, the Kruskal-Wallis test was applied again:

- H = 2.589
- p-value = 0.460

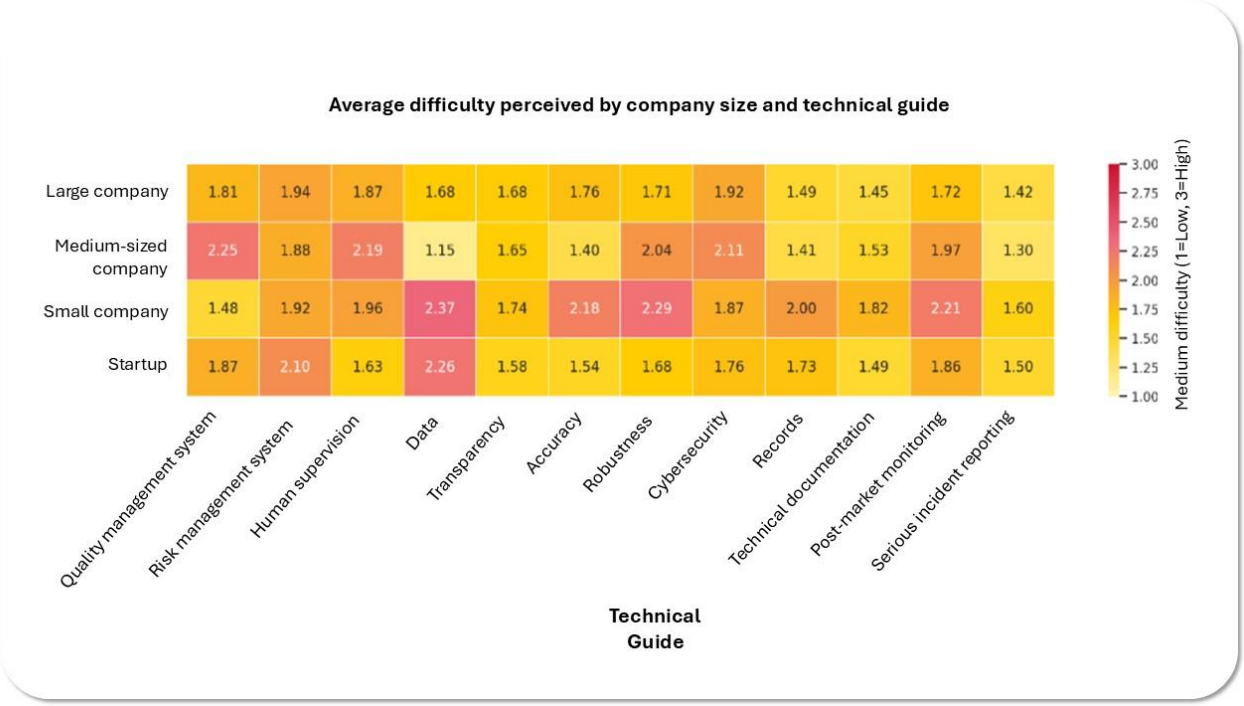

Image 8 - Heat Map: Medium difficulty perceived by size and technical guide

The p-value is greater than 0.05, therefore statistically **Image 8 - Heat Map: Medium difficulty perceived by size and technical guide** is significant differences according to the size of the company.

In this case, the same verification process was carried out to analyse whether the perceived difficulty varied according to company size. Although small differences were observed between the different categories, it was necessary to determine whether they were sufficiently consistent to be considered real.

The results indicate that the observed variations are compatible with the usual variability that may exist between the entities analysed. In other words, there is insufficient evidence to affirm that the company size significantly influences the perceived difficulty during the adaptation process. This

suggests that the challenges identified are, in general terms, similar regardless of whether the entity is small, medium, or large.

After carrying out the relevant analysis for the indicator of perceived difficulty by sector and company size, the following conclusions can be drawn:

- The **perceived difficulty varies clearly according to the sector** and, to a lesser extent, according to company size.
- By sector, the greatest difficulties are concentrated in **employment, machinery and critical infrastructure**, with values close to or greater than 2.2 in measures found in technical guides such as data, accuracy, robustness, cybersecurity and risk management. This difference is supported by the Kruskal-Wallis test, which shows statistically significant differences according to sector ($H = 51.312$; p-value < 0.001), indicating that the observed differences do not appear to be due to chance.
- By company size, descriptive differences are observed, particularly among small companies and *startups*, which experience greater difficulties in some technical guides such as data, accuracy, robustness and post-market monitoring. However, the Kruskal-Wallis contrast does not show statistically significant differences according to company size ($H = 2.589$; p-value = 0.460). Although the differences by company size are not statistically robust, the greater descriptive difficulties observed in *startups* suggest that they could benefit from more targeted support in technical requirements, which should be explored in future initiatives.
- Overall, the indicator suggests that **the sector is the factor that best explains the differences in perceived difficulty**, while the effect of company size should be interpreted with caution.

Indicator 3: Correlation between Maturity and Perceived Difficulty

- **Metric**: Spearman's correlation between the perceived difficulty of adopting a measure and its level of maturity, considering its degree of implementation and documentation.
- **Analysis**: Determines whether measures with lower levels of maturity present greater perceived difficulty and identifies the main compliance gaps.

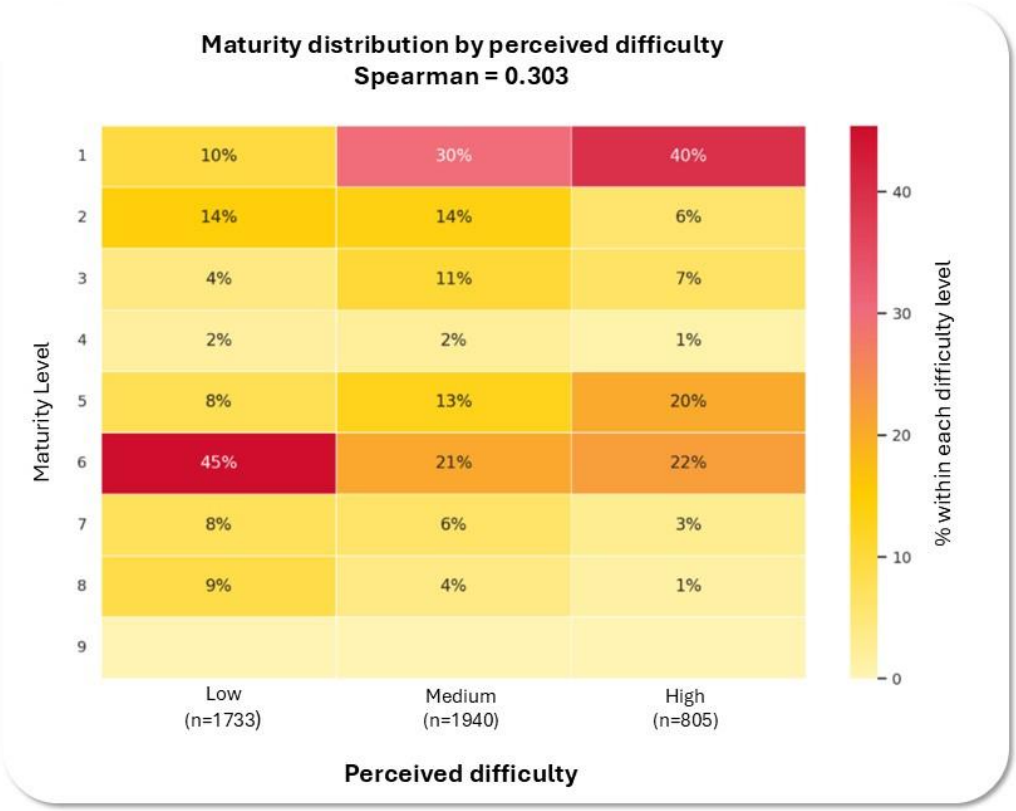

Image 9 - Heat Map of the correlation between maturity and perceived difficulty

*n = number of measures counted

Maturity level	Maturity level value
1	L1. Not documented or implemented
2	L2. Documentation in progress and not implemented
3	L3. Not documented and implementation in progress
4	L4. Documented and not implemnted
5	L5. Documented and implementation in progress
6	L6. Documented and implemented
7	L7. Not documented and implemented
8	L8. Documentation in progress and implemented
9	L9. Measure not required for my system

Table 5 – Maturity Level of the Measures Identified in the Adaptation Plan

The observation of the **Table 5 – Maturity Level of the Measures Identified in the daptation Plan** allows the following conclusions to be drawn: **Image 9 - Heat Map of the correlation between maturity and perceived difficulty**

The Spearman correlation coefficient obtained is **–0.303**. This result shows a negative relationship of moderate-to-low intensity: in general, **the lower the level of maturity of a measure, the greater the perceived difficulty in adopting it**.

- The main gap is at **level 1: measures neither documented nor implemented**, which accounts for **40% of the measures assessed as having high difficulty**. On the other hand, measures with low difficulty are mainly concentrated at **level 6: measures documented and implemented**, which represents **45% of measures with low difficulty**.
- In conclusion, the results show that **less advanced measures tend to be perceived as more difficult**, while measures already documented and implemented are usually associated with lower perceived difficulty.
- However, the relationship observed is not particularly strong. Therefore, **the level of maturity alone does not explain the perceived difficulty**. Other factors may also play a role, such as the type of technical guide, the sector of activity, the technical complexity or the resources available.
- Overall, the indicator identifies the measures that have not yet been documented or implemented as the main area requiring attention, for which additional support may be needed.

Indicator 4: Analysis of the Measures to be Implemented (Adaptation Plan)

- **Metric**: Distribution of the actions planned by the entities for each measure **("Not Reported", "Document and Implement", "Document", "Implement", "Documented and Implemented").**
- **Analysis**: Identifies the main adaptation needs of the entities and determines whether they vary according to the technical guide, the sector of activity or the size of the company.

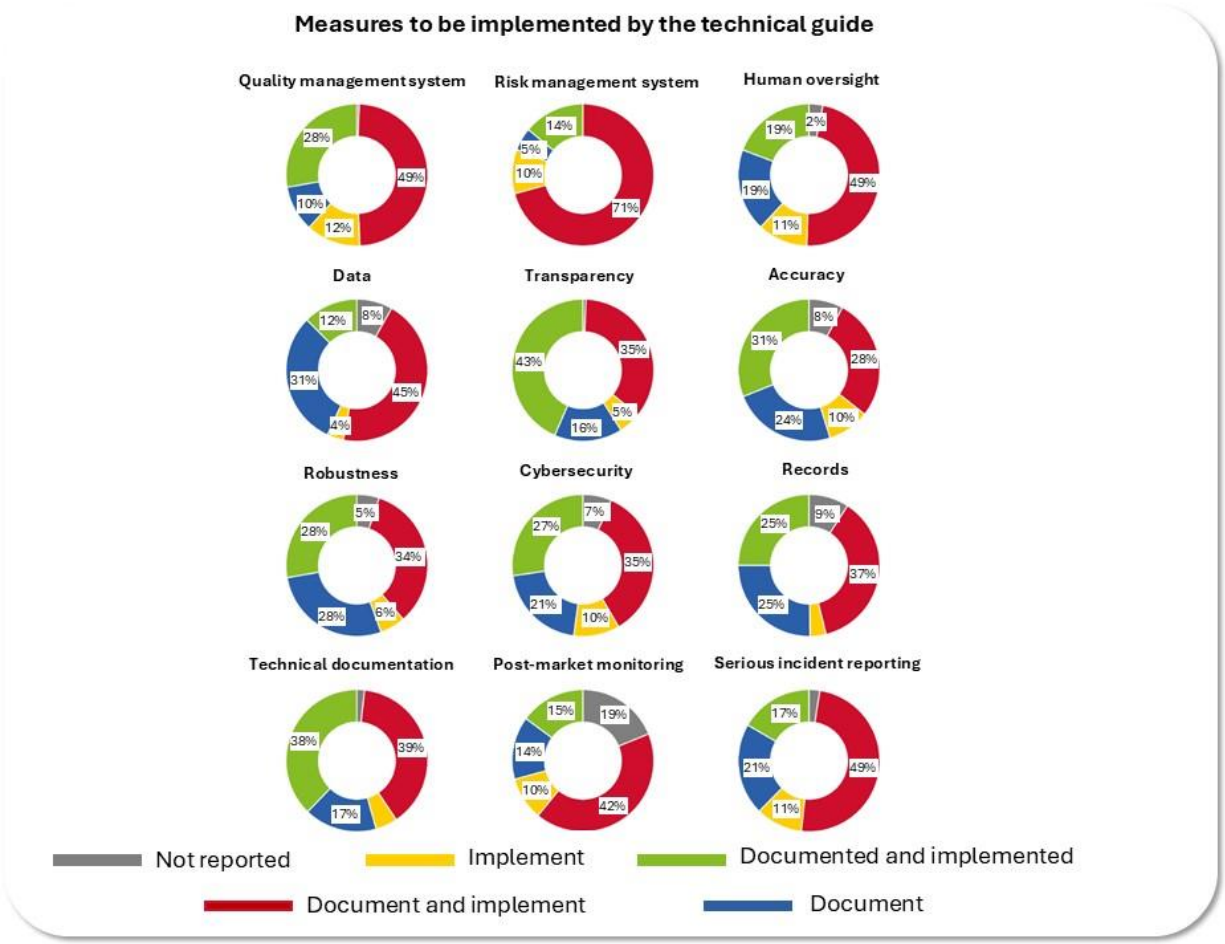

Image 10 - Measures to be implemented by technical guide

Thus, according to the **Image 10 - Measures to be implemented by technical guide**Image 10 - Measures to be implemented by technical guide:

- The analysis of the measures provided for in the Adaptation Plans shows that the **category "Document and Implement" is the most recurrent**. This shows that, for a significant proportion of the entities, the adaptation to the Sandbox framework consisted not only of formalising existing actions, but also of simultaneously developing operational measures and their corresponding documentary support.
- From the perspective of the technical guides, this need for comprehensive adaptation is especially observed in structural areas such as the **risk management system**, **human oversight**, **serious incident reporting**, **quality management system** and **data**. These results suggest that the more cross-cutting and governance requirements required a greater combined implementation and documentation effort.

After observing **Image 11 - Measures to be implemented by sector.** We can reach the following conclusions:

- By sector, there are significant differences in the degree of initial maturity. Sectors such as medical devices, access to aid and services and employment show a greater proportion of measures to be **"Document and Implement",** which points to a greater need for comprehensive adaptation. On the other hand, biometrics and critical machinery and infrastructure have a high proportion of measures already documented and implemented, which reflects a higher degree of previous progress compared to the rest of the sectors.

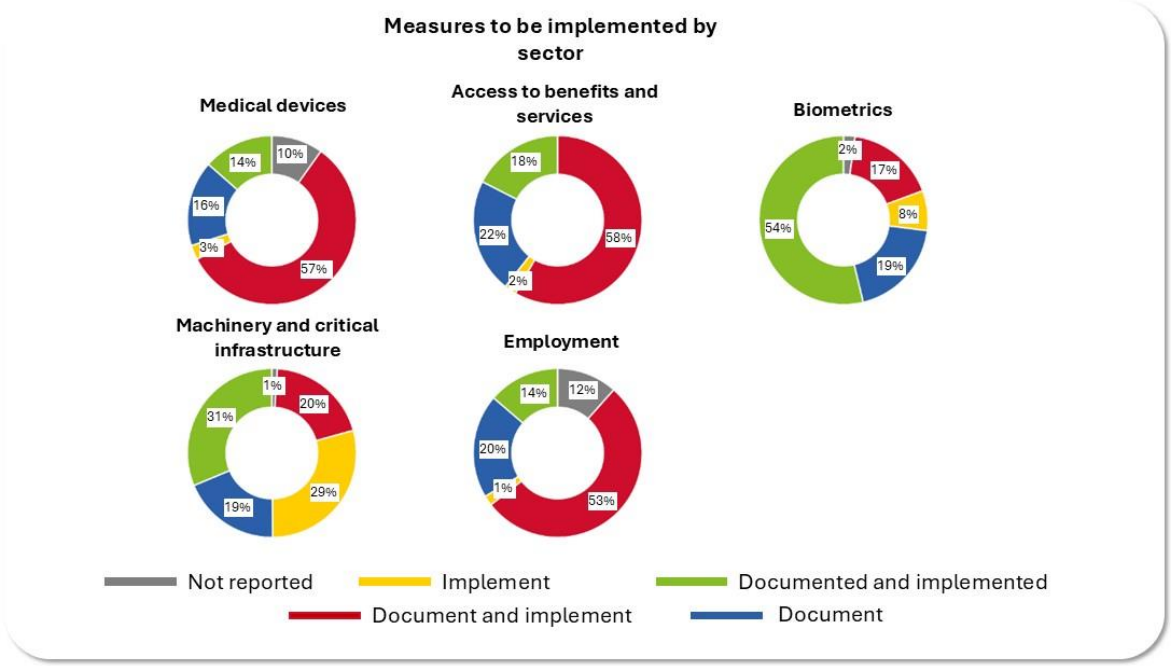

Image 11 - Measures to be implemented by sector

After observing **Image 12 - Measures to be implemented by company size** We can reach the following conclusions:

- In terms of company **size**, *startups* are the group with the highest concentration of measures to **be documented and implemented**, which can be associated with a lower degree of prior formalization. On the other hand, **small companies** show a greater proportion of measures pending only documentation, which suggests that part of the actions may already have been developed, but required greater traceability and formalisation.
- In the case of **Large and medium-sized companies**, the distribution is **More balanced**, combining measures to be documented and implemented with measures already documented and implemented or pending documentation only. This points to a greater presence of previous structures, although not always fully aligned with the specific technical guides of the Sandbox.

