6

Empresas desarrollando cumplimiento de requisitos

Guía 6. Supervisión humana

Reglamento Europeo de Inteligencia Artificial

Financiado por la Unión Europea
NextGenerationEU

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

Plan de Recuperación, Transformación y Resiliencia

España | digital 2026

Esta guía ha sido desarrollada en el marco del desarrollo del piloto español de sandbox regulatorio de IA, en colaboración entre los participantes, asistencias técnicas, potenciales autoridades nacionales competentes y el grupo asesor de expertos del sandbox.

La guía tiene como objetivo servir de apoyo introductorio a la normativa europea de Inteligencia Artificial y sus obligaciones aplicables. Si bien **no tiene carácter vinculante ni sustituye ni desarrolla la normativa aplicable, proporciona recomendaciones prácticas** alineadas con los requisitos regulatorios a la espera de que se aprueben las normas armonizadas de aplicación para todos los estados miembros.

El presente documento está sujeto a un **proceso permanente de evaluación y revisión**, con actualizaciones periódicas conforme al desarrollo de los estándares y las distintas directrices publicadas desde la Comisión Europea, y será actualizada una vez se apruebe el Ómnibus digital que modifica el Reglamento de Inteligencia Artificial.

Entre las referencias técnicas relevantes que son aplicables, destaca la norma "***prEN 18229-1 AI Trustworthiness Framework – Part 1: Logging, Transparency and Human Oversight***", que actualmente está en desarrollo.

Fecha de versión: 10 de diciembre de 2025

Contenido general

Índice detallado

1. Preámbulo

1.1 Objetivo de este documento

El Reglamento Europeo de Inteligencia Artificial (*AI Act*) dedica su artículo 14 a la Vigilancia Humana sobre sistemas de IA de alto riesgo. Dicho artículo se enmarca en el capítulo segundo (*Requisitos para los sistemas de alto riesgo*) del mencionado reglamento.

La versión del Reglamento Europeo de la IA tomada como referencia en el presente documento ha sido la publicada por el Consejo de la Comisión Europea el 13 de junio de 2024.

El presente documento proporciona medidas de implementación que faciliten el cumplimiento de los requisitos expresados en el mencionado artículo.

1.2 ¿Cómo leer esta guía?

Como se mencionaba anteriormente, el **presente documento proporciona medidas** de implementación para entidades proveedoras y responsables del despliegue de los sistemas de IA **que faciliten el cumplimiento** de las obligaciones expresadas en el artículo 14 del Reglamento, dedicado a la Vigilancia Humana.

Para ello el documento **recorre en orden** todos los apartados de dicho artículo, dando respuesta a las preguntas fundamentales necesarias para **facilitar el cumplimiento** de las obligaciones expresadas en dichos apartados.

1.3 ¿A quién está dirigido?

Este documento es una **guía de implementación** sobre la Transparencia en Inteligencia Artificial para conseguir los objetivos marcados por el Reglamento Europeo de la IA (*AI Act*).

Por tanto, este documento está dirigido a:

- Los responsables de la entidad proveedora que diseñan conceptualmente el sistema de IA atendiendo a los requisitos del responsable del despliegue, quienes podrán tener en cuenta las medidas descritas en el presente documento para crear un sistema de IA Transparente en función de los requisitos descritos en el Reglamento.
- Los responsables del despliegue, quienes deberán ser conscientes de los requisitos sobre Transparencia que tendrán en función del caso de uso y proceso que va a soportar el sistema de IA.

A lo largo de todo el documento se utiliza un lenguaje entendible por todos ellos, minimizando los tecnicismos necesarios para su comprensión.

1.4 Casos de uso

Para contextualizar cada una de las medidas expuestas que permiten cumplir los requisitos del reglamento, se utilizarán ejemplos sobre dos casos de uso:

- Concesión de ayudas económicas a familias sin recursos
- Gestión de enfermedades crónicas - Bomba de insulina inteligente

Dichos casos de uso se desarrollan en detalle en la **Guía práctica y ejemplos para entender el Reglamento IA.**

Los ejemplos sobre dichos casos de uso son expuestos a alto nivel, sin entrar en detalles ni ser exhaustivos, para intentar abarcar las mayores casuísticas posibles. Además, no responden a experiencias reales (pero sí con la intención de ser realistas desde un punto de vista didáctico), teniendo como objetivo únicamente aclarar un poco más las medidas, no pudiendo por tanto ser tomados como especificaciones en una implantación real.

2. Introducción

2.1 ¿Qué es la Vigilancia Humana en la Inteligencia Artificial?

Las personas deben ser capaces de tomar decisiones autónomas con conocimiento de causa en relación con los sistemas de IA. Para ello se les deberá proporcionar los conocimientos y herramientas necesarios para comprender los sistemas de IA e interactuar con ellos de manera satisfactoria y, siempre que resulte posible, permitirles evaluar por sí mismos o cuestionar el sistema. Los sistemas de IA deberían ayudar a las personas a tomar mejores decisiones y con mayor conocimiento de causa de conformidad con sus objetivos. El principio general de vigilancia humana **debe ocupar un lugar central en la funcionalidad del sistema.**

La vigilancia humana ayuda a garantizar que un sistema de IA no socave la autonomía humana o provoque otros efectos adversos. La vigilancia se puede llevar a cabo a través de **mecanismos de gobernanza**, tales como los enfoques de participación humana, control o mando humanos. La participación humana hace referencia a la capacidad de que **intervengan seres humanos en todos los ciclos de decisión del sistema**, algo que en muchos casos no es posible ni deseable. El control humano se refiere a la capacidad de que intervengan seres humanos **durante el ciclo de vida del sistema y en el seguimiento de su funcionamiento**. Por último, el mando humano es la capacidad de supervisar la actividad global del sistema de IA, así como la capacidad de **decidir cómo y cuándo utilizar el sistema en una situación determinada.** Esto puede incluir la decisión de no utilizar un sistema de IA en una situación particular, establecer niveles de discrecionalidad humana durante el uso del sistema o garantizar la posibilidad de ignorar una decisión adoptada por un sistema. Puede ser necesario introducir mecanismos de supervisión en diferentes grados para respaldar otras medidas de seguridad y control, dependiendo del ámbito de aplicación y el riesgo potencial del sistema de IA. Si el resto de las circunstancias no cambian**, cuanto menor sea el nivel de supervisión que pueda ejercer una persona sobre un sistema de IA, mayores y más exigentes serán las verificaciones y la gobernanza necesarias.**

La responsabilidad final de las acciones realizadas por un sistema de IA es competencia de las personas de la entidad proveedora y usuaria responsables del mismo. Por ello es necesario que dichas personas puedan **vigilarlo** (apartado primero del artículo 14). Para que dicha vigilancia sea efectiva, las personas deben tener el **control sobre el sistema** y poder gestionar los **riesgos** que pueden derivarse de su uso (apartado segundo del artículo 14). Para ello:

- Debe existir **presencia humana** en algún momento del proceso en el que intervienen el sistema de IA.
- El sistema debe estar diseñado y construido de tal manera que permita **interpretar** sus mecanismos de razonamiento y sus resultados.

