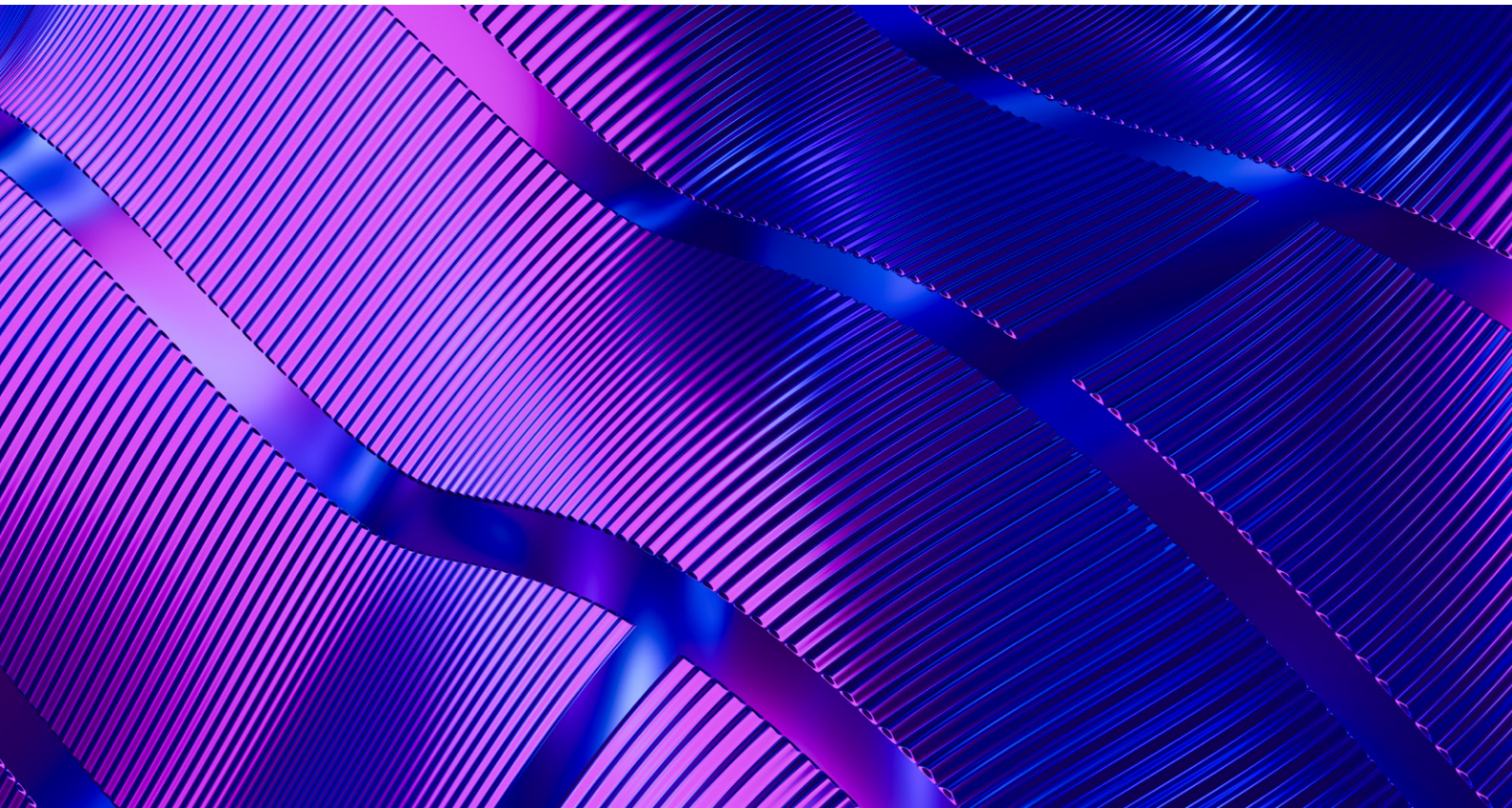
HERBERT SMITH
FREEHILLS
KRAMER

NUEVAS GUÍAS DE LA AESIA PARA CUMPLIR CON EL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL (RIA)

2026



Índice

Introducción

Nuevas Guías de la AESIA para cumplir con el Reglamento Europeo de Inteligencia Artificial

La Agencia Española de Supervisión de la Inteligencia Artificial (**AESIA**) ha publicado 16 Guías de apoyo diseñadas para facilitar el cumplimiento del Reglamento Europeo de Inteligencia Artificial (**RIA**).

Estas guías han sido desarrolladas en el marco del piloto español de *Sandbox* regulatorio de IA, con el fin de servir de apoyo para la implementación y cumplimiento del RIA y proporcionar recomendaciones prácticas y no vinculantes alineadas con los requisitos regulatorios, a la espera de que se aprueben las correspondientes normas armonizadas de aplicación a todos los estados miembros.