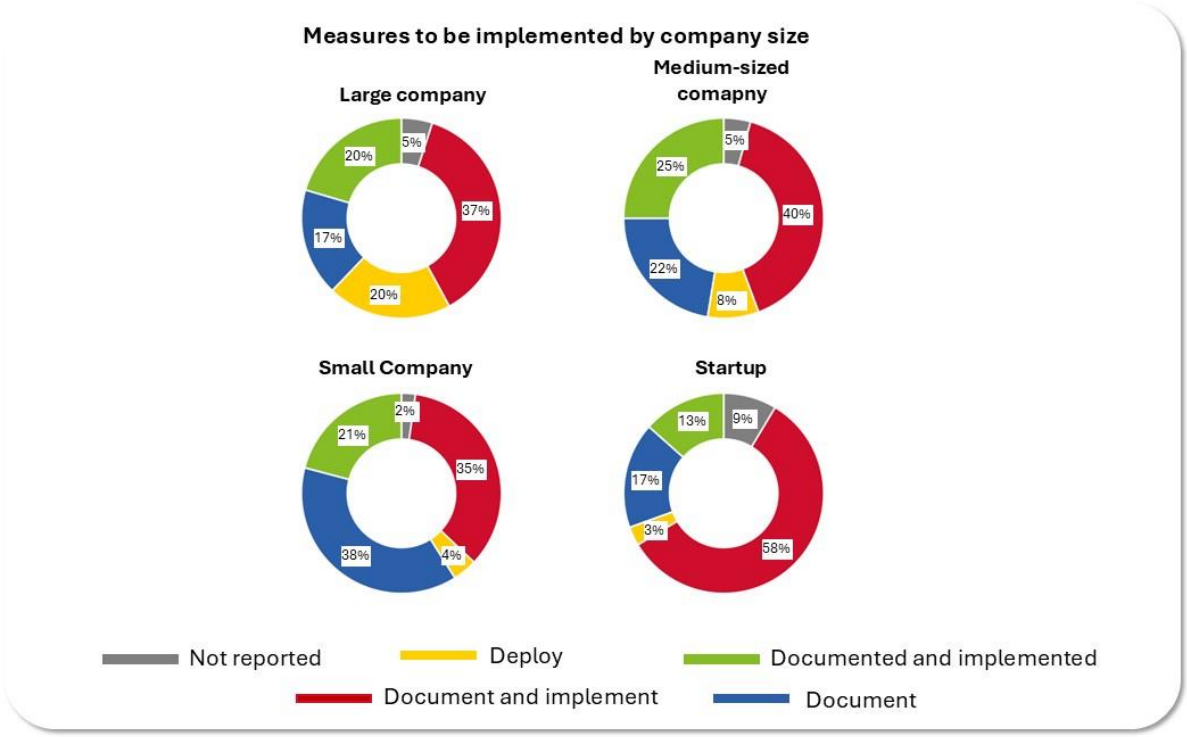

Image 12 - Measures to be implemented by company size

After observing the results of the indicator for the analysis of the measures to be implemented, the most frequent action in adaptation plans is **to document and implement**. Consequently, the adaptation to the Sandbox was not limited to formalizing existing practices, but also required the development of measures, establishing controls and generating the necessary documentation to prove their application.

- The greatest adaptation needs are concentrated in cross-cutting governance technical guides, such as risk management, human oversight, serious incident reporting, quality management system and data governance.
- Differences are also observed according to the sector. **medical devices, access to aid and services and employment** present a greater need for comprehensive adaptation, while **biometrics and machinery and critical infrastructure** show a greater degree of initial progress.
- In relation to size, ***startups*** concentrate more measures pending documentation and implementation, while **small companies** mainly need to formalize and document existing actions. Medium **and large companies** show a more balanced situation and a greater presence of previous structures.

- Overall, the indicator shows that adaptation needs depend both on the initial level of progress of the entities and on their sector and size. These results allow the identification of areas where more targeted support might be needed.

Summing up:

- The joint analysis of the indicators shows that Phase II made it possible to clearly identify the main adaptation gaps of the participating entities, both from an operational and documentary point of view. In general terms, the entities focused their efforts on complying with the measures defined in the guides, while the incorporation of additional continuous improvement measures was limited.
- The perceived difficulty was not homogeneous between entities. The **sector is confirmed as the most relevant factor** in explaining the observed differences, since the Kruskal-Wallis contrast shows statistically significant differences in the mean difficulty according to sector ($H = 51.312$; p-value < 0.001). In particular, the employment and critical machinery and infrastructure sectors concentrated higher levels of difficulty, especially in technical guides such as data, accuracy, robustness, cybersecurity and risk management.
- On the other hand, company **size did not show statistically significant differences** ($H = 2.589$; p-value = 0.460). Although small companies and *startups* present greater difficulties in some technical guides containing measures requiring greater technological involvement, these differences should be interpreted with caution, since there is not enough evidence to affirm that size robustly explains the perceived difficulty.
- Likewise, an **inverse relationship between maturity and perceived difficulty is observed**. Spearman's correlation of -0.303 indicates that measures with lower maturity tend to be perceived as more difficult. This gap is especially concentrated in level 1: undocumented and non-implemented measures, where 40% of the measures with high difficulty are located. On the other hand, 45% of the measures with low difficulty are located at level 6: documented and implemented, associated with measures already implemented and/or documented.
- In relation to Adaptation Plans, the most frequent category was **"Document and Implement",** which shows that many entities not only needed to formalize existing actions, but also to develop operational measures from an initial phase. This need was especially visible in structural technical guides such as the **risk management system, human oversight, serious incident reporting, quality management system and data**.
- Overall, Phase II confirms that the main challenge for entities was not only to understand the technical guides, but also to translate them into concrete, documented, traceable and

implementable measures. The results suggest that the greatest support needs are concentrated in technical and governance requirements, especially in sectors with more complex or critical systems.

Phase III: Implementation of the Adaptation Plan

This phase focused on the implementation of adaptation plans by the entities to evaluate the measures that the systems intended to address, those that did not intend to address and those that were still undecided. This resulted in the partial or complete implementation of the prioritized measures, the generation of documentation and an understanding of the real effort invested according to the resources and capacities of each entity.

Indicator 1: Status of Measures to be Addressed by the Technical Guide

- **Metric**: Percentage of measures that systems committed to address, versus those that remained pending decision or were not applicable.
- **Analysis**: Identifies the technical guides in which the entities assumed a greater commitment to action during the implementation phase.

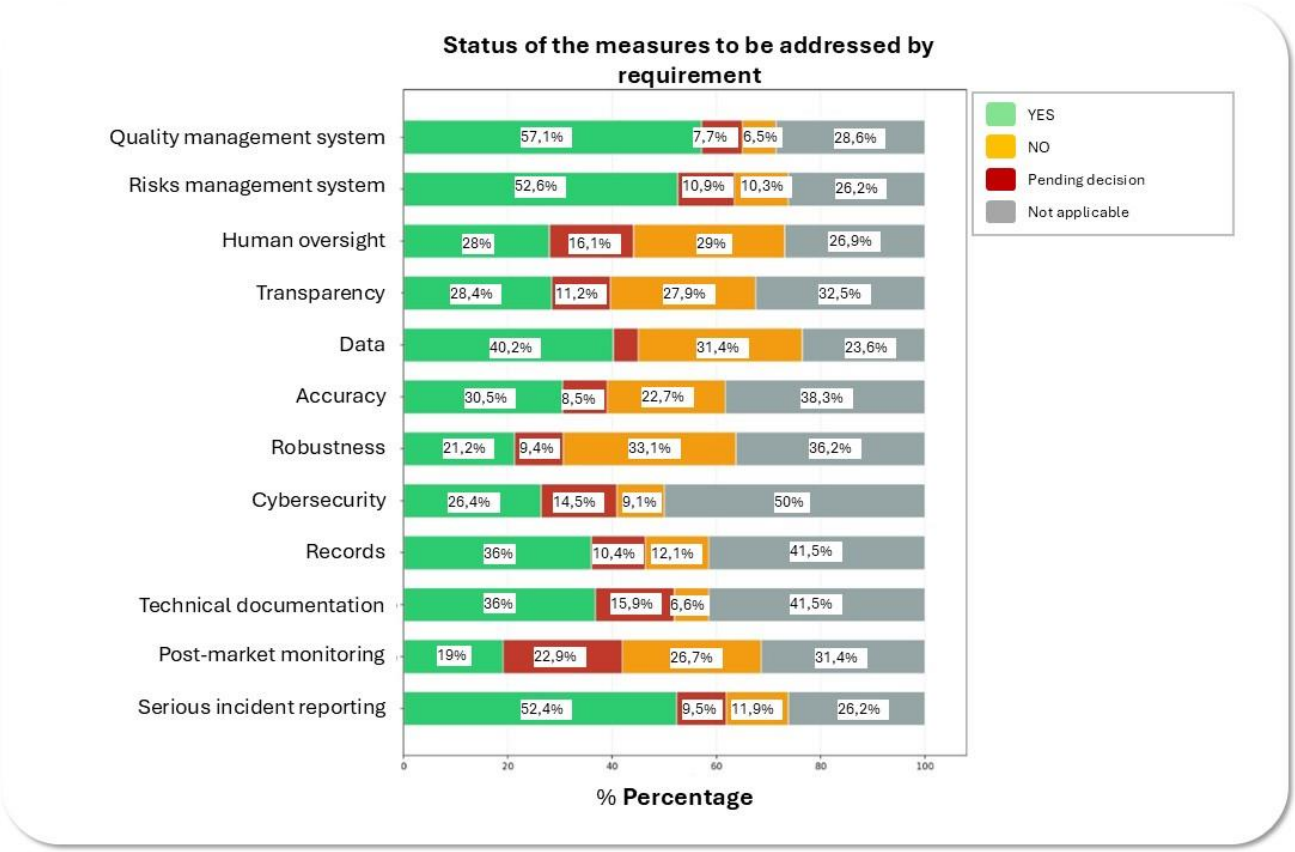

Image 13 - Status of the measures to be addressed by requirement

After observing the **Image 13 - Status of the measures to be addressed by requirement** .We can conclude:

- The was a **high degree of commitment**, especially in technical guides linked to consolidated management frameworks such as quality (57.1% of measures addressed), risk management (52.6%) and serious incident reporting (52.4%).
- However, a significant number of measures remained pending decision in technical guides with more technical requirements such as transparency, soundness and cybersecurity, reflecting a more cautious approach on the part of the entities.
- The **high percentage of "Not applicable" measures** in some technical guides demonstrates a **correct interpretation of the regulatory framework** and its applicability to each specific use case.

Indicator 2: Estimation of the Effort for the Implementation of Measures

To analyse and compare the effort required to implement the adaptation measures provided for in the technical guides, it was decided **to classify the participating entities into different sizes**: large companies, medium-sized enterprises, small enterprises and startups, as described above, in order to assess how the size of the organisation influences the application of these measures.

The analysis focuses on the actual effort expressed in working hours, in order to compare the magnitude of the dedication needed to implement the measures analyzed. Based on the available information, an average effort has been calculated to provide an aggregated and more representative view of the volume of work required.

The comparison shows that **smaller entities have relevant levels of dedication and, in some cases, levels comparable to those of larger entities**. In particular, large, medium, and small companies report very similar actual efforts, while *startups* present a somewhat lower average effort.

Entity Type	Effort in working hours
Large Company	170.5 hours
Medium-Sized Company	164 hours
Small Company	177 hours
Startup	151.5 hours

Table 6- Estimation of time effort based on the size of the entity

In the case of **large companies**, an effort of 170.5 hours has been taken as a reference. This value allows us to contextualise the dedication observed in the rest of the entities and offers a point of comparison with larger organisations.

In the case of **medium-sized enterprises**, the actual declared effort amounts to 164 hours**.** In the case of **small companies**, the actual declared effort amounts to 177 hours. The difference between the two is small, at 13 hours, which indicates that, despite being classified as entities of different sizes, the effort required to address the measures analysed is of a very similar magnitude.

In the case of ***startups***, the average effort is 151.5 hours. This shows that, even in smaller organizations, adapting to the measures can require a significant time investment. It also indicates

that the average effort of *startups* is close to the effort declared by medium-sized companies, small companies and the reference used for large companies.

Considering the available references, the **estimated average effort** is **approximately 165.8 hours** of work. This allows us to contextualise the dedication required to address the measures and shows that, in general terms, adaptation requires a significant workload, even when it comes to limited projects or smaller entities.

The comparison between the different entities allows us to observe that the effort of large companies, medium-sized companies, small companies and *startups* is within a similar range, around **150-180 hours**. This proximity suggests that the size of the entity alone does not determine the volume of work required to address adaptation measures.

Overall, the data analysed suggest that **the adaptation effort does not depend exclusively on the size of the entity**. The dedication required seems to be conditioned by a combination of factors, including the level of risk of the system, the complexity of the project, the degree of previous maturity of the organization, the documentation already available, the number of users affected and the depth of the compliance measures that must be implemented.

Phase IV: Self-Assessment

This phase focused on the self-assessment by the entities to verify the degree of implementation and maturity of the systems, the identification of strengths and shortcomings (in systems and methodology), the obtaining of information on quality and traceability and the validation of the proposed self-assessment mechanisms.

Indicator: Evolution of Maturity (Self-Assessment)

- **Metric**: Final state of implementation of the measures (full implementation and documentation, partial implementation and documentation, insufficient implementation/documentation).
- **Analysis**: Evaluates the real progress of the entities and determines to what extent they managed to implement and document the actions foreseen in their Adaptation Plans.