En primera instancia podemos identificar una importante relación entre el concepto de vigilancia humana y otros dos desarrollados en el capítulo segundo del Reglamento Europeo de la IA (*Requisitos para los sistemas de alto riesgo*):

- **Transparencia**. Ya que para que un sistema pueda ser vigilado es necesario entender su funcionamiento y, para que dicho entendimiento pueda existir, es fundamental que el sistema sea transparente.
- **Gestión de riesgos**, ya que para que la vigilancia tenga garantía efectiva y plena es necesario tener el **control sobre el sistema** y poder gestionar los **riesgos** que pueden derivarse de su uso (apartado segundo del artículo 14).

A modo de resumen introductorio visual, se facilita una infografía que trata de dar una visión general del diseño de los sistemas de IA de alto riesgo que permitan realizar sobre ellos una vigilancia efectiva:

2.2 Por qué la necesidad de Vigilancia humana

La Inteligencia Artificial es una de las tecnologías software más complejas de todas cuantas hemos asimilado. Ello es debido a que sus capacidades (realización de predicciones, toma de decisiones, e incluso emulación de capacidades cognitivas del ser humano para realizar dichas predicciones y toma de decisiones), son similares a las del ser humano. **Los mecanismos necesarios para automatizar de manera masiva estas capacidades** mediante sistemas de IA conllevan una **complejidad que requiere de ser vigilada** para generar confianza debido a la criticidad de las acciones que pueda realizar.

Por otra parte, por razones de eficacia, las personas pueden decidir utilizar sistemas de IA. Sin embargo, **esta cesión de control** no puede ser total, manteniendo siempre una componente humana para, además. facilitar la **gestión de responsabilidad y de rendición de cuentas** debidas a las acciones del sistema de IA.

En el presente documento, para cada apartado del artículo 14 del Reglamento Europeo de la IA (dedicado a la Vigilancia humana), se exponen medidas que permiten proporcionar vigilancia humana a un sistema de IA conforme al requisito expuesto en dicho apartado. Cada una de dichas medidas se ejemplifica de manera detallada mediante el caso de uso descrito al inicio del documento.

Para introducir al concepto de Vigilancia Humana y a la importancia de su necesidad, basta con hacernos una pregunta mediante uno de los casos de uso que servirá de hilo conductor a lo largo de la siguiente guía:

¿Qué ocurriría si, como consecuencia de un error en la recolección de los parámetros en sangre debido a un fallo físico del dispositivo, el sistema de IA propusiera al paciente una dosis de insulina no adecuada a su dolencia, un médico no supervisara la prescripción realizada por el sistema, y el paciente procediera a su inoculación?

La respuesta es sencilla: se habría delegado por completo el control médico en un sistema de IA sujeto a errores, y las consecuencias de esta acción tendría un impacto negativo sobre un derecho fundamental de las personas: la salud.

Y aunque en los casos de sistemas de biometría, un fallo en la vigilancia humana tiene consecuencias muy evidentes y llamativas, este requisito no es dictado por el Reglamento de IA solo para ellos, sino que lo es para todos los sistemas IA de alto riesgo, aunque en los sistemas de biometría con unas condiciones particulares.

A lo largo del presente documento veremos cómo evitar este tipo de situaciones.

3. Reglamento de Inteligencia Artificial

La puesta en servicio o la utilización de sistemas de IA de alto riesgo debe supeditarse al cumplimiento de determinados requisitos obligatorios, entre los cuales está el de vigilancia humana. Estos requisitos tienen como objetivo garantizar que los sistemas de IA de alto riesgo disponibles en la Unión o cuyos resultados de salida se utilicen en la Unión no representen riesgos inaceptables para intereses públicos importantes reconocidos y protegidos por el Derecho de la Unión.

En este apartado se incluye los artículos referentes a la generación de registros del Reglamento 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 (Reglamento Europeo de Inteligencia Artificial) y se detalla en que secciones de esta guía se abordan los diferentes elementos de dichos artículos.

3.1 Análisis previo y relación de los artículos

Las obligaciones sobre la generación de registros se encuentran principalmente en el artículo 14 "Supervisión humana"

- **Artículo supervisión humana** → Establece que los HRAIS deberán incorporar las capacidades técnicas necesarias para habilitar la supervisión humana. Este artículo se divide en cinco apartados:
    1. Vigilancia efectiva durante el período que estén en uso.
    2. Establecer mecanismos para minimizar los riesgos para la salud.
    3. Fijar responsabilidades.
    4. Establecer mecanismos de transparencia, explicabilidad y trazabilidad.
    5. Asegurar vigilancia humana en sistemas con identificación biométrica.

3.2 Contenido de los artículos en el Reglamento de IA

AI Act

Art.14 – Supervisión humana

1. Los sistemas de IA de alto riesgo se **diseñarán y desarrollarán** de modo que puedan ser **vigilados de manera efectiva por personas físicas** durante el período que **estén en uso**, lo que incluye dotarlos de **herramientas de interfaz humano-máquina** adecuadas.
2. El **objetivo** de la supervisión humana será **prevenir o reducir al mínimo los riesgos** para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo **conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible**, en

particular cuando dichos riesgos persistan a pesar de la aplicación de otros requisitos establecidos en la presente sección.

3. Las medidas de supervisión serán proporcionales a los riesgos, al nivel de autonomía y al contexto de uso del sistema de IA de alto riesgo, y se garantizarán bien mediante uno de los siguientes **tipos de medidas**, bien mediante ambos:
   a) las medidas que el **proveedor defina y que integre**, cuando sea **técnicamente viable**, en el sistema de IA de alto riesgo antes de su introducción en el mercado o su puesta en servicio;
   b) las medidas que **el proveedor defina antes** de la introducción del sistema de IA de alto riesgo en el mercado o de su puesta en servicio y que sean **adecuadas para que las ponga en práctica el responsable del despliegue**.
4. A efectos de la puesta en práctica de lo dispuesto en los apartados 1, 2 y 3, el sistema de IA de alto riesgo se ofrecerá al responsable del despliegue de tal modo que las personas físicas a quienes se encomiende la supervisión humana puedan, según proceda y de manera proporcionada a:
   a) **entender** adecuadamente las capacidades y limitaciones pertinentes del sistema de IA de alto riesgo y poder vigilar debidamente su funcionamiento, por ejemplo, con vistas a detectar y resolver anomalías, problemas de funcionamiento y comportamientos inesperados;
   b) **ser conscientes** de la posible tendencia a **confiar automáticamente** o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo («sesgo de automatización»), en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión;
   c) **interpretar correctamente** los resultados de salida del sistema de IA de alto riesgo, teniendo en cuenta, por ejemplo, los métodos y herramientas de interpretación disponibles;
   d) **decidir**, en cualquier situación concreta, **no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir** los resultados de salida que este genere;
   e) intervenir en el funcionamiento del sistema de IA de alto riesgo o **interrumpir el sistema** pulsando un botón de parada o mediante un procedimiento similar que permita que el sistema se detenga de forma segura.
5. En el caso de los sistemas de IA de alto riesgo mencionados en el **anexo III, punto 1, letra a)**, **las medidas a que se refiere el apartado 3** del presente artículo garantizarán, además, que el responsable del despliegue no actúe **ni tome ninguna decisión** basándose en la identificación generada por el sistema, **salvo si al menos dos personas** físicas con la competencia, formación y autoridad necesarias han verificado y confirmado por separado dicha identificación.

El requisito de la verificación por parte de al menos dos personas físicas por separado no se aplicará a los sistemas de IA de alto riesgo utilizados con **fines de garantía del cumplimiento del Derecho, de migración, de control fronterizo o de asilo** cuando el Derecho nacional o de la Unión considere que **la aplicación de este requisito es desproporcionada.**

3.3 Correspondencia del articulado con los apartados de la guía

En la siguiente tabla se listan las secciones del documento en el que se encuentran la explicación y las medidas aplicables a cada apartado/subapartado del artículo dedicado a la vigilancia humana.