Las 16 guías se estructuran en los siguientes bloques:

1. **Guías introductorias (Guías nº 1 y 2):** Proporcionan una visión general y práctica para facilitar una comprensión inicial del RIA mediante materiales de apoyo y ejemplos hipotéticos de casos de uso para ilustrar la aplicación de la normativa;
2. **Guías técnicas especializadas (Guías nº 3 a 15):** Abordan requisitos jurídicos y técnicos específicos aplicables a los sistemas de IA de alto riesgo. Incluyen aspectos críticos como los procedimientos de evaluación de la conformidad, los sistemas de gestión de la calidad, la gestión de riesgos, la gobernanza de datos, la transparencia, la supervisión humana, la precisión, la robustez del sistema y la ciberseguridad, así como documentación técnica; y
3. **Manual de *checklist* (Guía 16):** Proporciona una metodología estructurada en forma de lista de verificación que permite a las organizaciones evaluar fácilmente su nivel de cumplimiento, identificar deficiencias y diseñar un plan de adecuación estructurado.

La AESIA ha indicado que estos documentos no sustituyen a la normativa vigente y serán revisados a medida que se emitan o se vayan desarrollando las normas y directrices europeas. Se espera que se actualicen conforme se aprueben las modificaciones legislativas del "Paquete Ómnibus Digital".

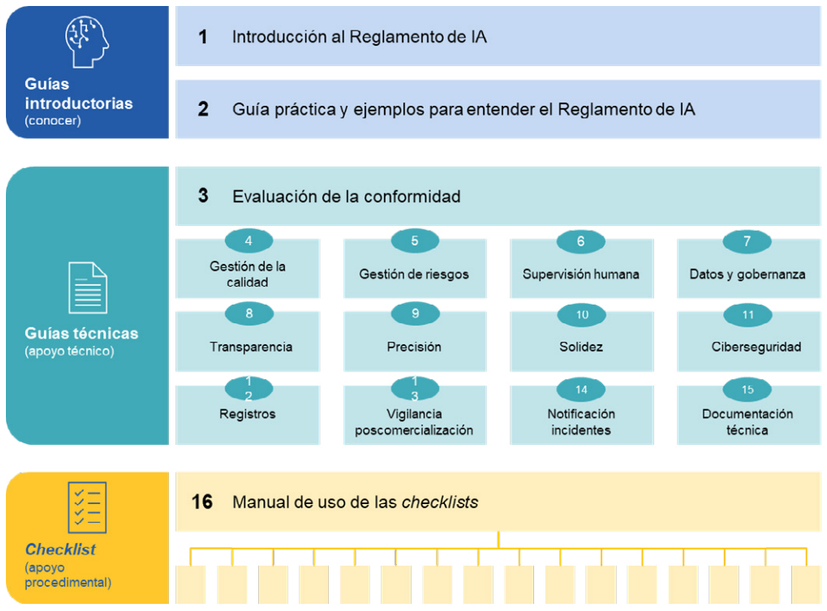


Ilustración. *'Relación de Guías del RIA'. Fuente: Guía n.º 1, p. 4.*

Guías introductorias

Estas guías proporcionan una visión general y práctica para facilitar una comprensión inicial del RIA mediante materiales de apoyo y ejemplos hipotéticos de casos de uso para ilustrar la aplicación de la normativa.

A continuación, exponemos los principales aspectos a tener en cuenta en relación con estas Guías:

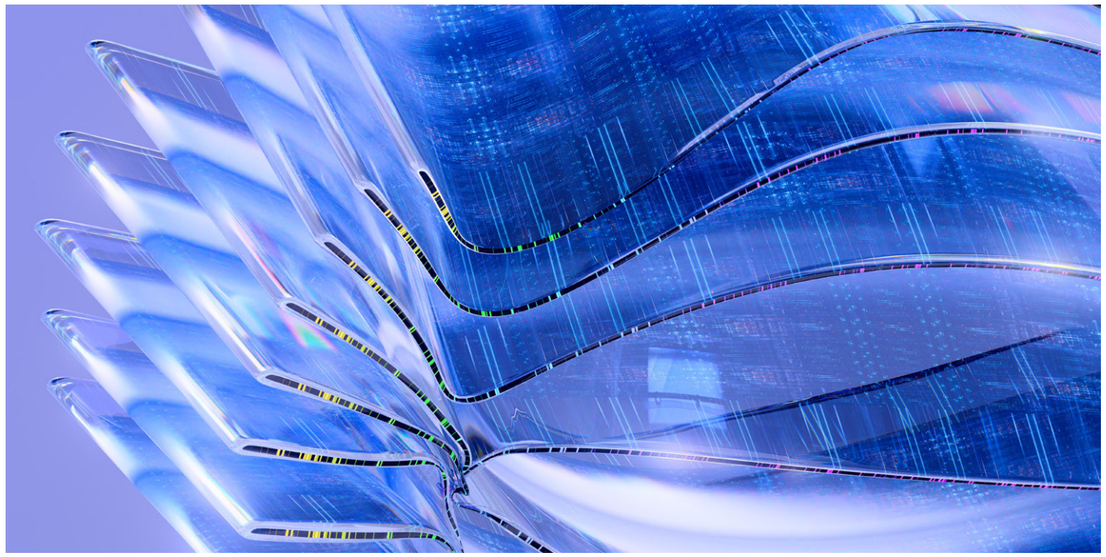
Guía 1

La **Guía 1**, denominada "**Introducción al RIA**", ofrece una visión estructurada del marco normativo y aborda, entre otros, los siguientes puntos clave:

- **Objetivos** del RIA;
- **Ámbito de aplicación y exclusiones:** detalla los principales ámbitos de exclusión bajo el artículo 2 del RIA como, por ejemplo, que el RIA no se aplica a los sistemas o modelos de IA desarrollados exclusivamente para la investigación científica y desarrollo, fines militares, de defensa o de seguridad nacional, ni al uso personal no profesional por parte de personas físicas;
- **Basado en riesgo:** explica el sistema de gobernanza basado en el nivel de riesgo asociado a cada tecnología, con sistemas prohibidos por su nivel de riesgo inaceptable, sistemas de alto riesgo, y resto de sistemas, siendo sobre los dos primeros sobre los que recae el grueso de las obligaciones;
- **Operadores:** Define las principales obligaciones del RIA para los distintos operadores, como proveedores, responsables del despliegue, importadores y distribuidores; y
- **Fomento de la innovación:** Para fomentar la innovación responsable, la guía destaca el papel de los *sandboxes* regulatorios y ofrece condiciones adaptadas para facilitar la integración de pequeñas y medianas empresas (pymes) en este ecosistema.

Guía 2

La **Guía 2**, denominada **"Guía práctica y ejemplos para entender el RIA"**, tiene como objetivo facilitar la comprensión del RIA a través de ejemplos prácticos. En particular, esta guía proporciona ejemplos enfocados en los "sistemas de alto riesgo" como, por ejemplo, sistemas de IA de identificación biométrica en el trabajo, sistemas de IA en la gestión de personal (promoción), o la detección de denuncias falsas. Además, la guía se centra en varios de los conceptos y términos generales descritos en el artículo 3 del RIA más relevantes para un mayor entendimiento pues son utilizados a lo largo del resto de guías y materiales del *sandbox*. La mayoría de los conceptos y términos vienen acompañados con ejemplos que sean aplicables a sistemas de IA. Por último, esta guía muestra un resumen de las obligaciones y funciones que establece el RIA para los proveedores y responsables de despliegue, vinculándolas directamente con las guías técnicas especializadas (Guías 3 a 15) que desarrollan cada requisito específico del RIA.



Guías técnicas especializadas

Estas guías abordan requisitos jurídicos y técnicos específicos aplicables a los sistemas de IA de alto riesgo. Incluyen aspectos críticos como los procedimientos de evaluación de la conformidad, los sistemas de gestión de la calidad, la gestión de riesgos, la gobernanza de datos, la transparencia, la supervisión humana, la precisión, la robustez del sistema y la ciberseguridad, así como documentación técnica.

A continuación, exponemos los principales aspectos a tener en cuenta en relación con estas Guías:

Guía 3

La **Guía 3**, denominada **"Evaluación de la conformidad"**, ofrece una visión detallada y orientativa sobre el proceso obligatorio de evaluación de conformidad al que deben someterse los sistemas de Inteligencia Artificial (**IA**) de alto riesgo bajo el RIA (artículo 43) antes de su comercialización o puesta en servicio en la Unión Europea. Su **objetivo** es proporcionar un itinerario práctico para que el proveedor (el responsable legal del proceso) demuestre que su sistema cumple con los requisitos esenciales de seguridad, transparencia y gobernanza de datos exigidos por el RIA.

Esta Guía da luz sobre los **dos procedimientos principales** para realizar la evaluación de conformidad:

- **Control interno (autoevaluación):** Es la vía general para la mayoría de los sistemas de alto riesgo (puntos 2 a 8 del Anexo III), como los destinados al empleo, educación, y servicios financieros. Bajo este esquema, el proveedor verifica internamente que su Sistema de Gestión de la Calidad (SGC) y su documentación técnica cumplen con el RIA; o
- **Intervención de un tercero (organismo notificado):** Es obligatorio para los sistemas de identificación biométrica (Anexo III, punto 1) cuando el proveedor no ha aplicado, o solo ha aplicado parcialmente, las normas armonizadas o especificaciones comunes disponibles. En estos casos, una entidad independiente debe auditar y certificar el sistema.

También detalla los **requisitos técnicos y documentales** que deben cumplirse para demostrar la conformidad con el RIA, integrando aspectos como la implantación de un SGC, la elaboración de la documentación técnica, y el desarrollo de planes de vigilancia poscomercialización. También proporciona ejemplos y metodologías que facilitan la aplicación de estos requisitos a través de instrumentos como las normas armonizadas y las especificaciones comunes (respectivamente, estándares técnicos creados por organismos de normalización a petición de la Comisión, y reglas técnicas dictadas directamente por la propia Comisión cuando las primeras no existen o resultan insuficientes).