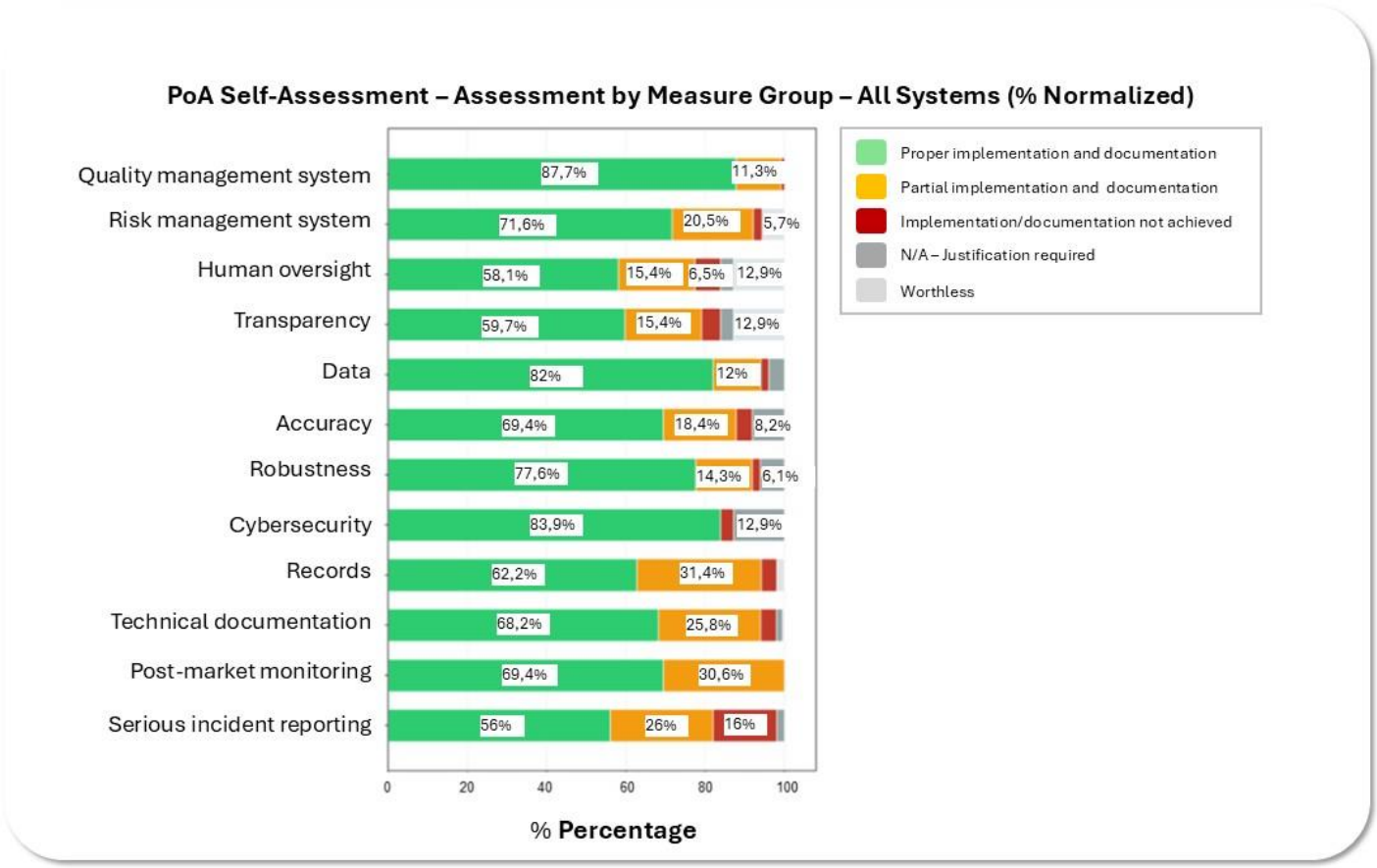

Image 14 -Self-assessment of the measures agreed in the PoA

As it can be seen in the **Image 14 -Self-assessment of the measures agreed in the PoA** is concluded that:

- Overall, the entities managed to fully implement and document most of the measures that were proposed, which indicates that their **Adaptation Plans were realistic and well-designed**.
- The technical guides with the highest success rate in terms of proper implementation and documentation, were quality (87.8%), data (82%) and cybersecurity (83.9%).
- The technical guides in which it was observed that it was most difficult to achieve full implementation were serious incident reporting, human oversight and transparency.
- This suggests that, although the planning was good, the execution and documentation in these areas were more complex than expected, possibly because they require not only technical changes but also deep organizational and cultural processes.
- The **main difficulty** was not only the technical execution, but also the **correct documentation and justification of the measures**, especially in the more complex technical guides.
- The **support of mentoring teams** was key to progress, but practical guidance still needs to be strengthened.
- Looking ahead, it is advisable **to strengthen guidance on technical guides with more technical requirements**, provide examples of documentary evidence and clarify the criteria for applicability.

Phase V: Post-Market Monitoring

This phase ended with the majority of the entities carrying out the initial diagnosis of their post-market monitoring

and reviewing the risk analysis that they had previously conducted, to identify indicators. In addition, some of them also took the opportunity to review or design their monitoring processes.

Indicator 1: Task Status Distribution

- **Metric**: Evaluates the progress status of the different tasks included in the post-market monitoring phase, considering the different participating systems as a whole.
- **Analysis**: This indicator allows us to obtain a global view of the degree of progress achieved in Phase V, identifying which tasks are completed, in progress or pending initiation.

This indicator allows us to observe differences between entities, derived from the different levels of previous development of their processes, resources and monitoring mechanisms.

After observing the **Image 15 - Average progress by sector** We can reach the following conclusions:

- The sector of **access to aid and services shows the highest level of progress** (2.78), being very close to the state of full implementation of post-market monitoring processes.
- The medical devices (2.41) and employment (2.39) sectors show a high degree of progress, reflecting an advanced implementation of the technical guides planned during the phase.
- Machinery and critical infrastructure (1.75) and ICT and biometrics (1.69) are the sectors with the **lowest level of progress**, remaining in an intermediate phase of development.

The results show differences in the speed of implementation between sectors, probably associated with the different levels of regulatory, organisational and technological maturity of the participating entities.

Sectors with a higher technological component and greater technical complexity seem to require greater effort to design and implement post-market monitoring mechanisms adapted to their systems.

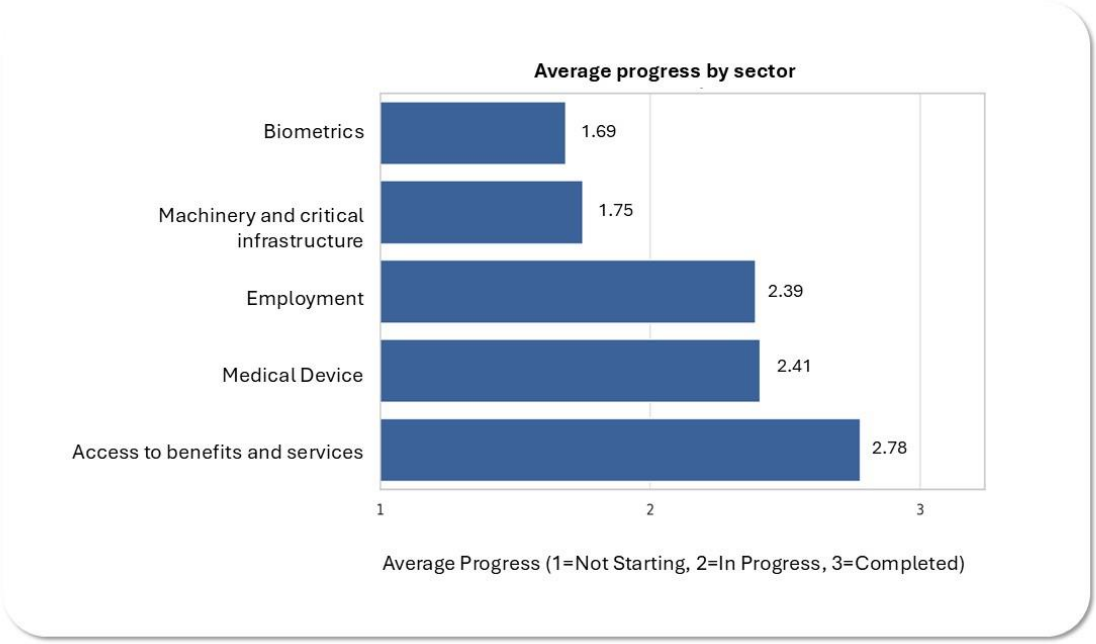

Image 15 - Average progress by sector

After observing the **Image 16 - Average progress by size .** We can reach the following conclusions:

- Medium-sized companies achieve the **highest level of progress** (3.00), completing, on average, all the activities planned during Phase V.
- Large companies (2.36) and startups (2.25) show a high degree of progress, being at an advanced stage of implementation of their surveillance systems.
- Small companies (1.84) show the **lowest level of progress**, mostly remaining in the development phase of the required processes.
- Smaller companies appear to **require more technical support** to complete the implementation of the processes required by the AI Act.

The results suggest that the availability of resources, previous experience in regulatory compliance and organizational capacity significantly influence the pace of implementation of post-market monitoring systems.

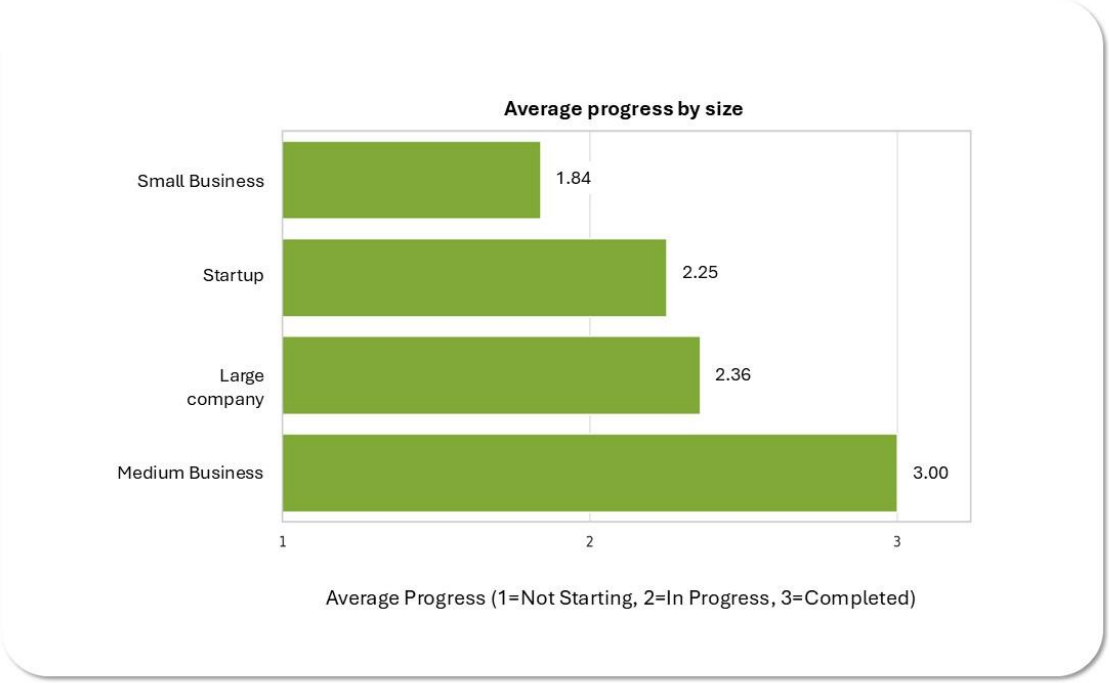

Image 16 - Average progress by size

Indicator 2: Average Progress by Sector and Size

- **Metric**: Assesses the number of entities that have completed, are in progress, or have not started the various tasks in Phase V.
- **Analysis**: This indicator allows the comparison of the degree of progress.

This indicator provides a comparative view of the operational maturity achieved by the entities and their ability to integrate post-market monitoring mechanisms into their internal processes.

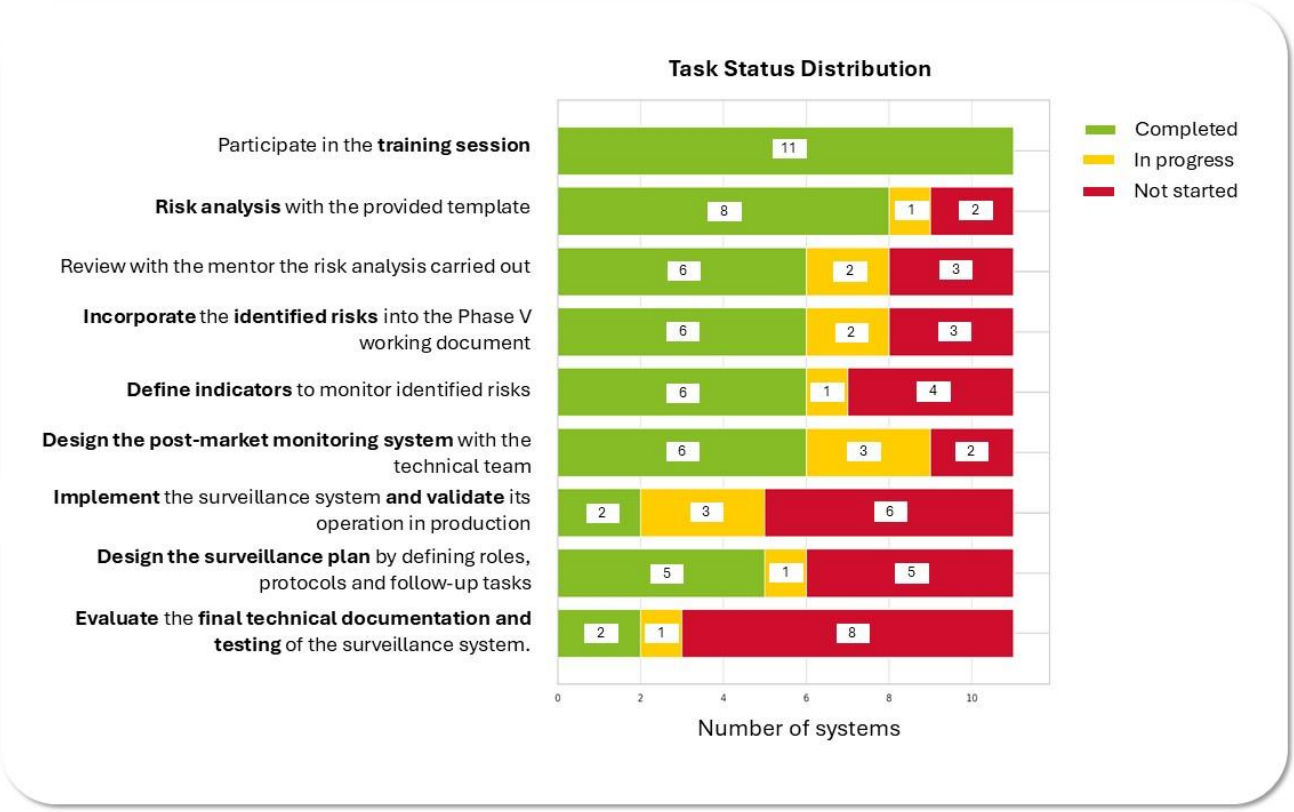

Image 17 - Task status distribution

After observing the **Image 17 - Task status distribution,** the following conclusions can be drawn:

- The **initial training was completed by 100% of the entities**, guaranteeing a common starting point for the development of the phase.
- The **risk analysis activity reached the highest levels of execution** (72.7% of completed tasks), which demonstrates an adequate assimilation of the proposed methodology.
- More than half of the entities (54.5%) completed key tasks such as reviewing the analysis with the mentoring team, incorporating risks into the working document and defining monitoring indicators, consolidating the design of the post-market monitoring system.
- The **design of the post-market monitoring system** showed a high degree of progress, with more than 80% of entities having initiated or completed it, demonstrating a good progression from planning to the structuring of the system.
- The main difficulties are concentrated in the **operational implementation of the system in production** and in the **evaluation of the technical documentation**, activities that register the lowest levels of completion (18.2%), as they require evidence of real operation and a greater degree of organizational maturity.

- The evolution of the tasks reflects a **progressive maturity curve**, where entities advance more easily in the planning and design phases than in those related to the implementation, validation and documentation generation.

The results show that the main challenge of Phase V does not lie in understanding the requirements specified in the technical guides, but in their **operationalization and effective integration** into the technical and organizational processes of the participating entities.

The overall findings of the post-market monitoring phase are presented below:

- Phase V has enabled most entities to **make significant progress in the design and implementation of post-market monitoring processes**, laying the groundwork for continuous monitoring of AI systems.
- The degree of progress achieved is **heterogeneous**, both by sector and by company size, confirming the importance of adapting support to the level of maturity of each entity.
- The sectors of access to aid and services, medical devices and employment have the **highest levels of maturity**, while biometrics and machinery and critical infrastructure show the greatest difficulties in completing implementation.
- In terms of business size, medium-sized companies achieve the highest degree of progress, followed by large companies and *startups*, while **small companies** require a greater effort to complete the planned activities.
- The entities show a high degree of progress in the initial stages of the training, risk analysis and planning process, while activities related to operational implementation, production testing and documentation generation concentrate the **main difficulties**.

The results show that the main barrier does not lie in the understanding of post-market monitoring, but in its **effective integration into the organisational and technical processes of the entities**.

The segmentation of the entities into different itineraries has made it possible to adapt the support provided to the needs and level of maturity of each organization, facilitating the advancement of very diverse profiles.

The methodology applied during Phase V has proven to be adequate to carry out the transition from a theoretical approach to the implementation of a post-market monitoring system aligned with the technical guides.

4. Lessons Learned and Conclusions

The first pilot of the Spanish AI Sandbox has made it possible to generate practical evidence on the application of the requirements provided for in Regulation (EU) 2024/1689 in conditions close to a real environment. Through the participation of high-risk artificial intelligence systems belonging to different sectors and levels of maturity, the initiative has contributed to identifying implementation challenges, validating regulatory support methodologies and generating useful knowledge for the entire artificial intelligence ecosystem.