Artículo Reglamento	Requerimiento Reglamento	Sección guía
14.1	Diseño y desarrollo para una vigilancia efectiva	Apartado 4.1.1
14.2	Riesgos sobre derechos fundamentales	Apartado 4.1.2
14.3.a	Tipos de medidas defina y que integre, cuando sea técnicamente viable	Apartado 4.1.3
14.3.b	Tipos de medidas antes de la introducción del sistema de IA	Apartado 4.1.3
14.4	Entendimiento y Autonomía	Apartado 4.1.4
14.4.a	Entender capacidades y limitaciones	Apartado 4.1.4.1
14.4.b	Sesgo de automatización	Apartado 4.1.4.2
14.4.c	Interpretar la información de salida	Apartado 4.1.4.3
14.4.d	Autonomía para decidir	Apartado 4.1.4.4
14.4.e	Interrupción	Apartado 4.1.4.5
14.5	Identificación biométrica	Apartado 4.1.5

4. ¿Cómo abordar los requisitos?

Recordamos que la presente guía toma como referencia el Reglamento 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 (Reglamento Europeo de Inteligencia Artificial).

El artículo sobre Vigilancia humana tiene **cinco apartados.** Este es un **resumen** de los **requisitos** expresados sobre los sistemas de IA de alto riesgo:

1. Que sean **diseñados y desarrollados para que, cuando estén en uso**, se pueda realizar sobre ellos una vigilancia efectiva.
2. Que existan mecanismos que permitan prevenir o **reducir al mínimo los riesgos** para la salud, la seguridad o los derechos fundamentales.
3. **Fijar la responsabilidad** de proveedor y responsable del despliegue.
4. Que las **medidas utilizadas para conseguir los requisitos** de los anteriores apartados **permitan a las personas** responsables de la vigilancia del sistema de IA:
   - Entender el sistema (apartados 4a y 4c).
   - Ser conscientes de su autonomía sobre el sistema (apartados 4b,4d y 4e).
5. Asegurar la vigilancia humana en un tipo de sistema de IA particular: los dedicados a la **identificación biométrica remota.**

A continuación, se desarrollan cada uno de los apartados y subapartados del artículo 14 dedicados a la vigilancia humana, indicando para cada uno de ellos las medidas necesarias para abordar los requisitos planteados por el Reglamento.

4.1 Requisitos sobre Vigilancia Humana en el Reglamento

4.1.1 Apartado 1. Diseño y desarrollo para una vigilancia efectiva

AI Act

Art.14.1- Supervisión humana

Los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas.

Qué entendemos

El artículo comienza con un objetivo global: que los sistemas de IA de alto riesgo sean diseñados y desarrollados para que, cuando posteriormente estén en uso, se pueda realizar sobre ellos una vigilancia efectiva por parte de las personas responsables del sistema.

Medidas para llevarlo a cabo

Por tanto, primero para cumplir este requisito es necesario detallar las medidas de diseño y desarrollo para una vigilancia efectiva del sistema de IA.

Una vez diseñado y desarrollado el sistema bajo dichas medidas, el sistema pasa a estar en uso. Y entonces es necesario habilitar mecanismos que permitan monitorizar el resultado de dichas medidas. Para ello el apartado anticipa la necesidad de habilitar una "interfaz humano-máquina".

4.1.2 Apartado 2. Riesgos sobre derechos fundamentales

AI Act

Art.14.2 - Supervisión humana

El objetivo de la supervisión humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persistan a pesar de la aplicación de otros requisitos establecidos en la presente sección.

Qué entendemos

En este apartado pone énfasis en cómo la vigilancia humana debe tener en cuenta la **gestión de riesgos** que puedan **afectar a la salud, la seguridad y los derechos fundamentales** de las personas ya sea por la finalidad prevista del sistema de IA o por otros usos indebidos razonablemente previsibles que se hagan del mismo.

De hecho, el artículo 9, dedicado al Sistema de gestión de riesgos, comienza sus requisitos remarcando igualmente en su apartado segundo estos tres ámbitos: salud, seguridad y derechos fundamentales.

En cuanto a los **derechos fundamentales,** este concepto se cita en el Reglamento Europeo de la IA más de cincuenta ocasiones ya que los sistemas de IA de **alto riesgo**, objetivo del reglamento, deben asegurar la salvaguarda de dichos derechos. Estos derechos son los recogidos en la Carta de los Derechos Fundamentales de la Unión Europea.

En cuanto a la mención sobre el dominio de la **salud**, es una especificación concreta de la mencionada carta (Artículo 35). La gestión de riesgos en este dominio es especialmente importante ya que, por ejemplo, el riesgo en un sistema de IA encargado de administrar un medicamento a un paciente es alto ya que dicha dosis tiene un impacto directo sobre la salud y la vida de dicho paciente.

En cuanto a la mención sobre el dominio de la **seguridad**, es igualmente una especificación concreta de la mencionada carta (transversal, con mención concreta en su artículo sexto). La gestión de riesgos en este dominio es igualmente importante ya que, por ejemplo, el riesgo

en un sistema de IA encargado de seguridad en un vehículo autónomo puede provocar accidentes que afecten a la vida de las personas.

Medidas para llevarlo a cabo

La vigilancia humana debe ser capaz de gestionar los riesgos en estos tres dominios. Para ello se deberá realizar una gestión de riesgos del sistema de IA conforme a las medidas desarrolladas en la **guía del Artículo 9** (Sistema de gestión de riesgos), aplicando dichas **medidas tanto a la finalidad prevista del sistema, como a los usos indebidos razonablemente previsibles** que se puedan hacer del mismo.

Las medidas definidas en dicha guía incluyen un **ejemplo ilustrativo del desarrollo de un sistema de gestión de riesgos para varios casos de uso**, soportado por una *checklist* documentada de sencillo interfaz.

Esta gestión de riesgos tiene que estar enmarcada dentro del modelo de gobernanza, una medida específica para asegurar la Vigilancia Humana, desarrollado en el apartado **Modelo de gobernanza** del presente documento.

4.1.3 Apartado 3. Tipos de medidas

AI Act

Art.14.3 - Supervisión humana

Las medidas de supervisión serán proporcionales a los riesgos, al nivel de autonomía y al contexto de uso del sistema de IA de alto riesgo, y se garantizarán bien mediante uno de los siguientes tipos de medidas, bien mediante ambos:

a) las medidas que el proveedor defina y que integre, cuando sea técnicamente viable, en el sistema de IA de alto riesgo antes de su introducción en el mercado o su puesta en servicio;

b) las medidas que el proveedor defina antes de la introducción del sistema de IA de alto riesgo en el mercado o de su puesta en servicio y que sean adecuadas para que las ponga en práctica el responsable del despliegue.

Qué entendemos

Fijar la responsabilidad de proveedor y responsable del despliegue:

- El proveedor debe proporcionar al responsable del despliegue las medidas que faciliten la vigilancia humana del sistema.
- El responsable del despliegue debe ponerlas en práctica.

Medidas para llevarlo a cabo

Todas las indicadas en el apartado de medidas aplicables del presente documento.