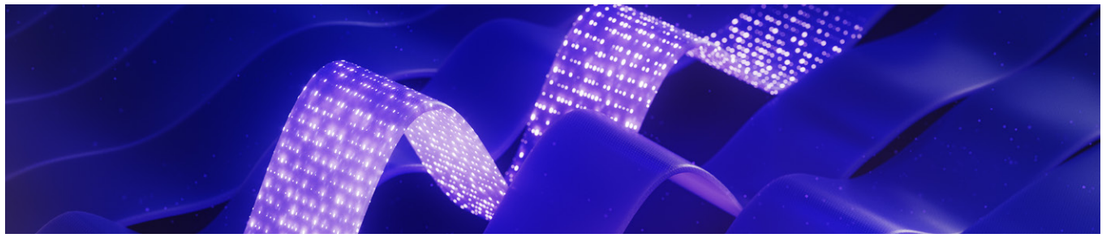
El éxito de este proceso culmina con la elaboración de la **Declaración UE de conformidad** y la colocación del **marcado CE** (indicador clave, pero no una prueba absoluta por sí mismo, de la conformidad de un producto con la legislación de la UE).

Guía 4

La **Guía 4**, denominada **"Sistema de Gestión de Calidad"**, es la pieza que articula todas las obligaciones operativas del RIA dentro de una organización. Tiene como **objetivo** analizar las medidas organizativas y técnicas que servirán a los proveedores (y, en casos específicos de desarrollo conjunto, a los responsables del despliegue) para cumplir con el artículo 17 del RIA. La finalidad del SGC es garantizar que los sistemas de IA de alto riesgo sean seguros, fiables, auditables durante todo su ciclo de vida.

Esta Guía detalla los elementos esenciales que una entidad debe integrar para cumplir adecuadamente con el RIA mediante políticas, procedimientos e instrucciones. Entre los 13 apartados obligatorios establecidos en el artículo 17 destacan: la elaboración de un marco de rendición de cuentas y gobernanza clara que defina las responsabilidades del personal directivo y técnico, procesos documentados para el diseño, desarrollo, pruebas y validación de los sistemas, y medidas de ciberseguridad y precisión integradas desde el diseño.

Un aspecto fundamental es que la implementación del SGC debe ser **proporcional al tamaño de la organización** del proveedor. Este principio busca equilibrar la carga administrativa con la capacidad de la empresa.



Guía 5

La **Guía 5**, denominada **"Gestión de riesgos"**, desarrolla operativamente el cumplimiento del artículo 9 del RIA bajo la rúbrica "Sistema de gestión de riesgos". Este artículo establece la obligación de implementar un proceso iterativo y continuo durante todo el ciclo de vida de los sistemas de IA de alto riesgo para identificar y mitigar peligros potenciales. La responsabilidad de cumplir con esta obligación recae principalmente en el proveedor como encargado del desarrollo del sistema. No obstante, la Guía 5 aclara que, si un responsable del despliegue participa en el desarrollo del sistema, este deberá aplicar las medidas desarrolladas por el proveedor. El foco de este sistema debe ser la protección de la salud, la seguridad y a los derechos fundamentales de las personas.

Para desarrollar un sistema de gestión de riesgos adecuado, la Guía 5 detalla un itinerario compuesto por ocho fases interconectadas.

Ilustración. *'Fases sistema gestión de riesgos'. Fuente: Guía n.º 5, p. 15*

Un punto clave de la guía es la determinación del apetito al riesgo, que se define como el volumen de riesgo que la organización tolera para lograr su misión. Esta valoración se realiza de forma cuantitativa mediante una escala de niveles (eg del 1 al 15). Si un sistema tiene un impacto crítico, como la administración de insulina en el ámbito médico, el apetito al riesgo debe ser muy bajo; por el contrario, en sistemas de bajo impacto como la recomendación de películas basada en gustos, la tolerancia puede ser mucho mayor.

Por último, la guía advierte sobre **peligros críticos** como la discriminación algorítmica y el impacto negativo de los sistemas biométricos. Utilizando casos de estudio, se ilustra cómo sesgos en los datos pueden penalizar injustamente a colectivos por raza o género en el acceso a la salud o el empleo.

Guía 6

La **Guía 6**, denominada **"Supervisión humana"**, se centra en el desarrollo operativo de las obligaciones establecidas en el artículo 14 del RIA. Su objetivo principal es garantizar que los sistemas de IA de alto riesgo estén diseñados de tal manera que puedan ser vigilados de forma efectiva por personas físicas durante su uso.

La supervisión humana no es solo un **requisito técnico**, sino un mecanismo de gobernanza para evitar que la IA socave la autonomía humana o provoque daños a la salud, la seguridad o los derechos fundamentales. La guía subraya que el control humano permite la rendición de cuentas por las acciones del sistema. En concreto, para que el humano entienda realmente el funcionamiento del sistema, la guía establece una serie de medidas que el diseño debe incorporar, incluyendo, entre otras:

- **"Contrafactualidad":** El sistema no debe limitarse a explicar por qué tomó una decisión, sino también detallar qué cambios en los datos de entrada habrían variado el resultado. Esto es esencial para que un supervisor o un afectado comprendan, por ejemplo, bajo qué condiciones se habría aceptado una solicitud denegada.
- **Desglose jerárquico:** La interfaz debe permitir navegar por la información de manera estructurada, desde lo más global (el mecanismo de razonamiento general del modelo) hasta lo más particular (la lógica detrás de una decisión individual concreta).
- Asimismo, la guía explica los tres enfoques fundamentales para calibrar cuánta autonomía se le otorga al sistema:
- *Human-in-the-loop* (**HITL**): El humano interviene en cada ciclo de decisión. Es el nivel recomendado para casos de alto riesgo donde cada acción debe ser validada individualmente.
- *Human-on-the-loop* (**HOTL**): La vigilancia se realiza a posteriori o durante el funcionamiento, permitiendo la intervención si se detecta una anomalía.
- *Human-in-command* (**HIC**): El ser humano mantiene el mando global y la responsabilidad última sobre las decisiones críticas y el funcionamiento seguro del sistema.