The accumulated experience confirms the value of **combining training, expert support, practical tools, collaboration spaces and structured evaluation mechanisms** to facilitate the early adaptation of organizations to new regulatory requirements. This conclusion is especially relevant for organizations with less internal compliance capacity, where the technical and documentary complexity of the AI Act can become a real barrier to entry.

It has also generated **relevant lessons** for future regulatory sandbox initiatives and for the authorities in charge of the application and oversight of the AI Act.

Good Practices Identified

The main strength of the Sandbox has been its **eminently practical approach**, based on continuous support for organizations during their regulatory adaptation process. The combination of training, specialized mentoring and applied work has made it possible to understand more precisely the real needs of the participating entities and generate knowledge directly transferable to future support and supervision initiatives.

The initial phase of training allowed entities with very different levels of knowledge and maturity to start from a common base. This stage contributed to homogenizing concepts, clarifying relevant aspects of the technical guides and facilitating the subsequent work of self-diagnosis, adaptation and implementation.

Another element that has been particularly valued has been the **quality of the resources** made available to organizations. Technical guides, training materials, templates, checklists and mentoring sessions have worked as effective tools to reduce interpretative uncertainty, structure the internal work of each entity and identify gaps.

The experience has also shown the **high involvement** of all the actors involved in the pilot, whose collaboration facilitated the early identification of difficulties, the joint search for solutions and the generation of a shared learning environment. In this regard, it is worth noting the autonomous work of the teams of the participating entities, especially during the implementation phase, and the support of the mentoring teams during the self-evaluation phase, which made it possible **to translate the plans into real progress**.

The availability of a **flexible methodology** in the stages in which this was possible, such as the post-market monitoring, with a segmentation by maturity levels, fostered high participation with the result of a robust and well-planned risk analysis.

Likewise, **the support provided by the advisory group of experts** was decisive in addressing complex issues, providing legal and technical certainty, improving the quality of the materials developed and enriching the exchange of knowledge among the participants.

In summary, the experience has confirmed the importance of having strong governance structures, clearly defined responsibilities and effective internal coordination mechanisms.

Challenges for Future Sandbox Initiatives

One of the main challenges observed has been related to the timing of activities. **Initial estimates** did not always reflect the actual effort required to complete certain tasks, especially during the self-diagnosis and implementation phases, whose schedules were not as realistic as they should have been in some cases. Factors such as the technical complexity of the systems, the size of the organizations, their level of maturity or the needs for internal coordination had a significant influence on execution times.

Experience has also shown that **support needs differ significantly depending on the profile of the participating entities**. While large organizations tend to have more specialized resources, their internal processes are usually more complex and can slow down the adoption of certain changes. On the other hand, SMEs and startups have greater organizational agility, although they have fewer capacities dedicated to regulatory compliance and are more sensitive to resource constraints or changes in priority.

In relation to training activities, it is advisable **to extend the duration of certain technical modules and adapt part of the content to the different professional profiles** involved in the application of the AI Act, including technical, legal, compliance, business and risk management profiles, in order to increase the practical applicability of the knowledge acquired. In addition, this training should be

reinforced with specialized support on complex topics, such as cybersecurity, for example, through **ad hoc *sessions*** for the resolution of more advanced questions.

Likewise, opportunities for **improvement have been identified in document and knowledge management systems**. The creation of more agile tools, dynamic repositories of frequently asked questions, updated interpretative criteria and examples of validated evidence would facilitate the reuse of experiences, the traceability of decisions and a more efficient management of queries made by entities.

The challenge in the last phase, once the systems are launched on the market or put into operation, will be the possibility of **transferring the implementation to real production environments**, defining quality indicators and supporting entities with a lower degree of maturity.

Another relevant lesson is the convenience of **strengthening coordination mechanisms** between participants, mentoring teams, expert professionals and competent authorities **through more structured communication channels** and homogeneous procedures for the management and resolution of queries.

From a regulatory perspective, the experience highlights the need to **continue developing practical guides** complementary to the technical guides, incorporating document models, templates, examples and use cases that facilitate a homogeneous application of regulatory requirements.

Likewise, it is necessary to move towards greater **harmonisation between the** AI Act**, the General Data Protection Regulation and the rest of the relevant sectoral regulations,** reducing possible regulatory overlaps and uncertainties through coherent and coordinated reference frameworks.

Finally, it is considered advisable to promote European networks of experts and permanent cooperation mechanisms **between *sandboxes*** that facilitate the exchange of experiences, interpretative criteria, methodologies and good practices, as well as to develop work schemes adapted to the specific needs of large companies, SMEs and *startups*.

Satisfaction Surveys

In order to complement the analysis carried out through the operational indicators of the Sandbox, the perception of the participating entities was collected through two satisfaction surveys carried out at different times throughout the project.

The first survey was carried out **at the end of the training phase** and made it possible to evaluate the usefulness of the content taught, the quality of the training materials and the adequacy of the methodology used.

Subsequently, **during the post-market monitoring phase**, a second survey was carried out aimed at obtaining an overall assessment of the experience of participation in the Sandbox and the set of activities carried out during the pilot. This survey covered the assessment of Phases II to V and also incorporated a specific block of questions aimed at measuring overall satisfaction with the experience of participation in the programme.The questions associated with each phase focused on assessing the perceived usefulness, the effort required and the level of satisfaction achieved, as well as determining the extent to which the Sandbox had contributed to the organizations' understanding and implementation of the measures contained in the technical guides.

For its part, the global satisfaction block aimed to assess the general experience of participation in the Sandbox, analysing the degree to which initial expectations were met and the perceived usefulness of the program in improving the organizations' readiness for regulatory requirements. Likewise, aspects such as the quality of the work environment generated, the role played by the Administration, the usefulness of the support activities - including mentoring, events and resources made available to the participants - and the level of collaboration achieved between the different actors involved were examined.

The results obtained, collected **ANNEX III: Results of the Satisfaction Survey** and presented during the closing ceremony of the AI Sandbox, show a positive assessment of the initiative by the participating entities. They also reflect a general perception of greater readiness to face future adaptation processes derived from the application of the AI Act, as well as a favourable assessment of the training, mentoring and support activities carried out throughout the pilot.

Conclusions

The first pilot of the Spanish AI Sandbox has demonstrated the usefulness of controlled test environments as instruments to **translate the requirements of Regulation (EU) 2024/1689 into practice and generate applied knowledge on its implementation**. The experience developed during the project has made it possible to validate work methodologies, identify regulatory, technical and organisational challenges, and generate practical resources aimed at facilitating the adaptation of organisations to the new European regulatory framework.

Among its most relevant results are the continuous improvement of technical guides, the creation of self-diagnosis and monitoring tools, the consolidation of collaboration mechanisms between public administrations, companies and experts, and the generation of practical evidence on the difficulties and needs associated with compliance with the requirements applicable to high-risk artificial intelligence systems.

The pilot has also contributed to **strengthening the institutional capacities necessary for the future implementation of the AI Act**, providing practical experience both to the participating entities and to the different actors who will perform support, evaluation and supervision functions. In the same way, it has made it possible **to place Spain among the most advanced Member States** in the development of regulatory *sandbox* experiences in the field of artificial intelligence and to generate transferable knowledge for future national and European initiatives.

After the completion of the pilot and the publication of its main results, the challenge now is to consolidate the knowledge acquired, facilitate the reuse of the methodologies developed and continue to promote initiatives that allow organizations to be accompanied in the effective application of the AI Act. In this context, the technical guides, the lessons learned and the accumulated experience constitute a solid basis on which to build the next stage of artificial intelligence governance in Spain and contribute to the development of future regulatory *sandboxes* at European level.

The Sandbox experience also acquires special relevance in the context of the evolution of the Spanish institutional framework for the application of the AI Act. At the closing date of the pilot, the processing of the draft Organic Law on the proper use of governance of artificial intelligence had begun in the Congress of Deputies, among other aspects, at defining the national system of supervision, coordination and governance of artificial intelligence, has begun. In this scenario, the results obtained during the Sandbox, the methodologies developed and the capabilities acquired by the different actors involved constitute an asset of special value for the deployment of future controlled test environments.

In short, the main legacy of the Sandbox does not lie only in the results obtained by the participating entities during the pilot, but also in the **creation of a practical, methodological and institutional knowledge base** that will accelerate the effective application of the Regulation, reduce adaptation barriers for organizations and strengthen Spain's capacity to lead responsible innovation and artificial intelligence governance initiatives at the European level.

Annex I: Relevant Consultations within the Framework of the Project

This annex contains the queries and issues raised during the development of the Sandbox. These were documented for incorporation into the concluding reports and, where appropriate, for their eventual referral to the competent institutions.

Enquiries About the Guides

During the course of the Sandbox, the questions that arose regarding the application of the content of the technical guides were resolved. In this section, the most relevant questions are detailed, together with the responses provided on the basis of the interpretation of the technical guides.

Consultation 1. Existing risk management systems:

If, due to the operational nature, sector, or geography in which my entity operates, I already have a risk management system in place, can that risk management system be reused to comply with the risk management guide?

According to the guide, it is possible to reuse the risk management systems that the entity has already had to prepare to comply with other sectoral or regulatory requirements.

However, it is particularly critical to ensure that such a risk management system is aligned with the specific measures in the guide, in particular with regard to aspects such as health, safety or fundamental rights. Additional technical elements, such as the management of the AI system's lifecycle, should also be taken into account.

Consultation 2. Statistical significance tests:

Why is it necessary to perform statistical significance tests to justify the accuracy metrics and objective functions?

According to the guide, when we develop an AI system, especially a high-risk one, we must demonstrate that the accuracy metrics we choose (such as Accuracy, F1-Score, AUC-ROC...) and the objective functions (the criteria that the model optimizes) are:

- Suitable for the use case
- Reliable.
- Technically justified.

The problem:

The results of an AI system (accuracy, recall, etc.) can vary due to factors such as:

- Randomness in the data.
- Sample size.
- Variability inherent in training.

Therefore, one metric may seem better than another, but that difference could be the result of chance and not a real improvement.

What do significance tests contribute?

Statistical significance tests allow us to:

- Determine whether a difference between metrics or models is real and not random.
- Justify, on a mathematical basis, why we choose a specific metric as the accuracy benchmark.
- Avoid decisions based on irrelevant or insufficiently robust results.
- Meet the guide's requirement to demonstrate that the system achieves adequate and controlled levels of accuracy.

Practical example

Let's say we compare two accuracy metrics:

- Accuracy A: 87%
- Accuracy B: 89%

At first glance, B looks better. But is that difference significant or is it just random variation?

- With a significance test, we can calculate whether B's improvement is statistically relevant.
- If the test confirms that the difference is significant (e.g., $p < 0.05$), we have an objective basis for choosing B.
- If it is not significant, we cannot justify that B is really better.

Direct relationship with the guide

The technical guide requires that:

- High-risk AI systems achieve and document an appropriate level of accuracy.
- The metrics and objective functions are justified and their performance is traceable.

Significance tests are the standard tool to meet these requirements and provide technical and statistical evidence.

Consultation 3. Unit, stress and integration tests

Why should unit, stress, and integration tests be performed to measure accuracy?

- Unit tests evaluate the accuracy of individual components of the AI system (e.g., preprocessing modules, inference algorithms, specific layers). In addition, they allow errors or accuracy degradations to be detected in early stages of development. They ensure that each part correctly contributes to the required overall level of accuracy.
- Stress tests evaluate accuracy under extreme or adverse conditions (variations in input data, changes in the operating environment, extreme or high-load situations, ...). They are key to checking whether accuracy is maintained in non-ideal scenarios, as required in the technical guides in terms of robustness.
- Integration tests verify the accuracy of the entire system once the various components are integrated. In addition, they detect incompatibilities or degradations that do not appear in unit tests and ensure that the system, as a whole, maintains the promised levels of accuracy.

Justification based on the guide

- Continuous monitoring: accuracy levels must be known and monitored at all stages, not just in final validation.
- Technical evidence: Tests provide documented and verifiable records demonstrating that the required levels of accuracy are achieved
- Failure prevention: They detect deviations or degradations before the system reaches production, aligning with the safety and reliability principles of the AI Act
- Risk mitigation: They identify situations where the system loses accuracy, allowing safeguards or operating limits to be defined.

Planning and carrying out these tests is necessary because: it helps ensure that the system maintains the required levels of accuracy under any foreseeable circumstances. In addition, it prevents errors or degradations in components from impacting overall accuracy. It supports compliance with the technical guides by ensuring an adequate, traceable and controlled level of accuracy throughout the life cycle. In addition, it provides the technical evidence necessary to pass conformity assessments and regulatory audits.

Consultation 4. Shutdown in the event of a disaster

The question was raised as to whether the post-market monitoring technical guide requires that, like the provider, the deployer should also be able to shut down the system in the event of a catastrophe.

According to the guide, for the purposes of implementing paragraphs 1, 2 and 3 of the AI Act , the high-risk AI system shall be provided to the deployer in such a way that natural persons entrusted with human oversight can, as appropriate and in a proportionate manner: (a) adequately understand the relevant capabilities and limitations of the high-risk AI system and properly monitor its operation, for example with a perspective of detecting and addressing anomalies, operational problems and unexpected behaviour; (b) be aware of the potential tendency to automatically or excessively rely on the outputs generated by a high-risk AI system ('automation bias'), in particular in the case of systems that are used to provide information or recommendations for natural persons to make a decision; (c) correctly interpret the outputs of the high-risk AI system, taking into account, for example, the available interpretation methods and tools; (d) decide, in any particular situation, not to use the high-risk AI system or to discard, invalidate or reverse the outputs generated by it; (e) intervene in the operation of the high-risk AI system or interrupt the system by pressing a stop button using a similar procedure that allows the system to stop safely.

Therefore, it is mandatory for the deployer to have this shutdown or interruption capability.

Consultation 5. Information to share about datasets

What is the scope of the information to be provided to deployers regarding the datasets of a system?

The technical documentation guide mentions certain data issues to be documented, but the purpose of this documentation is to support the conformity assessment, it is not designed for the entire file to be shared with the deployer. On the other hand, the transparency guide establishes an obligation to share "*where appropriate, specifications relating to input data, or any other relevant information*

in relation to the training, validation and testing datasets used, taking into account the intended purpose of the high-risk AI system,".

If we cross-reference the transparency guide with the technical documentation guide *"where applicable, the data requirements, in the form of fact sheets describing the training methodologies and techniques, as well as the training datasets used, and including an overview of those datasets and information about their source, scope and main characteristics; how the data were collected and selected; labelling procedures (e.g. for supervised learning) and data cleansing methodologies (e.g. anomaly detection);".* Therefore, it is interpreted that the information provided should not be more detailed than that required for the technical documentation. Accordingly, the obligation would be fulfilled by providing a general description of the datasets, their origin, scope and main characteristics.

Consultations Referred to the Advisory Group of Experts

In addition, as a result of the concerns raised by the participating systems, a series of issues were detected whose analysis and response were particularly complex. In order to assess these questions and give their corresponding response a greater wight, certain questions were raised to the Advisory Group of Experts. In this section, the most relevant issues that were raised to the Advisory Group of experts for analysis are detailed.

Consultation 1. Delineating the perimeter of an AI system

Background

Within the framework of the Sandbox, the need arose to interpret and apply the definition of "AI system" provided for in Article 3.1 of the AI Act in real cases. During the analysis of specific systems and applicable safeguards, a recurring difficulty was identified in determining which components and functionalities should be considered part of the AI system and which should be considered external or environmental elements.

The consultation is based on the definition contained in Article 3.1 of the AI Act and the European Commission's guides on the main characteristics of AI systems, with the aim of establishing technical criteria that allow the perimeter of an AI system to be delimited in a consistent and practical manner from an engineering perspective.

Question raised

- How should the perimeter of an AI system be determined to identify which components and functionalities form part of it, considering the definition in Article 3.1 of the AI Act and the guides of the European Commission?

In particular, the question concerns how to establish whether certain modules, interfaces, explanation mechanisms, data input or other elements should be considered integrated into the AI system.

Reply

The definition of "AI system" contained in Article 3.1 of the AI Act does not set out an explicit notion of the "perimeter" of the system. Consequently, its delimitation must be carried out on a case-by-case basis, taking into account the architecture, functionality and intended purpose of the particular system, as it cannot be determined using automatic criteria or exhaustive lists of components.