4.1.4 Apartado 4. Entendimiento y Autonomía

AI Act

Art.14.4 - Supervisión humana

A efectos de la puesta en práctica de lo dispuesto en los apartados 1, 2 y 3, el sistema de IA de alto riesgo se ofrecerá al responsable del despliegue de tal modo que las personas físicas a quienes se encomiende la supervisión humana puedan, según proceda y de manera proporcionada a:

a) entender adecuadamente las capacidades y limitaciones pertinentes del sistema de IA de alto riesgo y poder vigilar debidamente su funcionamiento, por ejemplo, con vistas a detectar y resolver anomalías, problemas de funcionamiento y comportamientos inesperados;

b) ser conscientes de la posible tendencia a confiar automáticamente o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo («sesgo de automatización»), en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión;

c) interpretar correctamente los resultados de salida del sistema de IA de alto riesgo, teniendo en cuenta, por ejemplo, los métodos y herramientas de interpretación disponibles;

d) decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir los resultados de salida que este genere;

e) intervenir en el funcionamiento del sistema de IA de alto riesgo o interrumpir el sistema pulsando un botón de parada o mediante un procedimiento similar que permita que el sistema se detenga de forma segura.

Qué entendemos

En resumen, que las medidas utilizadas para conseguir los requisitos de los anteriores apartados permitan a las personas responsables de la vigilancia del sistema de IA **entender** el sistema (apartados 4a y 4c), y ser conscientes de su **autonomía** sobre el sistema (apartados 4b,4d y 4e).

A continuación, se desarrollan cada uno de los subapartados.

4.1.4.1 Apartado 4a. Entender capacidades y limitaciones

AI Act

Art.14.4a - Supervisión humana

a) entender adecuadamente las capacidades y limitaciones pertinentes del sistema de IA de alto riesgo y poder vigilar debidamente su funcionamiento,

por ejemplo, con vistas a detectar y resolver anomalías, problemas de funcionamiento y comportamientos inesperados;

Qué entendemos

Para que exista vigilancia humana es fundamental que las personas responsables del sistema de IA puedan entender con detalle las capacidades y limitaciones del mismo.

Este requisito está **igualmente expresado en el apartado 13.3.b** del Articulo 13 correspondiente a la Transparencia. Tal y como se refería anteriormente, esta relación tiene un sentido: para que un sistema pueda ser vigilado es necesario entender su funcionamiento y, para que dicho entendimiento pueda existir, es fundamental que el sistema sea transparente.

Medidas para llevarlo a cabo

Las medidas reflejadas en la guía de Transparencia sobre el aparado **13.3.b**

4.1.4.2 Apartado 4b. Sesgo de automatización

AI Act

Art.14.4b - Supervisión humana

b) ser conscientes de la posible tendencia a confiar automáticamente o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo («sesgo de automatización»), en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión;

Qué entendemos

Las personas pueden llegar a confiar en exceso en la salida de un sistema de IA, incluso de mayor manera que en otra persona. Este requisito pretende reforzar el concepto de vigilancia humana, haciendo **ser conscientes de que la responsabilidad final,** ya sea sobre una predicción o sobre una decisión, es propia de las personas que interaccionan con el sistema de alto riesgo, quienes deberán estar convenientemente formadas en el proceso de negocio soportado por el sistema, ya que ellas son quienes deben evaluar y tomar la decisión final.

Medidas para llevarlo a cabo

- Incluir en el sistema un modo de error forzado para testar el criterio y posible sobre confianza de las personas que utilizarán el sistema
- La formación indicada en el modelo de gobernanza.

4.1.4.3 Apartado 4c. Interpretar la información de salida

AI Act

Art.14.4c - Supervisión humana

c) interpretar correctamente los resultados de salida del sistema de IA de alto riesgo, teniendo en cuenta, por ejemplo, los métodos y herramientas de interpretación disponibles

Qué entendemos

Asegurar la existencia de métodos y herramientas que permitan la correcta interpretación de la salida del sistema.

Medidas para llevarlo a cabo

Este es un ejemplo más de la estrecha relación entre la vigilancia humana del sistema de alto riesgo y su transparencia, ya que para poder vigilar un sistema es necesario entender su funcionamiento y, para que dicho entendimiento pueda existir, es fundamental que el sistema sea transparente. Por ello, para cumplir este requisito hay que recurrir a **medidas ya disponibles y reflejadas en la guía de Transparencia**, especialmente las siguientes, ya que están directamente relacionadas con la interpretación de la información de salida del sistema:

- Detallar de lo más global a lo más particular.
- Adaptar el lenguaje.
- Utilizar contrafactualidad.

4.1.4.4 Apartado 4d. Autonomía para decidir

AI Act

Art.14.4d - Supervisión humana

d) decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir los resultados de salida que este genere;

Qué entendemos

Reforzar el concepto de **autonomía como parte de la vigilancia humana** ya que, en general, los sistemas de IA deben diseñarse de manera que aumenten, complementen y potencien las capacidades de las personas, dando cabida a que éstas puedan decidir cuándo y cómo utilizar el sistema en cada situación determinada. Esto puede incluir la decisión de no utilizar un sistema de IA en una situación concreta, establecer niveles de decisión humana durante el uso del sistema o garantizar la capacidad de imponerse a una decisión tomada por el sistema.

Tanto este requisito como el del siguiente apartado, relacionado con la interrupción del sistema, son los que determinan de manera absoluta **el nivel de vigilancia humana.**

Medidas para llevarlo a cabo

- Gobernanza Human in/on the loop

4.1.4.5 Apartado 4e. Interrupción

AI Act

> ### Art.14.4e - Supervisión humana
>
> e) intervenir en el funcionamiento del sistema de IA de alto riesgo o interrumpir el sistema pulsando un botón de parada o mediante un procedimiento similar que permita que el sistema se detenga de forma segura

Qué entendemos

Al igual que los subapartados b) y d), **reforzar el concepto de autonomía** como parte de la vigilancia humana. Este caso requisito planteado es una especificación adicional a la planteada en el anterior subapartado d (Autonomía para decidir), llegando al extremo fijar el requisito de disponer de un mecanismo de desactivación del sistema y/o de un procedimiento asociado.

Medidas para llevarlo a cabo

- Gobernanza Human in/on the loop

4.1.5 Apartado 5. Identificación biométrica

AI Act

> ### Art.14.5 - Supervisión humana
>
> En el caso de los sistemas de IA de alto riesgo mencionados en el **anexo III, punto 1, letra a)**, **las medidas a que se refiere el apartado 3** del presente artículo garantizarán, además, que el responsable del despliegue no actúe **ni tome ninguna decisión** basándose en la identificación generada por el sistema, **salvo si al menos dos personas** físicas con la competencia, formación y autoridad necesarias han verificado y confirmado por separado dicha identificación.
>
> El requisito de la verificación por parte de al menos dos personas físicas por separado no se aplicará a los sistemas de IA de alto riesgo utilizados con **fines de garantía del cumplimiento del Derecho, de migración, de control fronterizo o de asilo** cuando el Derecho nacional o de la Unión considere que **la aplicación de este requisito es desproporcionada.**

Qué entendemos

Los sistemas mencionados en el *Anexo III.1.a* son los **sistemas de identificación biométrica remota.** Para dichos sistemas en particular, se ha de garantizar que las medidas especificadas para el apartado 3 del presente artículo aseguren la verificación separada de dos personas

por separado, a excepción de cuando la legislación de la Unión Europea o la nacional española considere que esa medida es desproporcionada al utilizar dichos sistemas para la garantía de cumplimiento del Derecho, migración, control fronterizo o asilo.
Es importante **diferenciar entre los conceptos de identificación y reconocimiento biométricos**, ya que en muchas ocasiones son utilizados de manera equivalente pero que tienen matices relevantes.