Un punto importante es el énfasis en la necesidad de interfaces adecuadas. El proveedor debe facilitar al responsable del despliegue herramientas que permitan interpretar los resultados y los razonamientos de la IA de forma sencilla, visual y en lenguaje natural.

Por último, la guía aborda un **riesgo psicológico crítico**: la tendencia humana a confiar ciegamente en las decisiones de la máquina. Para combatirlo, propone medidas como implementar un modo de "error forzado", es decir, introducir fallos deliberados en entornos de prueba para testar si el supervisor mantiene su criterio crítico y garantizar la concienciación mediante la formación de los supervisores para que entiendan las limitaciones y fallos típicos del sistema.

Guía 7

La **Guía 7**, denominada **"Datos y gobernanza del dato"**, desarrolla operativamente las exigencias establecidas en el artículo 10 del RIA. Su objetivo fundamental es garantizar que los conjuntos de **datos de entrenamiento, validación y prueba** utilizados en sistemas de IA de alto riesgo sean adecuados, pertinentes y suficientemente representativos y cumplen los requisitos de calidad para **evitar sesgos y resultados discriminatorios.**

La gobernanza de datos se define como un conjunto de elementos (políticas, procesos y normas) integrados en un modelo de gestión que abarca **cinco fases críticas del ciclo de vida del dato**, explicadas en detalle en la guía:

1. **Requisitos de información:** Definir con qué información necesitamos alimentar el sistema de IA para lograr el objetivo perseguido.
2. **Recopilación:** Obtención de los datos asegurando su adecuación y representatividad. Es aconsejable que provengan de diferentes fuentes.
3. **Preparación:** Operaciones de etiquetado, depuración, enriquecimiento y transformación.
4. **Disposición:** Puesta a disposición de los datos para el desarrollo del sistema mediante herramientas técnicas adecuadas.
5. **Eliminación:** Retirada segura de los datos una vez cumplida su finalidad.

En relación con la preparación de los datos, la guía indica que es importante decidir en qué fase del ciclo de vida se definen e implementan los controles de calidad. En este sentido, la AESIA recomienda evaluar la calidad directamente en los repositorios de origen y centrar los controles posteriores en la calidad del proceso (verificando que los datos se copian, ingestan o transfieren correctamente). De este modo se evitan controles redundantes en distintas capas y se simplifica su gestión y remediación.

Un punto clave abordado en la guía es el **tratamiento de las categorías especiales de datos personales** (como origen étnico, salud o religión). El artículo 10.5 del RIA permite excepcionalmente el tratamiento de este tipo de datos exclusivamente para detectar y corregir sesgos, sujeto a una serie de condiciones. La guía precisa que la anonimización debe ser siempre la premisa inicial; solo si esta afecta "significativamente" a la precisión en la detección de sesgos se justifica recurrir a la seudonimización.

Por último, la importancia de la **documentación técnica** es otro pilar relevante de la guía. Como buena práctica, la AESIA recomienda, además de incluir los elementos contemplados en el anexo IV RIA en materia de gobernanza de datos, ampliar la documentación incorporando y justificando las fases del ciclo de vida mencionadas más arriba. Se debería especificar cada medida implementada y detallar cómo se ha implementado, además de identificar al responsable de dicha implementación.

Guía 8

La **Guía 8**, denominada **"Transparencia y provisión de información a los usuarios"**, desarrolla operativamente las exigencias del artículo 13 del RIA. Su objetivo central es garantizar que los sistemas de IA de alto riesgo se diseñen y desarrollen de modo que su funcionamiento sea suficientemente transparente para que los responsables del despliegue y los usuarios puedan interpretar los resultados y utilizarlos correctamente.

Esta guía traduce la obligación legal de "ser transparente" a un conjunto de medidas técnicas y documentales que deben cumplir el proveedor y el responsable del despliegue de sistemas de IA de alto riesgo. El **objetivo final** es eliminar la opacidad de los sistemas para que el humano encargado de la supervisión pueda ejercer un control real y efectivo sobre la tecnología.