The central technical criterion for defining the perimeter is the **inference capacity of** the system, understood as the ability to generate, from input data, outputs such as predictions, content, recommendations or decisions that may influence an environment.

In general, the system includes those components that are essential to this inference capacity, including machine learning models, both during training and once deployed, logical or symbolic inference engines, expert rule systems that perform non-trivial reasoning and data processing flows directly linked to the inference process. On the other hand, components that only execute static, deterministic rules, conventional heuristics without adaptive learning or reasoning, and reference models of a merely trivial nature are excluded.

To delimit the perimeter of the system, it is advisable to follow a structured methodology that includes, at least, the following actions:

- Identify the minimum system, mapping the complete flow from data input to output generation through the inference process.
- Determine the intended purpose of the system and the context of use, including both intended use and reasonably foreseeable misuse.
- Develop a context diagram that identifies the system, its input and output interfaces, and interactions with people, data sources, APIs, actuators, and other external systems.

- Analyze the end-to-end functional flow (input, data preparation, inference, post-processing, presentation or action, and monitoring) to identify the components that influence the final result.
- Apply a functional test of whether a component belongs to the system, considering that a component is part of the perimeter when its elimination prevents the system from performing inference, significantly reduces its autonomy or prevents it from achieving the intended purpose through its influence on its environment.

Likewise, the delimitation of the perimeter must consider the different phases of the system's life cycle. In the construction phase, the components necessary for the training and tuning of the model may be part of the system, while in the use phase, the deployed model itself, the processing of inputs and outputs and, where appropriate, the adaptation mechanisms that are essential to the inference capacity may form part of the system.

Consequently, the determination of the perimeter of an AI system requires a technical analysis of its specific architecture and operation, based on the elements provided for in Article 3.1 of the AI Act and the European Commission guidelines, as it is not possible to establish a uniform delimitation applicable to all systems. This interpretation is merely for information purposes and does not replace or further define the applicable regulations.

Consultation 2. Using general-purpose AI systems in high-risk contexts

Background

Within the framework of the regulatory sandbox for artificial intelligence, several questions arose about the application of Article 25.1.c) of the AI Act when an organization uses a general-purpose AI system for a purpose considered to be high risk. In particular, it was raised whether such use entails the acquisition of the status of provider of the high-risk AI system and what practical implications arise from complying with the obligations provided for in the AI Act .

Question raised

1. If a company or a Public Administration uses a general-purpose AI system, such as ChatGPT, for a purpose considered to be high-risk (for example, CV screening according to Annex III.4.a), does it become a provider of that high-risk AI system?
2. If so, is it realistic for an organization to assume this status in the light of the requirements of the AI Act and the limitations inherent in these systems? What approach should I take to comply with the applicable obligations?

Reply

Article 25(1)(c) of the AI Act provides that any operator shall be considered a provider of a high-risk AI system when it modifies the intended purpose of a general-purpose AI system placed on the market or put into service so that it becomes a high-risk system in accordance with Article 6 of the AI Act .

This modification of the intended purpose is not limited to technical alterations to the system but also includes its use for a purpose other than that intended by the original provider. Therefore, when a company or a Public Administration uses a general-purpose AI system for a high-risk purpose not initially contemplated, it acquires the status of a provider of the high-risk AI system with respect to that specific use.

As a result, it will have to assume the obligations applicable to providers of high-risk AI systems, including carrying out the conformity assessment, registering the system, preparing the technical documentation, implementing a risk management system, complying with transparency obligations and adopting human oversight measures. It shall also comply with the obligations set out in Article 16 of the AI Act, including compliance with safety and quality requirements, maintaining documentation and records, obtaining the CE marking, taking corrective action and cooperating with competent authorities.

The original provider will no longer be considered the provider with respect to that specific use, without prejudice to its duty to cooperate by providing the necessary information, unless it has expressly established limitations on the use of the system.

From a practical point of view, assuming the status of provider can be particularly demanding, as it requires sufficient resources, technical and legal capabilities and multidisciplinary teams to ensure compliance with the AI Act. This can make it difficult to use general-purpose AI systems in high-risk contexts, especially by small and medium-sized enterprises. However, this requirement responds to the purpose of the AI Act of preventing non-certified systems from being used in high-risk applications without the corresponding guarantees.

As a general criterion, it is recommended to first assess the possibility of using solutions already marketed as high-risk AI systems that comply with the AI Act. When this is not possible and a general-purpose AI system is chosen, it is advisable to involve the provider from the outset and to have the necessary information on the operation of the system to facilitate compliance with regulatory obligations.

In addition, the organisation should strengthen its internal AI governance, assigning clear responsibilities, training staff, implementing continuous monitoring procedures, recording incidents and documenting all actions related to compliance with the AI Act. From a technical point of view, additional risk mitigation measures can also be adopted, such as the incorporation of control mechanisms (*guardrails*), filtering of results through additional layers of *software*, human review of decisions (*human-in-the-loop*) and the performance of external audits.

Consultation 3. Determining the roles of operators in AI systems in accordance with the AI Act

Background

Within the framework of the AI regulatory Sandbox, various queries arose about the correct attribution of the roles provided for in the AI Act when different public or private entities are involved in the development, acquisition, commercialization and use of AI systems. In particular, different scenarios of collaboration between Entity **A** (which may be a territorial public administration, a public agency, a public sector foundation or a public owned company) and Entity **B** (which may be a public entity or a private company) were analysed to determine which operator assumes the status of provider, deployer or distributor, as well as whether there may be multiple providers for the same system.

Question raised

1. How should the roles of provider, deployer, and distributor be determined when different entities are involved in the development, commercialization, or use of an AI system?
2. Is it possible for there to be several providers of the same AI system and, if so, how are the responsibilities arising from the AI Act distributed?

Reply

The attribution of the roles of provider, deployer and distributor must be carried out in accordance with the definitions of the AI Act and the functions effectively performed by each operator.

Generally speaking, the **provider** is the operator who develops an AI system or has it development and places it on the market or puts it into service under its own name or brand; the **deployer** is the operator who uses the system under their its authority; and the **distributor** is the operator who participates in the supply chain without being a provider or deployer.

In the cases analysed, the determining element for identifying the provider is the entity under whose name the system is marketed or put into service. When an AI solution is developed for an

organization, the organization will be the provider if the system is placed into the market or put into service; otherwise, the provider will be the entity that developed it, and the user organization will act as the deployer. Supervising the development or involvement of third parties through outsourcing does not, by itself, alter this distribution of roles.

On the other hand, the role of distributor can only correspond to operators who intervene in the supply chain without acting as providers or deployers.

In addition, there may be multiple providers of the same AI system, for example, in joint development projects or where the system is marketed under multiple brands. While they may contractually distribute certain tasks, such as preparing the technical documentation, risk management or CE marking, they all retain responsibility before the competent authorities for compliance with the obligations under the AI Act.

Consultations on the Rules of Procedure

Within the framework of the pilot, the participants raised a series of questions that fell outside the scope of the Sandbox pilot itself, as they did not concern the technical guides as such. These questions generally concerned aspects for which the participants considered that greater clarification or specific guidance from the different competent bodies was necessary. In this regard, the most relevant issues identified and requiring further clarification have been compiled.

Consultation 1. Technical Documentation Requirements

Concerns were raised regarding the lack of clarity in the regulations regarding the requirements applicable to the technical documentation of AI systems. In particular, participants expressed doubts about:

- The level of detail that the technical documentation must contain.
- The scope of the **concept** of **"accessibility"** of information and whether, in certain cases, it involves providing elements such as the source code.
- The minimum contents that must be included to ensure compliance with regulatory obligations.

This issue was identified as one of the main challenges for the participating entities due to the absence of consistent criteria for its practical application.

Consultation 2. General ambiguity and absence of sectoral criteria

Participants stated that the regulations have a high degree of generality, which makes their practical application across different sectors of activity difficult. In particular, concerns were raised about:

- The lack of definition of the functions to be performed as part of human oversight.
- The absence of references to specialised profiles according to the field of application (healthcare, legal, educational, etc.).
- The lack of standardised templates for technical documentation and other documentary requirements.
- The limited inclusion of practical examples and specific criteria for each sector.

Consultation 3. Handling incremental updates

Participants raised doubts about the regulatory treatment of updates to AI systems, especially when they evolve through retraining, the addition of new functionalities or the expansion of their use cases.

In particular, questions were raised as to when a modification should be considered substantial and in what cases a new evaluation of the system is necessary.

Consultation 4. Processing of synthetic data

Doubts were raised about the regulatory treatment of synthetic data and the implications that their use may have from a regulatory compliance perspective.

In particular, it was pointed out that, even though these data may be excluded from certain obligations, they may reproduce biases present in the original data or generate other relevant risks.

Consultation 5. Risk definitions and categories

Participants reported uncertainty regarding certain definitions contained in the AI Act.

In particular, doubts were identified about:

- The interpretation of concepts such as "high-risk AI" or "general-purpose AI".
- The practical application of expressions such as "endangering life or health" or "interfering with social or economic activities".
- The classification of certain use cases that do not clearly fit into the categories provided for by the regulations.

Consultation 6. Administrative burden arising from compliance

Participants expressed concern about the administrative burden involved in complying with the obligations under the AI Act, especially for small and medium-sized companies and *startups*.

Likewise, the risk that these obligations may generate a competitive disadvantage against organizations with greater resources and may affect innovation was highlighted.

Consultation 7. Adapting regulations to technological developments

The difficulty of ensuring that the regulation can adapt to the rapid development of artificial intelligence technologies was raised, especially given the continuous emergence of new models, capabilities and use cases.

Consultation 8. Specific regulation of generative AI

Participants noted the lack of specific provisions for generative AI systems, foundation models, and AI agents.

Specific risks related to intellectual property, misinformation, bias, healthcare, customer service, and transparency regarding the data used to train these models were also identified.

Consultation 9. Monitoring and implementation of the AI Act

Participants raised doubts about the envisaged institutional model for the monitoring and implementation of the AI Act.

In particular, the need to clarify the competences and coordination mechanisms between the different authorities responsible for their implementation was noted.

Consultation 10. Security and cybersecurity

The lack of specificity in the regulations regarding the technical security and cybersecurity requirements applicable to AI systems was identified.

In particular, the need for minimum criteria regarding architecture, protection against incidents and required security measures was raised.

Consultation 11. Legal liability in the use of AI systems

Participants expressed doubts about the distribution of responsibilities between developers, providers and users of AI systems, especially when using solutions developed by third parties.

In particular, issues were raised regarding the documentation to be required from providers, compliance verification mechanisms and the allocation of responsibilities in the event of a system failure.

ANNEX II: Analysis of results by Entity

This annex presents, compiled in tables, the performance of each system during the development of the AI Sandbox. The following tables include the results of each participating entity, systematically organized according to the sector, the AI use case evaluated and an executive summary consolidating its evolution throughout the project.

The information incorporated in these tables has been treated in accordance with the confidentiality framework agreed upon during the execution of the AI Sandbox pilot. For the preparation of this document, the participating entities have expressly authorised the use and disclosure of the information contained herein, guaranteeing that its processing is carried out in a limited, proportional and respectful manner in accordance with the applicable confidentiality and privacy obligations.

Medical Devices

Table 7 - Participating entity Bit&Brain Technologies, S.L

CL@UDIA	
Participating Entity	BIT&BRAIN TECHNOLOGIES, S.L.
Sector	Medical devices
MSA	AEMPS
Collaborators	N/A
Description	CL@UDIA is a platform designed for the processing and analysis of multimodal data obtained through wearable devices in the home environment, combining physiological and behavioural information to offer a comprehensive view of the user's condition.
Development	• The pilot of the AI Sandbox has been **a significant** step **forward in bringing CL@UDIA into line with the** future **requirements of the AI Act.** • **The quality area stands out as the most mature**, with the implementation of most of the measures envisaged and a solid basis for regulatory compliance. • In areas such as risk management, human oversight, transparency, and accuracy, implementation remains partial, although **frameworks and procedures have already been defined for future** consolidation.

	• **The monitoring plan has become a central element of this evolution**, incorporating specific risk indicators, defined responsibilities and a clear periodicity for reviews. • Overall, the pilot has contributed to consolidating a more robust, orderly governance model aligned with future regulatory requirements.

Table 8 - Participating entity Made Of Genes, S.L.U

MoGIA-MetCare: AI for the Prevention of Metabolic Diseases	
Participating Entity	MADE OF GENES, S.L. U
Sector	Medical devices
MSA	AEMPS
Collaborators	Genomcore SL
Description	The purpose of the Made of Genes MetCare system is to assess the risk of developing metabolic diseases, such as diabetes or obesity, in a personalized way. To this end, it integrates clinical data from the anamnesis together with analytical and genetic markers, allowing heterogeneous and complex biomedical information to be transformed into an individualised risk assessment. This makes it possible to orient prevention towards specific risk profiles, facilitating the early identification of people with a greater predisposition to developing metabolic diseases and supporting preventive strategies adapted to the clinical, analytical and genetic characteristics of each user.
Development	• **Progress towards compliance with the AI Act:** The AI Sandbox has been a significant step forward in adapting to future requirements, especially due to its connection with the healthcare sector and to in vitro diagnostic medical devices. • **Positive regulatory progression:** The company, which already had a high level of knowledge of standards and regulatory alignment, has made special progress in quality, risk management and cybersecurity, although technical documentation and serious incident reporting require further consolidation. • **Process consolidation:** The main progress has been the formalization of frameworks, procedures, and basic documentation to sustain compliance with the system. • **Next steps:** The Sandbox has made it possible to identify more clearly the degree of alignment with the AI Act, and

	the following phases should focus on completing the technical documentation, strengthening the management of major incidents and fully deploying the defined procedures.

Table 9 - Participating Entity: Tucuvi Care SL

Tucuvi Health Manager	
Participating Entity	Tucuvi Care SL
Sector	Medical devices
MSA	AEMPS
Collaborators	Google Cloud EMEA Limited Indra Soluciones Tecnologias de la Informacion, S.L.U
Description	Tucuvi Health Manager (THM) is medical *software* that automates the clinical monitoring of patients through phone calls managed by LOLA, a virtual assistant based on conversational AI.
Development	Information not available under the confidentiality framework set forth in this Sandbox.

Access to Benefits and Services

Table 10 - Participating Entity Dedomena Artificial Intelligence, SL

AI system for credit rating, solvency analysis and default risk of individuals based on their bank transactional data	
Participating Entity	Dedomena Artificial Intelligence, SL
Sector	Access to benefits and services
MSA	BdE
Collaborators	BBVA Ikualo Zien Ideas
Description	The main purpose of the artificial intelligence system is to score an individual's credit and to estimate the risk of applying for credit or financing, as well as to extract insights. The history of financial transactions is analysed, using proprietary AI models based on neural networks with architectures

	developed after years of research, as well as machine learning models. It has innovative features such as the explainability of results, anonymization with proprietary technologies of the data, as well as offering a more complete response that includes default prediction, estimation of future balance and insights about the individual.
Development	• **Dedomena started from a solid base** in precision, solidity, technical documentation and record-keeping, but **the pilot has made it possible to bring these practices closer** to the formal requirements of the Artificial Intelligence AI Act . • **The greatest progress** has been made in the **formalisation of risk management and quality management systems**, which have become the structural axes of the operating model. • **The technical documentation and traceability** have been significantly reinforced **,** incorporating greater detail on the architecture, data, operation, limitations and explainability of the system. • **Post-market monitoring has evolved towards a structured and operational model in production**, based on continuous monitoring, quantifiable indicators and defined action protocols.

Table 11 - Participating Entity: Fundación Centro de Investigación en Tecnoloxías da Información e as Comunicacións de Galicia (FCITICG)

Intelligent transcription processing system for emergency communications at the 112 Galician operations centre	
Participating Entity	Fundación Centro de Investigación en Tecnoloxías da Información e as Comunicacións de Galicia (FCITICG)
Sector	Access to benefits and services
MSA	AESIA
Collaborators	Agencia para a Modernización Tecnolóxica de Galicia Cluster TIC Cinfo contenidos informativos personalizados
Description	The Intelligent Transcription Processing System for Emergency Communications at the AXEGA 112 Galicia Operations Centre aims to transcribe speech to text based on ASR (Automatic Speech Recognition), processing emergency telephone calls in both Spanish and Galician, used for data verification, sometimes for judicial purposes. The system allows automatic

	transcription in real time, editing and saving in dialogue format with separation of interlocutors.
Development	• **Evolution towards compliance with the AI Act** , starting from an initial base more developed in technical requirements such as precision and robustness. • The main **qualitative leap** has occurred in the consolidation of processes linked to **the risk management system, data governance, human oversight and technical documentation**. • **In terms of post-market monitoring**, progress has been made in the definition of a surveillance plan through the creation of specific indicators and a formal protocol that allows FCITICG to anticipate, monitor and mitigate risks associated with bias or automated discrimination.