- **Identificación biométrica** es el proceso de verificar la identidad de una persona comparando su información biométrica con una base de datos completa de datos biométricos. Es decir, se trata de encontrar una coincidencia en una población completa, lo que hace que este proceso sea más lento y requiera más recursos. La identificación biométrica se utiliza comúnmente en aplicaciones gubernamentales y de seguridad, como el control de fronteras y la vigilancia pública.
- **Reconocimiento biométrico** es el proceso de verificar la identidad de una persona comparando su información biométrica con un conjunto limitado de datos biométricos, que previamente se han seleccionado o se encuentran en un grupo de interés. Es decir, se trata de encontrar una coincidencia dentro de un grupo específico de personas. El reconocimiento biométrico se utiliza comúnmente en aplicaciones comerciales, como el control de acceso a edificios y el desbloqueo de dispositivos móviles.

Una vez diferenciados ambos conceptos, se remarca que el Reglamento Europeo de la IA hace referencia a los sistemas de IA de **identificación biométrica.**

Medidas para llevarlo a cabo

La medida de **modelo de gobernanza** sobre el sistema de IA para una vigilancia efectiva indicada en el apartado 3 deberá contemplar la casuística indicada para estos sistemas de identificación biométrica remota.

4.2 Medidas aplicables para conseguir la Vigilancia Humana

Este capítulo del documento recoge el detalle de las medidas necesarias para cubrir los requisitos de Vigilancia humana expuestos en el artículo 14 del Reglamento.

4.2.1 Medidas de diseño y desarrollo para una vigilancia efectiva

Para poder vigilar cualquier sistema *software* es necesario **conocer cómo está diseñado y construido**. El siguiente paso necesario es fijar las **dimensiones** de dicho diseño y construcción sobre las cuales realizar la vigilancia humana. En el caso de los sistemas de IA de alto riesgo, el Reglamento Europeo de la IA define dichas dimensiones en su **capítulo segundo** (*Requisitos para los sistemas de alto riesgo*). Dicho capítulo se divide en los siguientes artículos:

- Artículo 9. Gestión de riesgos.
- Artículo 10. Datos y su gobernanza
- Artículo 11. Documentación técnica
- Artículo 12. Registros

- Artículo 13. Transparencia
- Artículo 15. Precisión, solidez y ciberseguridad

Dadas las menciones que se realizan en varios apartados del artículo 14 al entendimiento y a los riesgos, de entre todas las dimensiones podemos destacar:

- Transparencia. Ya que para que un sistema pueda ser vigilado es necesario entender su funcionamiento y, para que dicho entendimiento pueda existir, es fundamental que el sistema sea transparente.
- **Gestión de riesgos**, ya que para que la vigilancia tenga garantía efectiva y plena es necesario tener el **control sobre el sistema** y poder gestionar los **riesgos** que pueden derivarse de su uso (apartado segundo del artículo 14).

Todas las medidas necesarias se encuentran detalladas en las guías de implementación de cada uno de los mencionados artículos.

A quién aplica

En las medidas reflejadas en cada una de las mencionadas guías se refleja a quien aplica, ya sea al proveedor o al responsable del despliegue del sistema de IA. Como resumen global:

- En las medidas de diseño y desarrollo del sistema utilizado por el responsable del despliegue:
  - El responsable del despliegue tiene la responsabilidad de identificar los requisitos del Reglamento Europeo de la IA que deba cumplir el proceso de negocio que va a soportar el sistema. Además, debe garantizar mediante pruebas de aceptación que dichos requisitos son finalmente soportados por el sistema.
  - El proveedor debe implementar las medidas técnicas necesarias alineadas con dichos requisitos.
- Cuando el sistema está en uso:
  - El responsable del despliegue tiene la responsabilidad de asegurar que el comportamiento del sistema en producción se ajusta a los requisitos del reglamento.
  - La labor del proveedor no termina con la implantación, sino que también continua a partir de que el sistema está en producción. Esta relación, habitual en cualquier sistema *software*, es especialmente importante en los sistemas de IA, dada su complejidad, y particularmente en los de alto riesgo dada la criticidad de los procesos que soportan.

Ejemplo

Utilizamos como ejemplo una medida utilizada para conseguir Transparencia y que es detallada en la guía de dicho artículo: *Adaptar el lenguaje*.

A modo de recordatorio en modo resumen, dicha medida expone que:

- El sistema de IA debe ser diseñado para que pueda proporcionar información a todos los perfiles que interaccionan con él y así asegurar de manera transparente su completo entendimiento.

- Son muchos los tipos de perfiles que interaccionan con el sistema de IA a lo largo de todo su ciclo de vida. Por tanto, se han de habilitar mecanismos técnicos que permitan mostrar dicha información de manera transparente y entendible por todos ellos, adecuando el tipo de lenguaje a su nivel de interlocución.

A continuación, lo particularizamos para los dos casos de uso, utilizando los reflejados en la guía de Transparencia. En ambos casos el sistema proporcionará una **interfaz de usuario**, alineada con el nivel de interlocución de cada uno de los perfiles que interactúa con el sistema, y que facilite la siguiente información de manera fácilmente entendible, visual y textualmente en lenguaje natural, sobre todo a aquellos perfiles no técnicos**, para facilitar así su Vigilancia Humana.**

Ejemplo - Concesión de ayudas

- **Los responsables de la concesión de las ayudas** deben recibir información del sistema, **en lenguaje natural**, que les permita asegurar que está cumpliendo las políticas de concesión de ayudas, el grado de precisión que está teniendo a la hora de predecir las exclusiones sociales que finalmente se producen y que dan lugar a las ayudas, cómo de homogéneas con las predicciones que está proporcionando en familias con características similares, detalles acerca de todas las predicciones realizadas y decisiones tomadas por el sistema de IA.
- **Los técnicos** que implementan el sistema y se encargan de su monitorización mientras está en funcionamiento, deben conocer la misma información anterior, pero además **desde una perspectiva técnica.**
- **A las familias solicitantes,** el sistema deberá poder explicarles **en lenguaje natural** los motivos por los cuales se les ha concedido un determinado importe y no otro, porqué se les ha denegado la solicitud y en qué condiciones aplicadas a su realidad se les aceptaría. Etc.

Ejemplo - Bomba de insulina

- Los **facultativos** responsables de la administración de las dosis sobre sus pacientes deben recibir información del sistema, **en lenguaje natural** con el detalle técnico necesario desde un punto de vista médico, que les permita asegurar que las dosis suministradas están surtiendo el efecto correcto desde un punto de vista médico, atendiendo a los niveles de precisión deseados.
- Los **técnicos software** que implementan el sistema y se encargan de su monitorización mientras está en funcionamiento deben conocer la misma información anterior, pero **desde una perspectiva técnica**.
- **A los pacientes**, el sistema deberá poder explicarles **en lenguaje natural** las dosis que están recibiendo, su efecto, etc., de la misma manera que se lo explicaría el facultativo en una de sus revisiones.