A continuación, destacamos algunos **puntos clave** establecidos por la AESIA en esta guía para conseguir la transparencia:

- **Instrucciones de uso claras y completas:** La guía detalla el contenido mínimo que deben tener los manuales para que el usuario comprenda las capacidades y limitaciones del sistema, así como su nivel de precisión y los riesgos previsibles para los derechos fundamentales.
- **Diseño para el entendimiento:** Se enfatiza que la información no debe ser solo técnica, sino comprensible para el perfil del usuario que va a operar el sistema. Esto incluye el uso de interfaces que permitan un desglose jerárquico de la información (de lo general a lo particular) y expliquen la contrafactualidad (por qué el sistema no tomó otra decisión distinta).
- **Visibilidad de la muestra de datos:** El objetivo es que el usuario pueda entender y evaluar por sí mismo si la muestra de entrenamiento es justa y representativa para el objetivo específico de su negocio o caso de uso. Para ello es importante enumerar las fuentes de datos utilizadas y realizar un análisis exploratorio de datos para conocer su esencia, metainformación asociada, o los valores críticos o atípicos.
- **Gestión de riesgos por usos no previstos:** Se obliga a documentar no solo el uso previsto, sino a identificar y advertir sobre usos indebidos razonablemente previsibles, proporcionando las métricas necesarias para que el usuario detecte fallos de rendimiento en tiempo real.
- **Canales externos de transparencia:** Como buena práctica para facilitar la comprensión continua, la guía sugiere el uso de medios externos al sistema, como páginas web, wikis o páginas de documentación (*doc pages*), que recopilen de forma accesible toda la información sobre las capacidades y limitaciones de la tecnología.

Por último, la Guía 8 establece una serie de medidas concretas para documentar el cumplimiento por parte del proveedor y del responsable del despliegue con los requisitos de transparencia.

Guía 9

La **Guía 9**, denominada **"Precisión"**, desarrolla la precisión que requiere un sistema IA de alto riesgo conforme a lo dispuesto en el artículo 15 del RIA. La precisión de un sistema de IA es fundamental para poder **mitigar al máximo los potenciales riesgos para la salud, la seguridad y los derechos fundamentales** que pueda implicar la utilización de sistemas de IA de alto riesgo.

A continuación, destacamos algunos **puntos clave** establecidos por la AESIA en la Guía 9 para conseguir que un sistema de IA cumpla adecuadamente con el requisito de precisión:

- **Analizar los aspectos del ciclo de vida** del sistema de IA es esencial porque pueden influir en la precisión final de estos sistemas, siendo el objetivo principal que la precisión sea continua y consistente. En este sentido, para que un sistema de IA consiga una precisión a lo largo del tiempo, se deberá **realizar un preprocesamiento de los datos** (entre otras cosas, se deberá analizar que los datos de entrenamiento del modelo estén libres de sesgo de muestreo y el uso de idénticas maneras de procesado de datos para comparar la precisión entre varios modelos); y **tomar medidas acordes con el tipo de modelo y finalidad prevista para evitar el sobreaprendizaje** (*overfitting*) (por ejemplo, que los hiperparámetros se reporten durante los procesos de entrenamiento/prueba y validación del modelo, así como sus valores para cada modelo, o que ninguna información del conjunto de datos de prueba sea usada cuando se ajusten los hiperparámetros).
- **Seleccionar adecuadamente las métricas para medir la precisión.** La Guía ofrece en su anexo 7.1 unas métricas de precisión (lista no exhaustiva) que pueden utilizarse para medir la precisión, así como, los tipos de modelos con los que se tienen que relacionar. Para el proceso de selección de una métrica de precisión adecuada al sistema, debe tenerse en cuenta dos puntos fundamentales: (i) la finalidad prevista del sistema; y (ii) los riesgos que se hayan encontrado en el sistema de gestión de riesgo, así como seleccionar una función objetivo que pueda permitir alcanzar la finalidad prevista. En este apartado, la Guía destaca la importancia de **tener un repositorio centralizado** en donde se gestione toda la información de métricas asociadas al modelo en cualquier punto de su ciclo de vida.
- **Implementar medidas necesarias para que el proveedor garantice la precisión a lo largo del ciclo de vida de manera consistente.** Para ello, el proveedor deberá implementar, entre otras, medidas técnicas en relación con los inventarios de métricas de precisión y funciones objetivo (por ejemplo: que la salida del sistema esté acompañada por una medida de incertidumbre asociada a la precisión de dicha salida); y utilizar métricas específicas que se acompañen de evaluaciones estadísticas que sirvan para poder considerar la distribución de los datos, la dependencia de estos y otras suposiciones de cada evaluación para así conseguir una evaluación relevante.
- **Documentar adecuadamente el proceso de selección y aseguramiento de la precisión** de acuerdo con toda la documentación técnica restante que se proporciona en la Guía 15 titulada "Documentación técnica". Entre la documentación enfocada en el proceso de selección y aseguramiento de la precisión destaca: la tarjeta del modelo usada por el sistema y la tarjeta de bases de datos que dan una visión más adecuada de las métricas de precisión que se han provisto y de la proveniencia de los datos usados para entrenar el modelo.

Guía 10

La **Guía 10**, denominada **"Solidez"**, continúa desarrollando el artículo 15 del RIA, centrándose en la solidez del sistema de IA de alto riesgo. En relación con la solidez, la Guía considera la **ciberseguridad** como un pilar clave de la solidez porque permite proteger frente a cualquier ataque que pueda alterar la respuesta del sistema de IA y, por ende, su precisión. Asimismo, establece que los mecanismos de solidez deben garantizar que las medidas de ciberseguridad no se degraden con el tiempo.