Machinery and Critical Infrastructure

Table 12 - Participating Entity: Airbus Operations SLU

Operator Assistant for the Use of SOIs	
Participating Entity	AIRBUS OPERATIONS S.L.U.
Sector	Machinery
MSA	AESA
Collaborators	Stack Labs SAS Siren Asociación Industrial para el Impulso de la Economía del Dato y de la Inteligencia Artificial – INDESIA
Description	The system is a conversational assistant designed to assist operators during aircraft manufacturing. Its objective is to allow workers to consult, through natural language, the technical documentation used to perform complex tasks, which includes text, images and diagrams.
Development	• **Mature evolution towards compliance with the AI Act :** the pilot has allowed the engineering team to complement its initial focus on technical performance with a compliance-oriented perspective, integrating Ethics-by-Design principles. • **Process consolidation:** The main qualitative advance has been the eradication of informal processes in favour of governed methodologies. Highlights include the consolidation of the "*Human-in-the-Loop*" protocol as an

	unbreakable security barrier, the technical formalization of data lineage, and the creation of continuous transparency channels.

Table 13 - Participating Entity Imotion Analytics, SL

GuardianMobility AI - Critical Infrastructure Monitoring	
Participating Entity	IMOTION ANALYTICS, SL
Sector	Critical infrastructure
MSA	AESIA
Collaborators	N/A
Description	GuardianMobility AI is an advanced real-time monitoring system designed for the security and optimization of critical transportation infrastructures. Its architecture combines local and cloud processing, integrating convolutional neural networks and hybrid AI models for the early detection of incidents and anomalies in vehicles, crowds and strategic spaces.
Development	• **Evolution towards regulatory compliance:** Imotion Analytics' participation in the AI Sandbox has been positive and transformative, allowing the system to progress towards compliance with the applicable technical guides, especially starting from a solid foundation in accuracy and quality management. • **Improvement of technical documentation:** During the implementation phase, the main progress has focused on the reinforcement and improvement of the system's technical documentation. • **Post-market monitoring:** Significant progress has been made in defining a risk-based surveillance plan, with specific indicators and a formal protocol to anticipate and mitigate risks of bias or automated discrimination.

Table 14 - Participating Entity Loradix Red, S.L

Cybersecurity Enhanced with AI	
Participating Entity	LORADIX RED, S.L
Sector	Critical infrastructure
MSA	AESIA
Collaborators	SG2 Group
Description	AI-Enhanced Cybersecurity in Loradix
Development	The entity voluntarily withdrew from the Sandbox after completing only the training phase. This circumstance was notified to SEDIA, which confirmed it through the corresponding resolution.

Biometrics

Table 15 - Participating Entity Herta Security, S.L

BioSurveillance	
Participating Entity	Herta Security, S.L
Sector	Biometrics
MSA	AEPD
Collaborators	N/A
Description	BioSurveillance is a high-performance video surveillance solution that works by facial recognition (using biometrics). It is specially designed for the simultaneous identification of subjects in crowded and highly changing environments and is specially designed for use in environments with large moving crowds, such as airports, metro or train stations, shopping centres, sports stadiums or urban centres.
Development	• **Evolution towards compliance with the AI Act:** Herta started from a solid base and a high level of initial maturity in most of the measures of the technical guides, and the Sandbox has made it possible to adapt these good practices to the formal requirements of the AI Act, especially in risk analysis, quality management and technical documentation.

	• **Process consolidation:** The pilot has strengthened the system's key processes, in particular the AI-specific quality management system, clarifying roles, responsibilities and procedures from the already existing integrated system. • **Documentation, governance and monitoring:** There have been significant advances in technical documentation, data governance, traceability, transparency and life cycle control, as well as in post-market monitoring through defined indicators, documented procedures and explicit responsibilities.

Table 16 - Participating Entity Veridas Digital Authentication Solutions, S.L.

Veridas Age Verification	
Participating Entity	Veridas Digital Authentication Solutions, S.L.
Sector	Biometrics
MSA	AESIA
Collaborators	Universidad Autónoma de Madrid
Description	Veridas Age Verification is an artificial intelligence system designed to verify the age of users in vending machines, ensuring regulatory compliance in the sale of regulated products such as tobacco or also pending regulations for vapers. Its purpose is to guarantee that only those over 18 years of age can purchase these products, protecting minors and promoting social responsibility.
Development	• **Evolution towards compliance with the AI Act:** The Sandbox has been key to adapting existing good practices to the requirements of the AI Act, especially in terms of data quality and risk analysis. • **Process consolidation:** The pilot has reinforced the internal processes linked to the categories prioritized in the adaptation plan, consolidating a more structured model aligned with regulatory compliance. • **Monitoring and scalability:** Post-market monitoring has evolved into a formalized model, defined indicators and responsibilities, providing the team with tools and procedural frameworks to scale the system in accordance with the AI Act .

Employment

Table 17 - Participating Entity Adevinta Jobs, S.L.U.

Registration Rankings via HyM (How you Match)	
Participating Entity	ADEVINTA JOBS S.L.U.
Sector	Employment
MSA	AESIA
Collaborators	InfoJobs
Description	The purpose of the system is to calculate and score the compatibility between candidates' CVs and job offers. This currently allows companies to automate the first CV screening and order those registered according to their degree of suitability for the position.
Development	• **Adaptation to the AI Act:** The Sandbox environment has made it possible to adjust Adevinta's previous solid technical foundation to the formal requirements of the Artificial Intelligence Act, improving risk analysis, quality and documentation. • **Formal quality and risk management:** There has been a notable improvement in moving from informal management to a structured system with defined roles and protocols. • **New post-market monitoring:** Monitoring has evolved from measuring only business metrics to a risk-focused plan, using specific indicators to anticipate and mitigate identified risks.

Table 18 - Participating Entity Shakers Global Solutions SL

Shakers AI Matchmaker	
Participating Entity	Shakers Global Solutions SL
Sector	Employment
MSA	AESIA
Collaborators	BBVA
Description	The Shakers AI Matchmaker system is a matching system that uses personalized recommendations to connect freelance talent with business projects, through a multi-stage recommendation system that calculates a "match" (talent-project fit) that identifies the three best candidates best suited for each role in the team that will carry out the project.
Development	• **Regulatory alignment**: Going through the Sandbox has been key to meeting the requirements of the AI Act, achieving comprehensive technical documentation and data governance. • **Robust risk management**: Evolved from a generic control to a granular and auditable system, with defined roles and impact analysis specific to high-risk systems. • **Areas for improvement**: The need to improve bias detection and the data recording model has been identified. • **Proactive post-market monitoring**: Monitoring has shifted from focusing only on algorithm performance to a risk-based plan, designed to prevent bias and employment discrimination.

ANNEX III: Results of the Satisfaction Survey

This annex is entirely dedicated to showing the **results obtained in the satisfaction survey for the AI Sandbox**.

During the training phase, the survey was addressed to participants attending the training sessions, with the aim of evaluating their assessment of the different training sessions delivered. Through a series of questions, the **degree of satisfaction with the content and materials provided** was analysed, paying special attention to whether they had met the expectations generated by the initiative and whether the knowledge acquired had been useful in **facilitating the adaptation of their systems to the AI Act.**

The main questions asked, with the aim of standardizing the aspects to be evaluated among the different requirements proposed by the technical guides, are:

- What has been your level of understanding of the content taught?
- Do you think that the content taught had sufficient depth?
- Do you consider that the audiovisual support material shown during the training was adequate and useful in facilitating the understanding of the contents?
- Do you consider that the topics discussed were relevant to the application of the AI Act to your system?

Image 18 - Heat Map overall Satisfaction by requirements

The analysis **Image 18 - Heat Map overall Satisfaction by requirements** shows that human oversight is the requirement that received the highest rating in the dimensions of understanding and relevance, reflecting a widely consolidated understanding of its role within the regulatory framework and its importance in ensuring the safe, ethical and compliant use of AI systems.

Likewise, data and conformity assessment consistently stand out across the four dimensions evaluated, showing that the participants not only consider these requirements especially relevant, but also positively assess the depth with which they were covered and the quality of the materials used during the training sessions.

Similarly, the requirements related to the risk management system, robustness, accuracy and cybersecurity and post-market monitoring received high ratings, especially in terms of the quality of the support material and the depth of the content taught. These results indicate that the training activities have facilitated the understanding of concepts of greater technical complexity and have made it possible to contextualize their application within the obligations established by the AI Act.

Overall, **the results show a very positive assessment of the training sessions and a high level of understanding of the technical guides**. The analysis of the average scores constitutes an objective indicator of the effectiveness of the training provided and allows the identification of those guides that were most effectively understood by the participants, providing useful information to guide future training and continuous improvement actions.

If we analyse each of the dimensions individually, the results are reflected in the following graphs. To facilitate the interpretation of the results, the average scores obtained for each requirement have been grouped into three ranges: **Very important** (≥ 4.50), which reflects a very high level of assessment; **Important** (between 4.00 and 4.49), which indicates a positive assessment; and **Moderate** (< 4.00), which identifies those aspects with the greatest room for improvement. This classification makes it easier to visually identify the requirements that have reached a higher level of understanding among the participants.

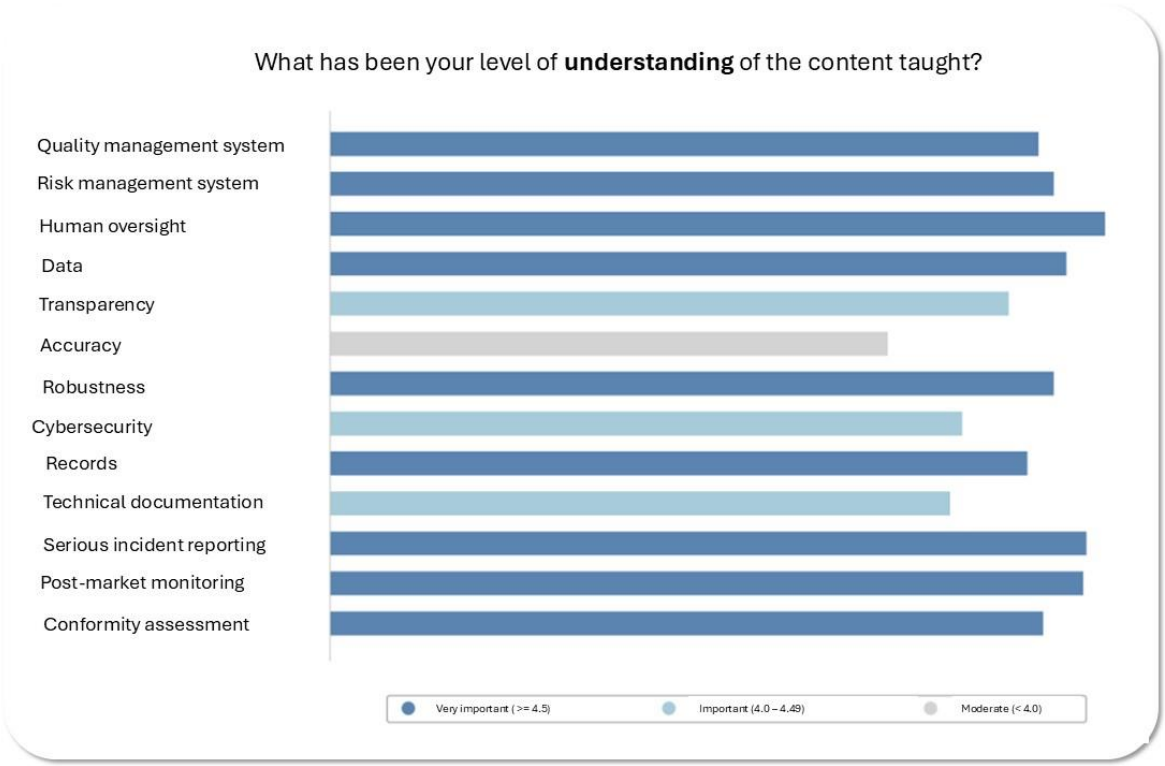

Image 19 - Level of understanding of the contents

In relation to the level of understanding of the contents taught, as shown in the **Image 19 - Level of understanding of the contents**, the results reflect a **very positive overall assessment.**

Most of the guides are in the highest range, which shows that the training sessions facilitated their proper understanding. In general terms, the scores show a standardized pattern, without significant differences between the different areas evaluated.

Among the **best-rated guides** are human oversight, serious incident reporting and post-market monitoring, although the rest classified as "Very Important" also achieve high scores, evidencing a good assimilation of the content taught.

On the other hand, transparency, cybersecurity and technical documentation are in the "Important" range, maintaining a positive rating, while accuracy is the only requirement classified as "Moderate". This result suggests that more technical or specialized content presented a greater difficulty in understanding for the participants.

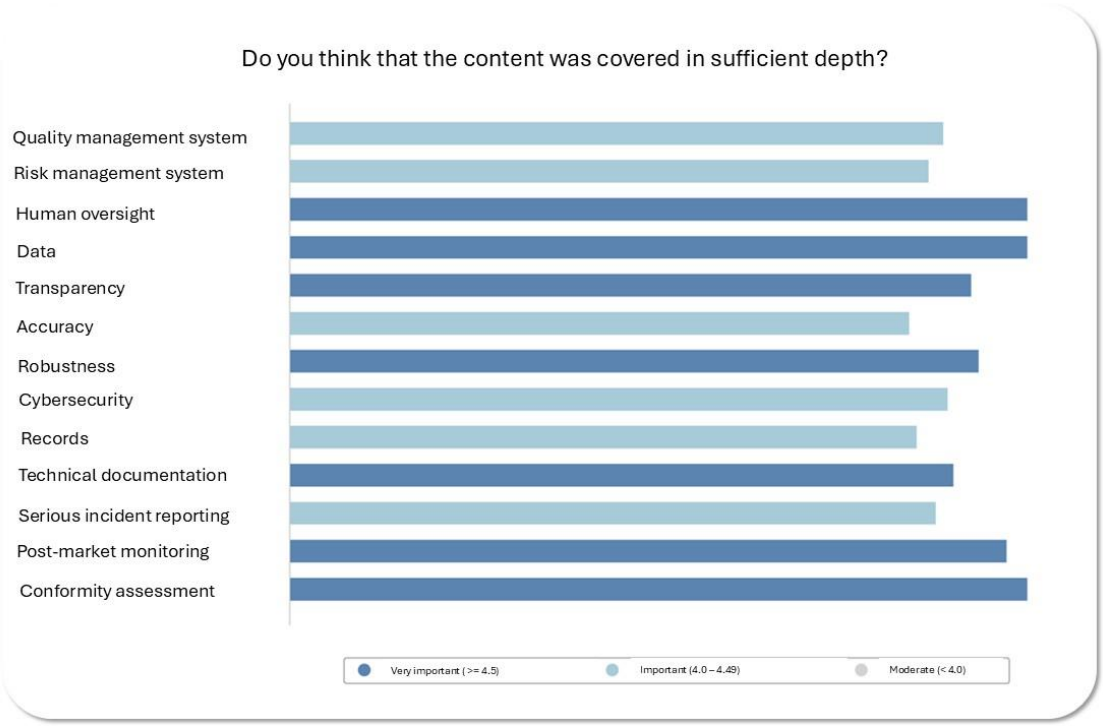

Image 20 - Depth of the content

The participants' perception of the depth with which the content was addressed during the training session is shown in the **Image 20 - Depth of the content** above.

The results reflect a **very positive overall assessment**, as most of the guides are in the "Very Important" category**,** which shows that the participants perceived that the sessions provided an adequate level of detail to understand the main requirements of the AI Act.

Among the content that received the highest ratings are human oversight, data, conformity assessment and post-market monitoring, followed by robustness, technical documentation and transparency. These results suggest that the sessions dedicated to these requirements achieved an adequate balance between the time available and the level of detail offered, allowing participants to acquire a sufficiently in-depth understanding of these subjects.

On the other hand, some requirements received slightly lower ratings, although they remain within the "Important" category. This is the case for the quality management system, risk management system, accuracy, cybersecurity, records and serious incident reporting. None of them falls into the “moderate” rating category, indicating that even in the relatively less highly rated areas, participants consider the depth of the content to be satisfactory.