A qué apartados aplica esta medida

- Apartado 1. Diseño y desarrollo para una vigilancia efectiva

4.2.2 Habilitar una interfaz humano-máquina (HMI)

Las medidas implementadas en las fases de diseño y desarrollo del sistema han de tener un reflejo cuando ése está en uso para poder supervisarlas y así garantizar que el funcionamiento es el adecuado. Y esta vigilancia debe poder ser realizada por perfiles técnicos como por los perfiles no técnicos. Dado que los sistemas de IA tienen un reciente despliegue, esta interfaz humano-máquina (HMI) para su monitorización y vigilancia (algo habitual en los sistemas *software* tradicionales) se suele encontrar en fase embrionaria. Este es, junto a la naturaleza de alto riesgo de los sistemas de IA, el motivo por el cual el Reglamento Europeo de la IA pone énfasis en la necesidad de dicha interfaz.

A quién aplica

El proveedor del sistema de IA será quien proporcione dicha interfaz al responsable del despliegue, quien hará uso de ella mientras el sistema está en uso.

Ejemplo

Ver ejemplo de la medida anterior, donde ya se describe la necesidad de una interfaz.

A qué apartados aplica esta medida

- Apartado 1. Diseño y desarrollo para una vigilancia efectiva

4.2.3 Modelo de gobernanza

El concepto de *gobernanza IT* es clave para conseguir el objetivo de vigilancia humana sobre sistemas de IA de alto riesgo. Dicha gobernanza es necesaria realizarla durante el diseño y desarrollo del sistema de IA para garantizar que la creación del sistema de IA está alineada con el reglamento. Pero **cuando el sistema ya está en uso es especialmente crítica la vigilancia del sistema** a través de dicho modelo de gobernanza, garantizando así que el comportamiento del sistema de alto riesgo sigue alineado con el reglamento. De esta manera se proporciona una vigilancia de principio a fin del sistema de IA a lo largo de todo su ciclo de vida.

El modelo de gobernanza debe incluir:

- **Una estructura organizativa multidisciplinar** incluyendo a todos los perfiles que forman parte del ciclo de vida del sistema (responsables de negocio, técnicos, juristas y auditores entre otros).
- **Procedimientos** que permitan supervisar la solución desde que se conceptualiza y diseña **hasta que está en uso** (responsabilidad desde el diseño). En cuanto al momento en que está en uso, si el sistema es uno de los mencionados en el *Anexo III. 1a* (sistemas de identificación biométrica remota), dicho modelo de gobernanza deberá incluir un procedimiento que asegure la verificación separada de dos personas por separado, a excepción de cuando la Unión Europea considere que esa medida es desproporcionada al utilizar dicho sistema para la garantía de cumplimiento del Derecho, migración, control fronterizo o asilo.
- Una parte fundamental de cualquier modelo de gobernanza IT es la **gestión de riesgos.** Es conveniente disponer de un marco de gestión de riesgos asociado al

sistema de IA y realizar evaluaciones periódicas de los riesgos y el impacto del sistema de IA a lo largo de todo el ciclo de vida de éste, **especialmente cuando está en uso.**

- La Inteligencia Artificial y las medidas necesarias para su vigilancia son campos de conocimiento relativamente nuevos para muchas personas. Como parte también del modelo de gobernanza, es necesario **habilitar planes formativos** que permitan capacitar a todas las personas que interactúen con el sistema a lo largo de su ciclo de vida acerca de:
  - Las medidas de diseño y desarrollo que se han tomado sobre dicho sistema, atendiendo a cada uno de los requisitos definidos en el Reglamento.
  - Cómo utilizar las interfaces de usuario que permiten monitorizar las medidas definidas en el diseño y desarrollo.
  - El propio sistema de gobernanza, incluyendo la gestión de riesgos anteriormente descrita.

Dada la criticidad de los sistemas de IA de alto riesgo, dicha formación debería acompañarse de un modelo de evaluación que permita asegurar la asimilación de los conceptos transmitidos.

A quién aplica

- El **proveedor** deberá tener un modelo de gobernanza durante la construcción y uso del mismo, y proporcionar las directrices de gobernanza que el responsable del despliegue deberá llevar a cabo cuando el sistema esté en uso.
- El **responsable del despliegue** adaptará dicho sistema de gobernanza a la realizad de su estructura organizativa y procedimientos.

Ejemplo - Bomba de insulina

Cuando el sistema esté en uso el modelo de gobernanza incluirá:

- Una estructura organizativa multidisciplinar que incluya perfiles que puedan dar cobertura a los tratamientos médicos, soportar la operación técnica del sistema para asegurar su funcionamiento, y gestionar los riesgos que puedan afectar a la salud de los pacientes (derecho fundamental sobre el que pone foco el *Reglamento*).
- Procedimientos que involucren a dichos perfiles y que, integrados en el proceso de negocio soportado por el sistema, permitan supervisar las funcionalidades del mismo tales como, por ejemplo:
  - La recepción y monitorización de los indicadores recogidos online por el sistema sobre el paciente con el objetivo de realizar un seguimiento del tratamiento.
  - La gestión de las alertas accionadas por el sistema o de forma directa por el propio paciente.
  - La validación de las dosis propuestas por el sistema antes de la inoculación que se realizará el paciente.
- Un plan de formación y evaluación continua que asegure:
  - Que los perfiles médicos y técnicos anteriormente descritos están capacitados para entender, usar, supervisar el sistema y gestionar los riesgos para la salud que puedan derivarse de su uso.
  - Que los pacientes entienden igualmente cómo utilizarlo a la hora de inocularse la dosis prescrita por el mismo y validada por el médico.

En ambos casos, prestando especial atención en los momentos en los que se introduce en el sistema una nueva versión o *release* ya sea de datos (por ejemplo, ante la incorporación de nuevas patologías o de nuevos compuestos) o de modelo (por ejemplo, ante la incorporación de una funcionalidad que modifique las pautas horarias de inoculación).

A qué apartados aplica esta medida

- Apartado 1. Diseño y desarrollo para una vigilancia efectiva
- Apartado 5. Identificación biométrica
- Apartado 4b. Sesgo de automatización

4.2.4 Concienciación. Error forzado

El sistema podrá incluir un modo de error forzado que podrá ser activado en la fase de pruebas del mismo, o en cualquier momento en un entorno con datos y casos iguales a los que tendrá el sistema cuando esté en producción. Dicho modo generará de manera controlada y deliberada algunas salidas erróneas, con el objetivo de **testar el criterio y el posible exceso de confianza** de las personas que utilizarán el sistema en producción y así **evaluar su capacidad de vigilancia** sobre el mismo. Se podrán analizar las respuestas realizadas por las personas ante las salidas erróneas deliberadas.

A quién aplica

- El proveedor del sistema proporcionará la funcionalidad de "error forzado" activable únicamente en entornos de pruebas, ya que hacerlo en un entorno productivo sería un factor de inducción a un error real.
- El responsable del despliegue utilizará dicho modo en las pruebas del sistema antes de su salida a producción, y también como actividad de formación a las personas que vayan a utilizar dicho sistema, analizando sus respuestas para poder evaluar su capacidad para supervisar el sistema.

Ejemplo - Bomba de insulina

Uno de los errores en este modo de error forzado consistirá, por ejemplo, en que el sistema proponga al médico una dosis no adecuada a uno de sus pacientes en función de valores anormales en sangre identificados igualmente por el sistema sobre el mismo. El médico deberá identificar dicha situación, marcarla como errónea, e indicar la dosis correcta para que el sistema se lo comunique al paciente, demostrando así que su vigilancia sobre las propuestas de decisión tomadas por el sistema es correcta.