If compared to the previous graph, one particularly interesting aspect can be observed. While in the analysis of understanding the precision requirement was the only one that appeared below the threshold, in this dimension its assessment improves, and it moves into the category of "Important". This result suggests that the lower understanding of this requirement does not seem to be due to a lack of depth in the training, but probably to the technical complexity of the concept itself and to the interpretative difficulties that existed when the sessions were held, prior to the publication of the harmonized standards.

Likewise, it can be seen that requirements such as serious incident reporting or records, which in the dimension of understanding had obtained very high scores, receive here a somewhat more moderate assessment in terms of depth. This difference can be interpreted as an indication that, although the participants had a good understanding of these requirements, they would have appreciated a more detailed explanation or a higher number of practical examples.

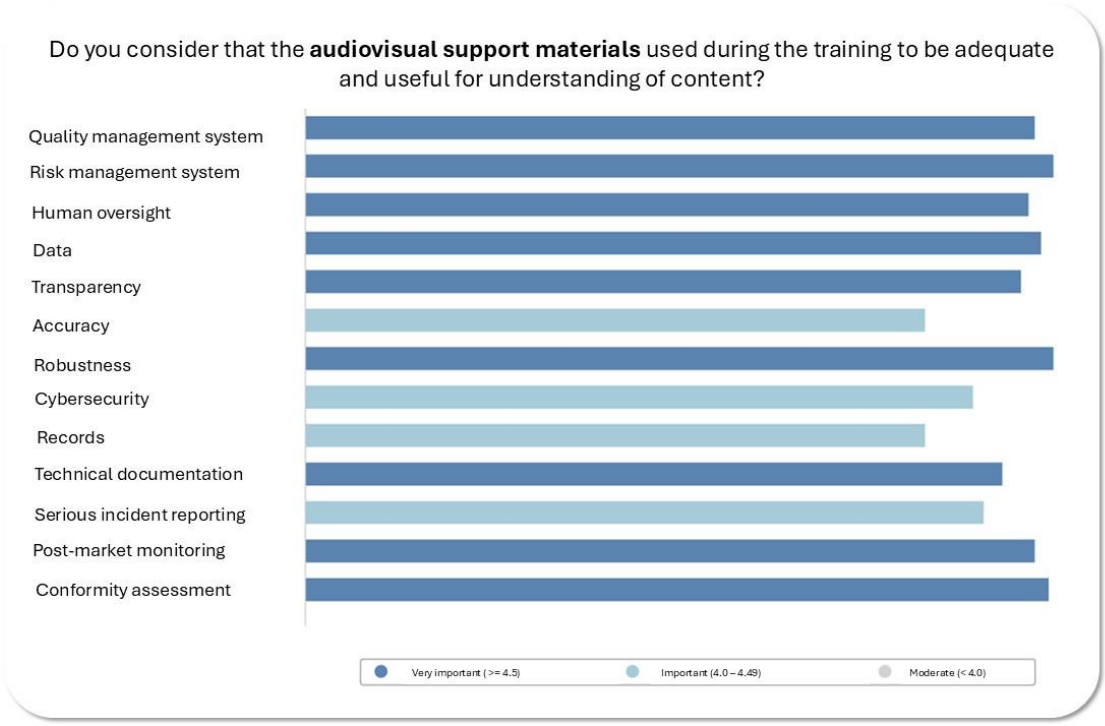

Image 21 - Usefulness of audiovisual materials

The following **Image 21 - Usefulness of audiovisual materials** shows the perception of the participants about the usefulness and adequacy of the audiovisual support material used during the sessions to facilitate the understanding of the content.

The results show a very favourable assessment of the resources used. Most of the requirements were categorized as "Very Important", indicating that the audiovisual materials contributed significantly to facilitating the understanding of the different requirements of the AI Act.

The highest ratings are concentrated on risk management system, robustness, conformity assessment, post-market monitoring, quality management system, data governance, human oversight, transparency and technical documentation. These results show that the materials designed for these subjects were perceived as especially effective in supporting the learning process and complementing the explanations given during the sessions.

On the other hand, **accuracy**, **cybersecurity**, **records** and **serious incident reporting** are placed in the "Important" category. While these assessments are still clearly positive, they suggest that the resources used on these issues could be optimized. Given that these are requirements with a strong technical component, it is reasonable to think that the incorporation of more detailed graphical representations, process diagrams, implementation examples or practical demonstrations would contribute to further improving their explanatory capacity.

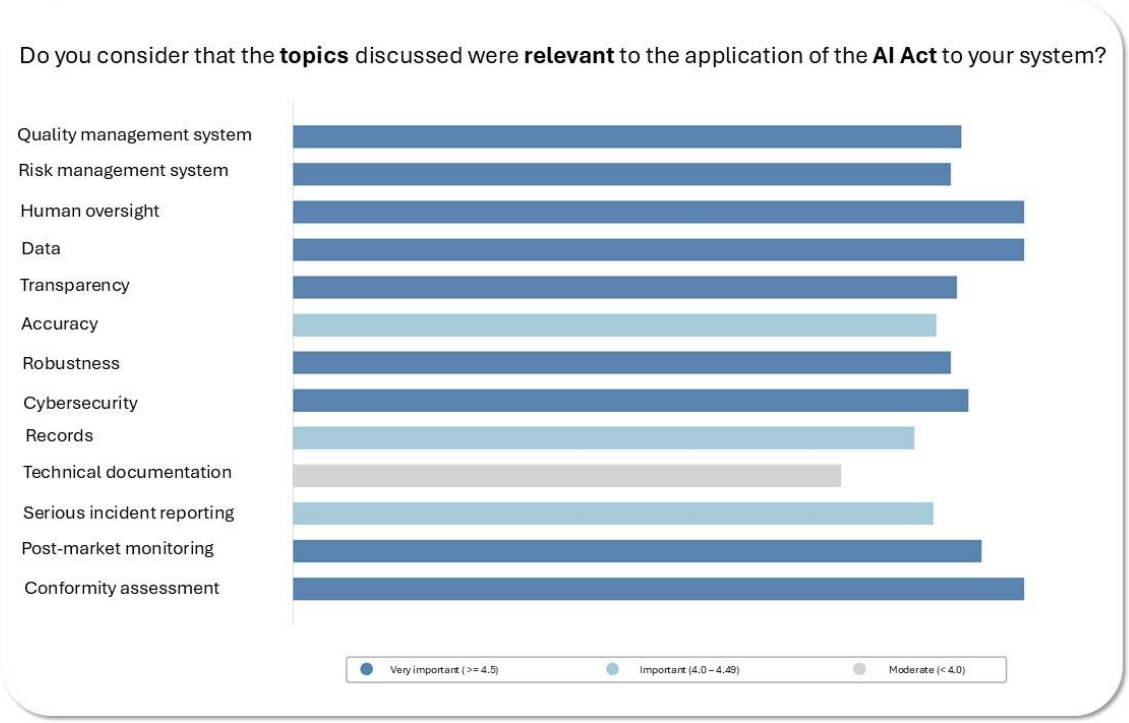

Image 22 - Relevance of the topics taught

It assesses the perceived relevance of the content taught for the application of the AI Act to the systems of the participating entities. This is a particularly significant dimension, as it goes beyond assessing the quality of the training to measure **its practical usefulness and alignment with real implementation needs.**

The results shown in the **Image 22 - Relevance of the topics taught** above show a broadly positive perception regarding the relevance of the contents addressed. Most of the requirements obtain scores corresponding to the "Very Important" category, indicating that participants consider that the sessions dealt with aspects directly applicable to their systems and relevant to addressing the obligations arising from the AI Act.

Among the requirements that receive the highest ratings are human oversight, data governance and conformity assessment, followed by post-market monitoring, cybersecurity, quality management system, risk management system, transparency and robustness. These scores show that participants identify these areas as essential elements for the adaptation of their systems to the new regulatory framework, confirming that the selected content adequately responded to the priorities of the AI Sandbox.

On the other hand**,** accuracy, records and serious incident reporting obtain ratings included in the "Important" category**.** While they still reflect a favourable perception of their usefulness, they suggest that this content could be reinforced through a more practical implementation-oriented approach or through specific examples that allow for a clearer visualization of their application in different types of AI systems.

The case of technical documentation deserves consideration, as it is the only requirement that is classified in the "Moderate" category. This result does not necessarily imply that technical documentation is perceived as an unimportant requirement from a regulatory point of view, but may reflect that, at the time the AI Sandbox was conducted, participants encountered greater difficulties in appreciating its practical application. This interpretation is consistent with the context in which the sessions took place, characterised by the absence of harmonised standards and consolidated technical references specifying the content and scope of the documentation required by the AI Act. Consequently, it is reasonable that this requirement was perceived as more abstract or less immediately transferable to the reality of the systems evaluated.

In the second phase, **Image 23 - Phase II Satisfaction: Self-Diagnosis and Adaptation Plan** shows the results of the following questions:

- Has the self-diagnosis process improved your understanding of the technical guidance measures that your AI system would need to comply with?
- Has the material provided (templates and checklists) been adequate?
- Considering that the self-diagnosis process has provided you with a complete view of how mature your system is, do you consider the workload it has entailed?
- Do you think that, on a global level, this phase has been satisfactory?

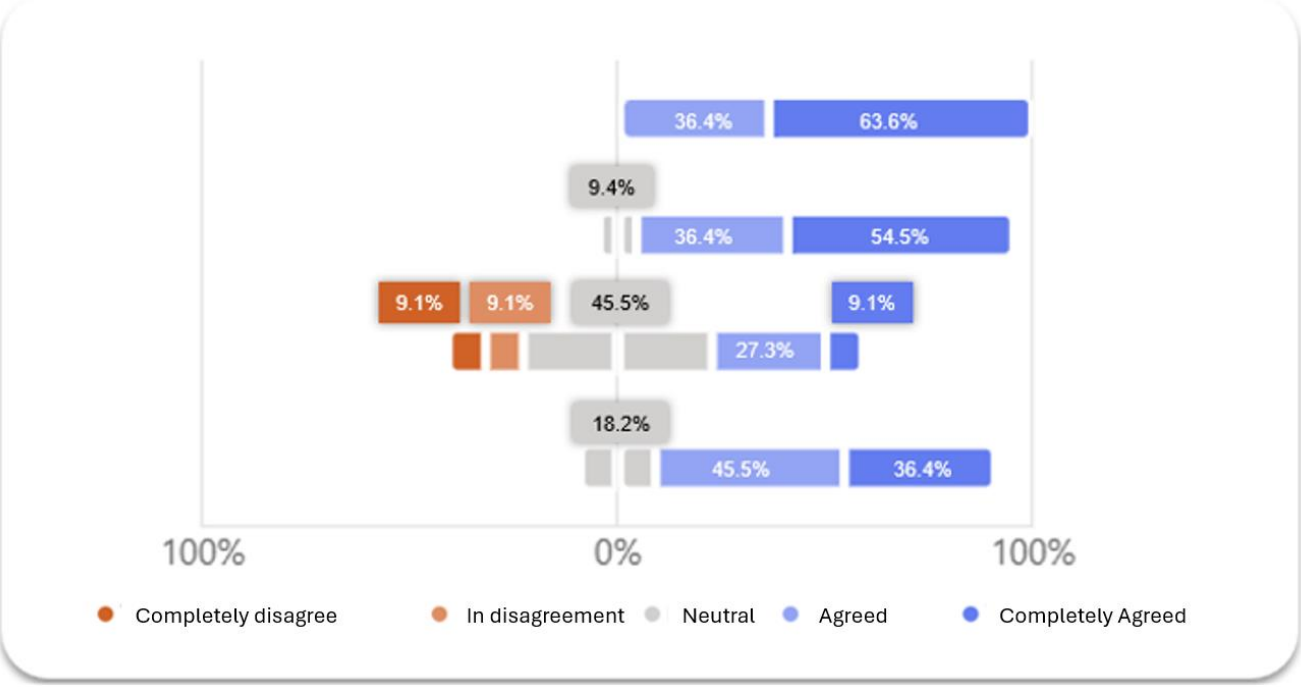

Image 23 - Phase II Satisfaction: Self-Diagnosis and Adaptation Plan

63.6% of the entities consider that the self-diagnosis process has been crucial in improving their understanding of the measures required by the technical guides and how they could be implemented so that each participating system could be adapted to those requirements. Only 36.4% consider that it is a process that should be carried out, but they do not regard it as essential.

In the third phase, the implementation phase of the adaptation plan**, Imagen 24 - Percentage of systems that have sought assistance from internal technical experts** and **Image 25 - Satisfaction Phase III: PoA Implementation** show the results of the following questions:

- Have you required the support of the guides' technical experts during this phase?

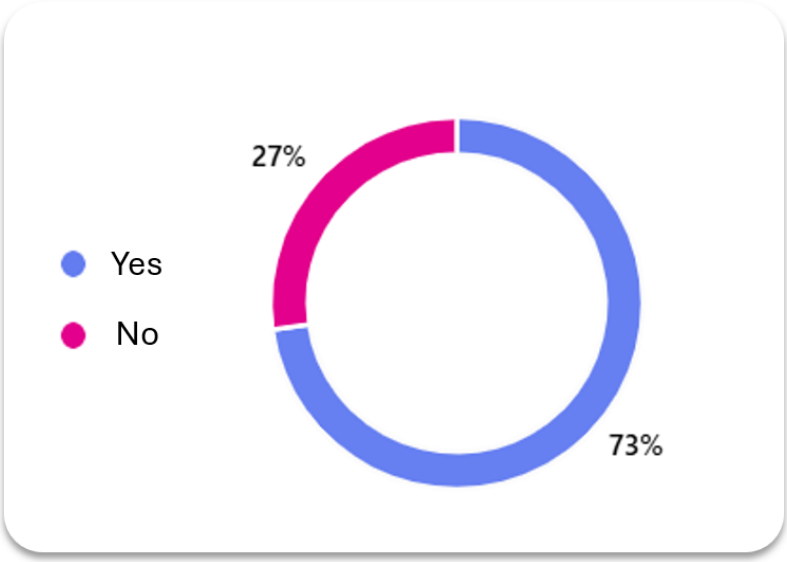

Imagen 24 - Percentage of systems that have sought assistance from internal technical experts

73% of the systems consider mentoring to be of vital importance for the resolution of queries and doubts about the technical-regulatory aspects of the phase. It is regarded as essential support for clarifying different aspects of the Sandbox.

- Has this phase facilitated the implementation of the measures in the guides planned in your adaptation plan?
- Has the actual effort dedicated to this phase been adapted to that planned in the previous phase?
- Do you think that this phase has been satisfactory on a global level?

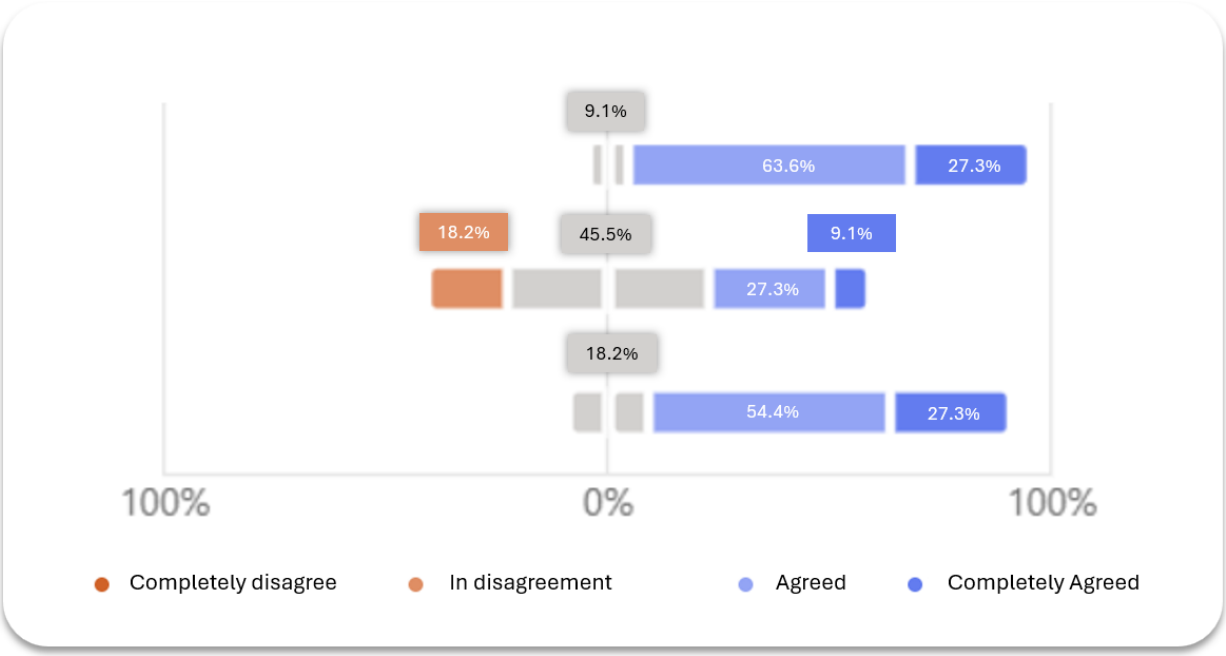

Image 25 - Satisfaction Phase III: PoA Implementation

In the fourth phase, **Image 26 - Phase IV Satisfaction: Self-Assessment** presents the results to the following questions:

- Has the results report received contributed to improving your understanding of the level of compliance achieved?
- Has preparing the self-assessment made it easier for you to identify the evidence foreseeably associated with the AI Act requirements for your system?
- Has the effort required by the self-assessment procedure been adequate?
- Do you think that this phase has been satisfactory on a global level?

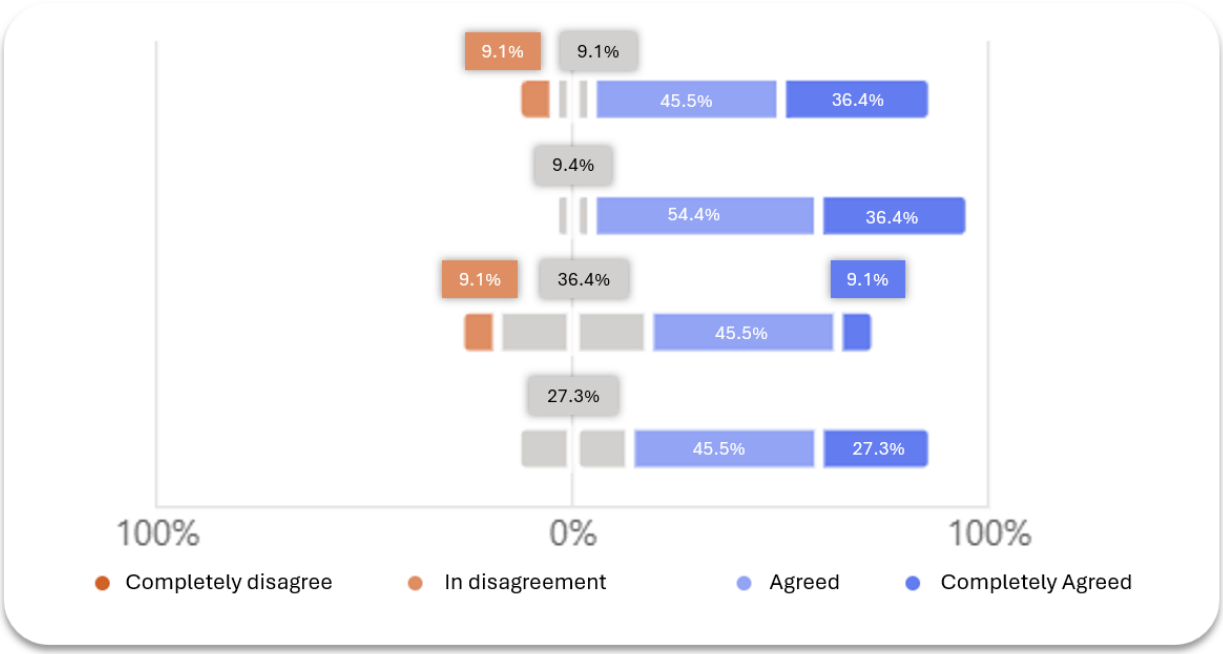

Image 26 - Phase IV Satisfaction: Self-Assessment

In the fifth phase, **Image 27 - Phase V Satisfaction: Post-Market Monitoring** .It includes the results of the following questions:

- Have you found the content taught in the training adequate?
- Were the risks foreseen during the implementation phase in line with those detected in this phase?
- Do you think your system is better prepared for post-market monitoring than it was in at the beginning of the phase?
- Do you think that this phase has been satisfactory on a global level?

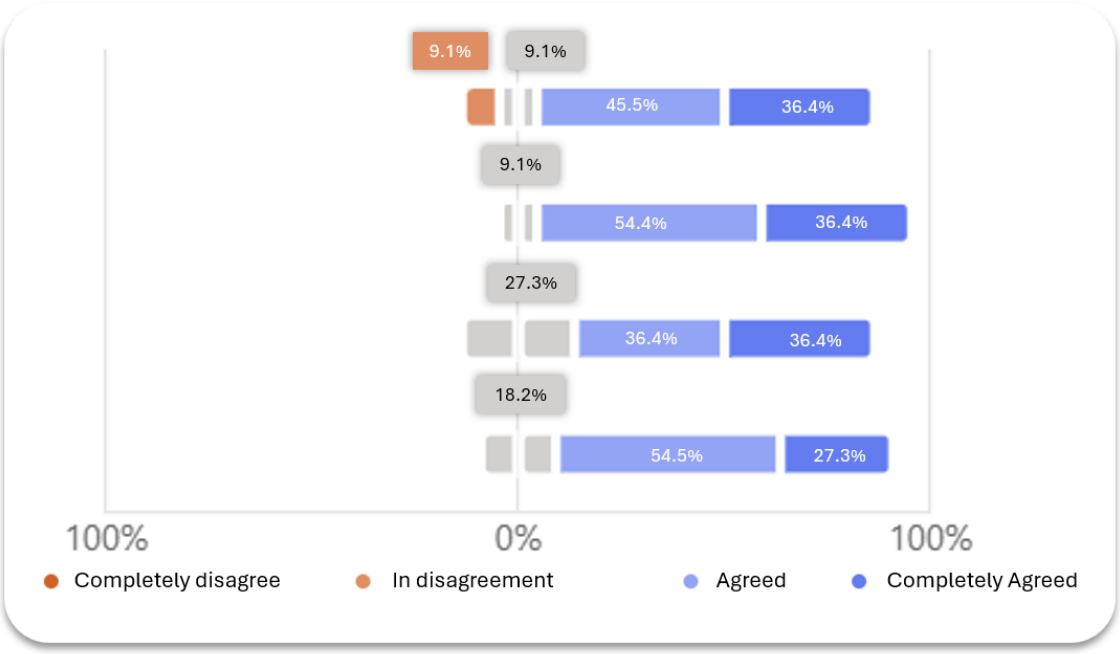

Image 27 - Phase V Satisfaction: Post-Market Monitoring

The Sandbox Global Satisfaction Survey, represented in **Image 28 - Overall satisfaction of the AI Sandbox** , includes the results of the following questions:

- Do you think that participating in the AI Sandbox has improved your organization's readiness to comply with the measures defined in the technical guides?
- Do you think the experience of participating in the first AI Sandbox has been satisfactory?
- Has the project met your initial expectations?
- Do you consider the role of the Administration clear and appropriate?
- Do you think that synergies have been generated between the entities?
- Do you think that the deadlines established for the development of each phase have been adequate and realistic?
- Do you think the shared work environment (SharePoint) was the right fit?
- Do you think that mentoring has facilitated the resolution of doubts that have arisen throughout the project?
- Have you found the events organized within the framework of the sandbox useful?

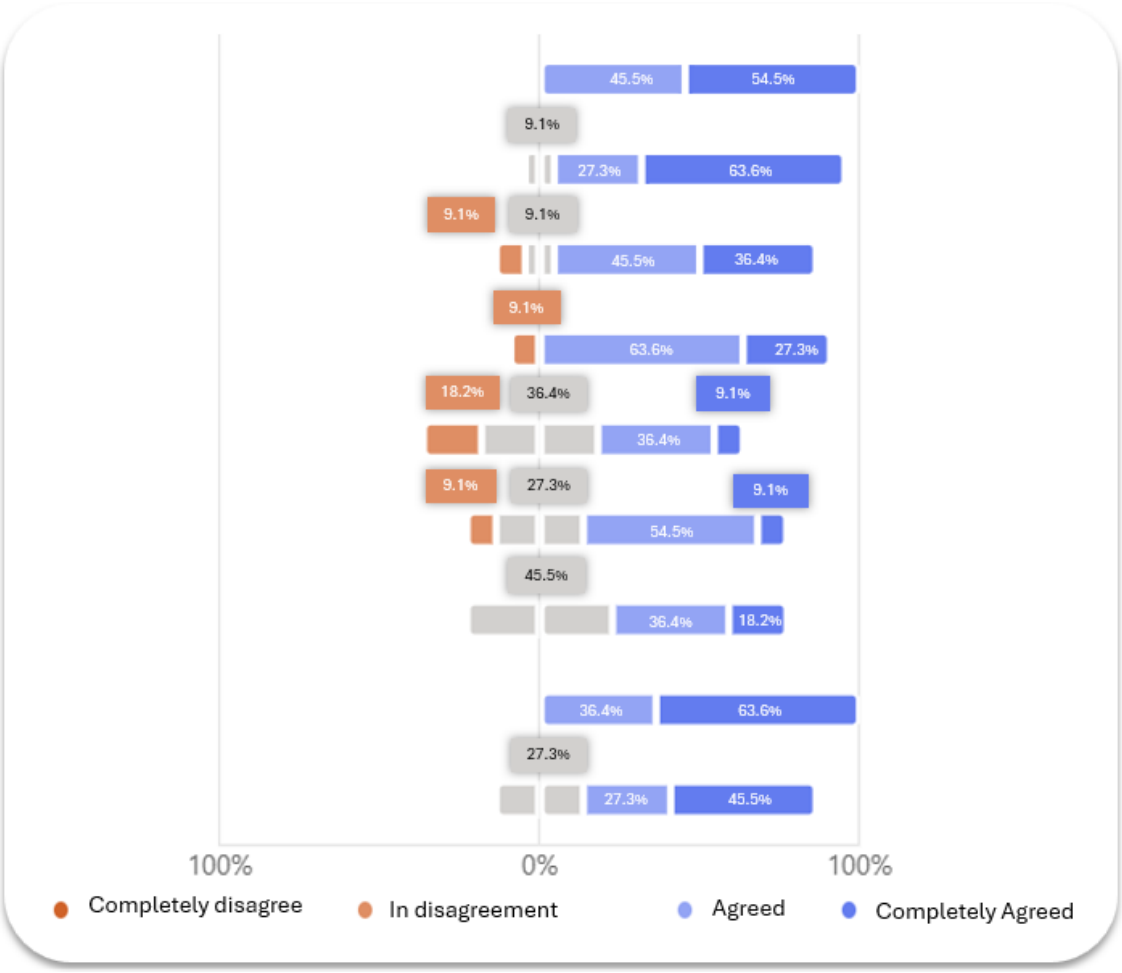

Image 28 - Overall satisfaction of the AI Sandbox

In this last section of the survey, 3 questions were posed with free text:

- Have you identified any significant obstacles?
- What do you think has been the most positive thing about the AI Sandbox pilot?
- What aspects do you think should be improved in future similar initiatives?

From the compendium of replies received, the following overall conclusions have been drawn:

The results show that the AI Sandbox has been a broadly positive experience for the participating organisations, providing a hands-on learning environment that has fostered a deeper understanding of the requirements of the technical guides of the European AI Act and has contributed to increasing the maturity of their internal systems and processes.

Among the most valued aspects are the mentoring and the possibility of having experts available to resolve questions relating to concepts or technical aspects, the close collaboration with the Administration, the quality of the guides prepared and the opportunity to exchange experiences with other participants, creating a highly enriching work model.

However, the experience has also made it possible to identify various challenges associated with the implementation process. These include:

- the high workload.
- the complexity and length of the documentation.
- the lack of greater specificity in some deliverables.
- and the evolution of the regulatory framework during the development of the pilot increased the effort required by organizations, especially those with less capacity.

Likewise, certain use cases with specific technical characteristics highlighted the need for more adapted and flexible methodologies.

In light of these results, the main opportunities for improvement are aimed at evolving the Sandbox towards a more efficient, practical and scalable model. To this end, it is advisable to simplify compliance processes, incorporate tools and templates that facilitate the preparation of documentation, offer more structured feedback on the degree of compliance achieved and strengthen the interaction between members of the mentoring team, the Administration and participants.

Similarly, greater customization of the implementation pathways according to the systems' use cases, the incorporation of sectoral specialists and the development of more operational test environments would contribute to consolidating the Sandbox as a reference instrument for the effective application of the European AI Act and for the promotion of responsible innovation.

Bibliography and Links of Interest

Regulation (EU) 2024/1689 of the European Parliament and of the Council, of 13 June 2024, laying down harmonised rules on artificial intelligence. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024. *Official Journal of the European Union, L, 12 July 2024.* Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Royal Decree 817/2023 of 8 November establishing an AI regulatory sandbox for testing compliance with the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence. *Boletín Oficial del Estado, BOE, No. 268, 9 November 2023.* Available at:. https://www.boe.es/eli/es/rd/2023/11/08/817

Resolution of 20 December 2024 of the State Secretariat for Digitization and Artificial Intelligence announcement for access to the AI regulatory sandbox for trustworthy artificial intelligence. *Electronic Office, Ministry for Digital Transformation and the Civil Service, December 2024.* Available at: https://digital.gob.es/content/dam/portal-mtdfp/avance-digital/sedia/sandbox-ia/enlaces-y-descargas/report_Sandbox%20IA%20Convocatoria%20v5.1%2020241218.pdf

Resolution of the State Secretariat for Digitization and Artificial Intelligence concerning the call for applications for access to the AI regulatory sandbox for trustworthy artificial intelligence, as provided for in Royal Decree 817/2023 of 8 November. *Electronic Office, Ministry for Digital Transformation and the Civil Service, 30 April 2025. Available at:* https://sedediatid.digital.gob.es/_layouts/15/GestorDescargas.ashx?s=AP&k=26886

Resolution of the State Secretariat for Digitization and Artificial Intelligence appointing the members of the Advisory Group of Experts referred to in Royal Decree 817/2023 of 8 November. *Electronic Office, Ministry for Digital Transformation and the Civil Service,10 June 2025. Available at:* https://sedediatid.digital.gob.es/_layouts/15/GestorDescargas.ashx?s=AP&k=26889

Resolution of the State Secretariat for Digitization and Artificial Intelligence accepting the voluntary withdrawal of Loradix Red S.L. from the AI regulatory sandbox, pursuant to Article 24 of Royal Decree 817/2023 of 8 November. *Electronic Office, Ministry for Digital Transformation and the Civil Service*, 12 March 2026. Available at: https://sedediatid.digital.gob.es/_layouts/15/GestorDescargas.ashx?s=AP&k=26896

State Secretariat for Digitization and Artificial Intelligence (Ministry for Digital Transformation and the Civil Service). Technical Guides Developed as Part of the Spanish AI Regulatory Sandbox Pilot. December 2025. Available at: https://aesia.digital.gob.es/en/guides

General Secretariat for Innovation (Ministry of Science, Innovation and Universities). *Sandbox: Understanding and Implementing. A Guide to Regulatory sandboxes*. *Publications Centre of the Ministry of Science, Innovation and Universities, 2024.* Available at: [Online] 2024. https://www.ciencia.gob.es/Innovar/sandboxes.html

Bank of Spain. FERNÁNDEZ ZAMANILLO, Concepción y TOLOBA GÓMEZ, Carolina. Spanish Regulatory Sandbox: Impact on the Promoters of Projects Monitored by the Bank of Spain. *Occasional Papers, No. 2427, September 2024.* Available at: https://www.bde.es/wbe/en/publicaciones/analisis-economico-investigacion/documentos-ocasionales/sandbox-regulatorio-espanol-impacto-en-los-promotores-de-los-proyectos-monitorizados-por-el-banco-de-espana.html

PERSET, Karine y PIETIKÄINEN, Anna . Why AI Sandboxes matter for responsible innovation and public trust. *OECD. AI.*2026.Availble at: https://oecd.ai/en/wonk/sandboxes-matter-responsible-innovation-public-trust

Organisation for Economic Co-operation and Development. Regulatory sandboxes in artificial intelligence. *OECD Digital Economy Papers, n.º 356, July 2023.* Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/07/regulatory-sandboxes-in-artificial-intelligence_a44aae4f/8f80a0e6-en.pdf