Ejemplo - Concesión de ayudas

Uno de los errores en este modo de error forzado consistirá, por ejemplo, en que el sistema proponga al responsable de las concesiones una horquilla de importe de ayuda no alineada con los ingresos económicos y necesidades de la familia. Dicho responsable deberá identificar dicha situación, marcarla como errónea, e indicar la horquilla correcta, demostrando así que su vigilancia sobre las propuestas de decisión tomadas por el sistema es correcta.

A qué apartados aplica esta medida

- Apartado 4b. Sesgo de automatización

4.2.5 Gobernanza. Human in/on the loop

Una de las decisiones que se ha de tomar en el diseño del sistema de IA es el nivel de autonomía que se le concede al mismo. Dicho nivel de autonomía se enmarca en los aspectos procedimentales a definir en el modelo de gobernanza del sistema de IA. Existen **tres niveles:**

- El primero, y más global que ha de tenerse siempre en cuenta ya que refuerza el concepto de Vigilancia Humana, es el de **Human in Command** (HIC), y se refiere a como un ser humano **es el responsable último** y toma las decisiones críticas en el funcionamiento de un sistema de IA. Este concepto profundiza en la idea de una posición de responsabilidad última de las personas en el funcionamiento del sistema de IA y de las decisiones críticas que se toman en su funcionamiento para garantizar decisiones finales y el logro de los objetivos deseados de manera segura y confiable.
- **Human-in-the-loop** (HITL) se refiere a la intervención humana en cada acción del sistema, lo que en muchos casos no es posible dado el alto volumen de acciones que gestiona. La vigilancia humana en este caso se denomina *ex ante,* e implica que la acción está parcialmente automatizada. En los casos de uso de **alto riesgo** (los que nos ocupan en el *Reglamento*), donde suele ser necesario que una persona valide cada una de las acciones, aunque el volumen de acciones sea alto, este es el nivel que suele ser el **recomendado**.
- **Human-on-the-loop** (HOTL) se refiere a una intervención humana de vigilancia sobre las acciones que es realizada a posteriori. La vigilancia humana en este caso se denomina *ex post,* e implica que la acción está **totalmente automatizada,** aunque se deba habilitar posteriormente la posibilidad de que una intervención humana permita revertirla. Este nivel de autonomía no es especialmente recomendado en los casos de alto riesgo que nos ocupan, debido al riesgo de impacto difícilmente reversible sobre los derechos fundamentales de las personas, uno de los objetivos principales del *Reglamento*.

Independientemente del nivel de autonomía aportado por ambos niveles, el sistema debe disponer de un mecanismo que permita la **desactivación inmediata** del mismo, sin perjuicio de la existencia de un *Plan de continuidad de negocio* que permita que el proceso soportado por el sistema pueda recuperar y restaurar sus funciones.

El nivel utilizado determinará el **nivel de vigilancia humana** que se ejerza sobre el sistema de IA, y tendrá impacto en los procedimientos de gobernanza asociados al sistema cuando esté en uso.

En ambos casos el sistema **registrará las dos acciones:** la propuesta o incluso ejecutada por el sistema, y la ejecutada finalmente por el responsable del mismo, siguiendo las medidas indicadas en la guía del artículo 12 (Registros).

A quién aplica

- Es el responsable del despliegue quien tiene que determinar el nivel de autonomía que se proporciona al sistema en función del caso de uso que soporte dentro de su negocio, así como el procedimiento de gobernanza asociado a dicho nivel cuando el sistema esté en uso.
- El proveedor tiene que implementar los mecanismos técnicos que permitan cada uno de los mencionados niveles.

Ejemplo - Bomba de insulina

Este es un caso donde, debido a que la aplicación inmediata de la decisión pudiera poner en riesgo un derecho fundamental de las personas (la salud), el sistema debe ser diseñado con un enfoque *Human-in-the-loop (HITL)*. Por tanto, cuando el sistema proponga la siguiente dosis a un paciente, su médico responsable deberá validarla o ajustarla en base a los parámetros en sangre del paciente recabados por el sistema, momento en el cual el paciente podrá recargar la bomba de insulina con la cantidad indicada por el sistema y administrársela.

Ejemplo – Concesión de ayudas

Este es igualmente un caso de uso que requiere de un nivel *Human-in-the-loop* (HITL), tanto por la naturaleza funcional "propositiva" del sistema que no requiere inmediatez como, principalmente, por el impacto de la decisión finalmente adoptada. El sistema analiza las peticiones de ayuda realizadas por las familias, y el gestor publico responsable de las mismas podrá analizarlas una a una, estableciendo la cuantía finalmente asignada. Por tanto, cuando el sistema, por ejemplo, proponga la cuantía asignada a una familia, el responsable público deberá validarla en base a la información económico-social recabada de la misma.

A qué apartados aplica esta medida

- Apartado 4e. Autonomía para decidir
- Apartado 4e. Interrupción

4.3 Resumen ejecutivo. Relación Apartado-medidas aplicables

	Medidas						
Apartados	MV1	MV2	MV3	MV4	MV5	MV6	MV7
1. Diseño y desarrollo para una vigilancia efectiva	X	X	X				
2. Riesgos sobre derechos fundamentales, salud y seguridad						X	
3. Tipos de medidas	X	X	X	X	X	X	X
4. Entendimiento y Autonomía							
4a. Entender capacidades y limitaciones							X
4b. Sesgo de automatización			X	X			
4c. Interpretar la información de salida							X
4d. Autonomía para decidir					X		
4e. Interrupción					X		
5. Identificación biométrica			X				

MV1	Medidas de diseño y desarrollo para una vigilancia efectiva
MV2	Habilitar una interfaz humano-máquina
MV3	Modelo de gobernanza
MV4	Concienciación. Error forzado
MV5	Gobernanza. Human in/on the loop
MV6	Artículo 09. Sistema de gestión de riesgos
MV7	Artículo 13. Transparencia

5. Documentación técnica

El Articulo 11 (Documentación Técnica) indica que se habrá de documentar el sistema de modo que demuestre que éste cumple los requisitos establecidos en la sección segunda (a la que corresponde el presente artículo de Vigilancia Humana) proporcionando de manera clara y completa a las autoridades nacionales competentes y a los organismos notificados la información necesaria para evaluar la conformidad del sistema de IA con dichos requisitos.

El mencionado artículo indica que dicha documentación contendrá, **como mínimo**, los elementos contemplados en el **anexo IV.** *[1]

Por otra parte, esta guía de Vigilancia Humana expone medidas para resolver los requisitos expuestos por el Reglamento Europeo de la IA en el artículo dedicado a dicha Vigilancia Humana sobre sistemas de IA. **Como resultado de dichas medidas se pueden documentar** aspectos del sistema que se exponen a continuación, que pueden ayudar a generar la documentación mínima requerida.

Diseño y desarrollo para una vigilancia efectiva

1. Responsable del despliegue. Documento con los requisitos que ha de soportar el sistema de IA en su caso de uso en lo que respecta a la gestión de riesgos del mismo, según los requisitos indicados por al Artículo 9 del Reglamento.
2. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a la gestión de riesgos del mismo, según los requisitos indicados por al Artículo 9 del Reglamento.
3. Responsable del despliegue. Documento con los requisitos que ha de soportar el sistema en su caso de uso en lo que respecta a los datos y la gobernanza del mismo, según los requisitos indicados por el Artículo 10 del Reglamento.
4. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a los datos y la gobernanza del mismo, según los requisitos indicados por al Artículo 10 del Reglamento.
5. Responsable del despliegue. Documento de requisitos que ha de soportar el sistema en lo que respecta a la documentación técnica del mismo, según los requisitos indicados por el Artículo 11 del Reglamento.
6. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a la documentación técnica del mismo, según los requisitos indicados por al Artículo 11 del Reglamento.

[1] Las pymes, incluidas las empresas emergentes, podrán facilitar los elementos de la documentación técnica especificada en el anexo IV de manera simplificada. A tal fin, la Comisión establecerá un formulario simplificado de documentación técnica orientado a las necesidades de las pequeñas empresas y las microempresas. Cuando una pyme, incluidas las empresas emergentes, opte por facilitar la información exigida en el anexo IV de manera simplificada, utilizará el formulario a que se refiere el presente apartado. Los organismos notificados aceptarán dicho formulario a efectos de la evaluación de la conformidad.

7. Responsable del despliegue. Documento de requisitos que ha de soportar el sistema en lo que respecta a los Registros del mismo, según los requisitos indicados por el Artículo 12 del Reglamento.
8. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a los Registros del mismo, según los requisitos indicados por al Artículo 12 del Reglamento.
9. Responsable del despliegue. Documento de requisitos que ha de soportar el sistema en lo que respecta a la Transparencia del mismo, según los requisitos indicados por el Artículo 13 del Reglamento.
10. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a la Transparencia del mismo, según los requisitos indicados por al Artículo 13 del Reglamento.
11. Responsable del despliegue. Documento de requisitos que ha de soportar el sistema en lo que respecta a la Precisión del mismo, según los requisitos indicados por el Artículo 15 del Reglamento.
12. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a la Precisión del mismo, según los requisitos indicados por al Artículo 15 del Reglamento.
13. Responsable del despliegue. Documento de requisitos que ha de soportar el sistema en lo que respecta a la Solidez del mismo, según los requisitos indicados por el Artículo 15 del Reglamento.
14. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a la Solidez del mismo, según los requisitos indicados por al Artículo 15 del Reglamento.
15. Responsable del despliegue. Documento de requisitos que ha de soportar el sistema en lo que respecta a la Ciberseguridad del mismo, según los requisitos indicados por el Artículo 15 del Reglamento.
16. Proveedor. Manuales de usuario y técnicos con los mecanismos para que el sistema de IA soporte los requisitos en lo que respecta a la Ciberseguridad del mismo, según los requisitos indicados por al Artículo 15 del Reglamento.

Habilitar una interfaz humano-máquina (HMI)

17. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA en lo que se refiere a los requisitos del Reglamento en su artículo 9 sobre la gestión de riesgos del mismo.
18. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA en lo que se refiere a los requisitos del Reglamento en su artículo 10 sobre los datos y su gobernanza.
19. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA accediendo a la documentación técnica del mismo exigida por el Reglamento en su artículo 11.
20. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA en lo que se refiere a los requisitos del Reglamento en su artículo12 sobre los Registros.
21. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA en lo que se refiere a los requisitos del Reglamento en su artículo 13 sobre Transparencia.

22. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA en lo que se refiere a los requisitos del Reglamento en su artículo 15 sobre Precisión.
23. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA en lo que se refiere a los requisitos del Reglamento en su artículo 15 sobre Solidez.
24. Proveedor. Manual de uso de la HMI que permita al responsable del despliegue vigilar al sistema de IA en lo que se refiere a los requisitos del Reglamento en su artículo 15 sobre Ciberseguridad.

Modelo de gobernanza

25. Proveedor. Documento con el Modelo de gobernanza sobre el sistema de IA que aplique durante la construcción y uso del mismo y que incluya, al menos, los contenidos indicados en el presente documento acerca de dicho modelo.
26. Responsable del despliegue. Documento con el Modelo de gobernanza sobre el sistema de IA que aplique durante el uso del mismo y que incluya, al menos, los contenidos indicados en el presente documento acerca de dicho modelo.
27. Proveedor y responsable del despliegue. Documento con el modelo de gobernanza que incluya el requisito específico para los casos de Identificación biométrica.

Concienciación. Error forzado

28. Proveedor. Manuales técnicos y de responsable del despliegue que describen al menos la funcionalidad de "error forzado" en el sistema de IA para evitar el sesgo de automatización.

Human in/on the loop

29. Responsable del despliegue. Documento que explicite el nivel de autonomía (HITL/HOTL) que utiliza el sistema de IA en su caso de uso.
30. Responsable del despliegue. Documento con el procedimiento de gobernanza asociado a dicho nivel de autonomía cuando el sistema de IA está en uso.
31. Proveedor. Manuales técnicos y de usuario que describen los mecanismos que permitan al responsable del despliegue utilizar el nivel de autonomía requerido.
32. Proveedor. Manuales técnicos y de usuario con la descripción del mecanismo y/o procedimiento que permita al responsable del despliegue la interrupción del funcionamiento del sistema.

6. Cuestionario de autoevaluación

Para realizar una autoevaluación del cumplimiento de los requisitos del Reglamento de Inteligencia Artificial referidos en esta guía, se ha generado un cuestionario de autoevaluación global con una serie de preguntas con los puntos clave a tener en cuenta respecto a las obligaciones que dictaminan los artículos del Reglamento de IA mencionados en esta guía.

Será necesario referirse a ese documento para realizar el apartado del cuestionario de autoevaluación correspondiente a esta guía.

7. Anexos

7.1 Glosario

El contenido de este documento pretende ser didáctico utilizando un lenguaje entendible y minimizando los tecnicismos, pero a la vez siendo preciso desde el punto de vista técnico y formal. Cuando se utilizan tecnicismos se explican en el mismo texto donde se exponen, pero en otros no se hace así ya que su explicación podría desviar el hilo argumental del documento. En esta sección se detallan dichos conceptos.

Ciclo de vida

El ciclo de vida de un sistema de IA son las fases por las que pasa dicho sistema desde su concepción hasta que es retirado.

Los estándares [ISO/IEC 22989] e [ISO/IEC 5338] definen en profundidad cuales son desde el punto de vista normativo, las fases del ciclo de vida de un sistema basado en IA. Por ejemplo, en el [ISO/IEC 22989:2022, cláusula 6.1] se definen fundamentalmente los siguientes estadios en la vida de un sistema basado en IA:

- Concepción.
- Diseño y desarrollo.
- Verificación y validación del producto o servicio.
- Despliegue.
- Funcionamiento y supervisión.
- Reevaluación.
- Retirada o desmantelamiento.

Fuente**:** ISO.org

8. Referencias, estándares y normas

8.1 Estándares

El presente documento recopila algunas de las recomendaciones de un conjunto de normas internacionales en el campo de la inteligencia artificial, siguiendo un enfoque basado en estándares normativos. Algunos de estos estándares han sido publicados, mientras que otros se encuentran en proceso de desarrollo, según aparece documentado en la publicación del Centro Común de Investigación (JRC), el servicio de ciencia y conocimiento de la Comisión Europea, titulada "*AI Watch: AI Standardisation Landscape. State of play and link to the EC proposal for an AI regulatory framework*".

Los Estándares Normativos, que recogen los contenidos del presente documento son:

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

Financiado por
la Unión Europea
NextGenerationEU

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

Plan de
Recuperación,
Transformación
y Resiliencia

España | digital 2026