Para evaluar adecuadamente la solidez de un sistema de IA, se debe tener en cuenta, entre otras cosas, lo siguiente:

- **Que el proveedor de un sistema de IA tiene la responsabilidad de establecer unas métricas adecuadas de solidez.** Para evaluar correctamente la solidez de un modelo, el proveedor deberá seguir los siguientes pasos:
  - Establecer requisitos u objetivos de solidez y métricas asociadas.
  - Planificar experimentos que demuestren la solidez.
  - Realizar experimentos de acuerdo con el plan establecido donde se registren los resultados, los datos utilizados y todos los valores de salida para que calculen las métricas de una manera más agregada.
  - Interpretar los resultados para informar de la decisión.
  - Decidir la solidez del sistema de acuerdo con los criterios e interpretación identificados anteriormente.
- Que para que se consiga un resultado adecuado de solidez, el proveedor debe contar con un **proceso de verificación** para confirmar que se ha cumplido con los requisitos especificados en el diseño, y un **proceso de validación** para confirmar que se cumplen los requisitos de la finalidad prevista del sistema IA utilizando conjuntos de **datos reales** y ejecutando el código.

- Que las métricas escogidas para validar las características de la solidez se deben validar y verificar en **entornos *hardware*** ya que se replican las capacidades de cómputo (memoria, CPU, velocidad de procesado, etc.) finales a las que tendrá acceso el sistema de IA. Asimismo, debe tenerse en cuenta que las métricas de solidez que estén asociadas al hardware deberán estar descritas en la documentación del sistema de IA con anterioridad a que sean diseñadas las pruebas de precisión y rendimiento de la etapa de ejecución del sistema, y, del mismo modo, monitorizar la solidez a través de estadísticas, distribución de datos y cambios en el uso de negocio.

Además, es muy importante que tanto el proveedor como, en su caso, el responsable del despliegue, dispongan de **una solidez técnica** para evitar que cualquier fallo, error o incoherencia pueda tener graves consecuencias para la seguridad del sistema o que afecte de manera negativa a algún derecho fundamental. Es por esto por lo que la Guía recalca la importancia de que el proveedor:

- Establezca estrategias y medidas de predicción de fallos, consecuencias no intencionadas de efecto negativo o que afecten a la seguridad de las personas o a derechos fundamentales.
- Para mitigar la falta de solidez en sistemas de IA frente a errores, fallos o incoherencias, el proveedor deberá:
  - **Organizativamente**, entre otras cosas: implementar técnicas de alerta a las partes responsables o, si no se recibe respuesta de éstas en un tiempo, implementar una función que permita el apagado automático del sistema, e introducir protocolos de fallos seguros que permita a los humanos predecir cualquier percance catastrófico.
  - **Técnicamente**, entre otras cosas: establecer diálogos con diferentes partes involucradas que permitan maximizar la diversidad e inclusión de distintos perfiles de dominio durante el diseño, desarrollo, mantenimiento, aplicación, supervisión y uso del sistema; y elaborar comités de evaluación del diseño del modelo para poder predecir incoherencias de diseño o implementación del sistema que pueda llevar a consecuencias inesperadas no deseadas.
- Para que exista una adecuada solidez, se deberá disponer de **técnicas de redundancia** que permitan alcanzar la solidez del sistema en donde se incluyan planes de respaldo o pruebas de fallos.

Por último, la Guía pone especial foco en la importancia que tienen los sistemas de IA que continúan aprendiendo tras ser comercializados, pues su aprendizaje puede derivar a la posibilidad de que aparezcan sesgos. Es por esto por lo que el proveedor y los responsables del despliegue deberán asegurarse de que el entrenamiento continuado en el tiempo de un sistema de IA no deteriora la solidez a la que se había llegado antes de su comercialización a través de estrategias para mitigar cualquier cambio que implique la degradación de la precisión y solidez del modelo y/o de sus datos.

Guide 11

La **Guía 11**, denominada **"Ciberseguridad"**, desarrolla las medidas que deben implementarse en los sistemas de IA para mitigar los riesgos y ataques a los que pueden enfrentarse a lo largo de su ciclo de vida.

La Guía subraya que, para aplicar eficazmente estas medidas, es imprescindible identificar previamente los tipos de ataques potenciales. A tal efecto, ofrece un esquema que relaciona las distintas fases del ciclo de vida del sistema con los posibles ataques en cada una de ellas:

Ilustración. *'Ciclo de vida del sistema inteligente'. Fuente: Guía n.º 11.*

Este enfoque permite que el proveedor o responsable del despliegue conozca los riesgos y ajuste adecuadamente las medidas de protección. Las principales medidas de ciberseguridad se agrupan en los siguientes bloques:

1. **Medidas de ciberseguridad organizativas**, que deben garantizarse y mantenerse a lo largo del ciclo de vida del sistema:
   - El proveedor deberá, entre otras acciones: planificar de manera global el nivel de ciberseguridad aplicado al sistema de IA durante el proceso de diseño y desarrollo; involucrar desde el inicio del sistema de IA al Delegado de Protección de Datos; acompañar a las instrucciones de uso del sistema de IA las recomendaciones de alto nivel en ciberseguridad aplicada a la IA; designar responsables del seguimiento; en el caso de que el sistema de IA sea entregado al responsable del despliegue en un formato *on-premise* o *in-cloud* gestionado por el responsable del despliegue, el proveedor debe proporcionar instrucciones adecuadas para su protección.
   - En este sentido, las medidas organizativas deben estar alineadas con medidas técnicas. Así, durante los procesos de instalación y/o configuración del sistema y en el manual de instrucciones, el proveedor deberá incluir información relativa a todos los riesgos de ciberseguridad concretos del sistema y cómo este se encuentra protegido. Asimismo, a lo largo del ciclo de vida deberán emplearse herramientas que automaticen las pruebas de seguridad y garantizar que las actualizaciones mantengan un nivel de ciberseguridad consistente y no degradado.
2. **Medidas para reforzar la resistencia del sistema frente a intentos no autorizados** de alterar su uso, resultados de salida o funcionamiento. Este tipo de medidas deben ir apoyadas de **inventarios de activos y de actores** del sistema durante todo el ciclo de vida. Destacamos, entre otras, las siguientes medidas:
   - El proveedor deberá implementar, entre otras medidas: inventariar todos los actores implicados en el proceso; establecer el nivel de accesos y permisos de cada uno de ellos; definir los roles implicados en la utilización de la herramienta; planificar y realizar el inventario de activos incluyendo herramientas, datos, procesos y modelos.

     Como soporte técnico, estos inventarios deberán contar con sistemas informáticos adecuados, mecanismos que permitan aplicar las políticas de acceso y un sistema documental centralizado, actualizado y accesible.
   - El responsable del despliegue deberá conocer los actores y activos que le resulten aplicables. A nivel organizativo, deberá integrar la documentación del proveedor con su organigrama interno y, cuando el sistema sea un activo propio, tratarlo como tal a efectos de ciberseguridad.
3. **Medidas para identificar y mitigar las vulnerabilidades asociadas a los datos de entrenamiento**, que deberán establecer controles de seguridad para evitar la alteración o manipulación:
   - El proveedor, entre otras medidas, deberá implementar controles de seguridad según la vulnerabilidad detectada. Por ejemplo, ampliar los conjuntos de datos mediante técnicas de aumento cuando sean insuficientes o establecer políticas adecuadas de control de accesos si existen deficiencias en la gestión de permisos.
   - Para el responsable del despliegue, la principal acción organizativa para proteger los sistemas frente a ataques de envenenamiento será realizar una lectura y comprensión del manual de instrucciones.
4. **Medidas de protección frente a ataques adversarios**, mediante controles de seguridad específicos:
   - El proveedor deberá, por ejemplo, evitar el uso de modelos ampliamente conocidos cuando exista riesgo de transferencia de ataques adversarios, e integrar la seguridad específica de sistemas de IA en sus estrategias de sensibilización.
   - El responsable del despliegue deberá conocer y analizar todas las vulnerabilidades y controles que sean de su responsabilidad y así asignar recursos humanos y técnicos para mitigarlas.
5. **Medidas frente a ataques dirigidos a descubrir y explotar los defectos del sistema de IA**, ya sean intrínsecos del modelo o derivados de su integración en el entorno software:
   - El proveedor deberá identificar e inventariar los defectos del modelo seleccionado, documentarlos en el modelo de amenazas y aplicar medidas para mitigar su explotación.
   - El responsable del despliegue deberá comprender el manual elaborado por el proveedor en relación con los defectos intrínsecos del sistema y **los mecanismos de configuración que le sean de aplicación de IA dentro del alcance de la finalidad prevista.**

Guía 12

La **Guía 12**, titulada **"Registros y archivos de registro generados automáticamente"**, detalla las medidas necesarias que deben implementar los proveedores y responsables del despliegue para cumplir con los requisitos del RIA en relación con la generación y conservación de registros en los sistemas de IA.

La Guía subraya la importancia de respetar los siguientes principios para una gestión adecuada de registros en los sistemas IA: confidencialidad, integridad, disponibilidad, autenticidad, accesibilidad y trazabilidad, responsabilidad, y retención y eliminación. Asimismo, destaca los siguientes aspectos clave:

- Debe tenerse en cuenta el tipo de agente responsable de los registros. Los proveedores o responsables del despliegue deberán encargarse de la conservación de los registros que genera el sistema siempre que estén bajo su control durante al menos seis meses, salvo que se disponga lo contrario en el Derecho de la Unión o nacional aplicable. En el caso de entidades financieras, los registros deberán conservarse como parte de su documentación obligatoria.

- Los registros deberán reflejar la información que se haya identificado como necesaria tras el proceso de evaluación.
- Cuando se trate de sistemas de identificación biométrica remota, se deberán incorporar elementos específicos y, como mínimo: un registro del período de cada uso del sistema (la fecha y la hora de inicio y la fecha y la hora de finalización de cada uso); la base de datos de referencia con la que el sistema ha cotejado los datos de entrada; los datos de entrada con los que la búsqueda ha arrojado una correspondencia; y la identificación de las personas físicas implicadas en la verificación de los resultados.

Dicho lo anterior, para un adecuado desarrollo y gestión de los registros, se deben abordar, de manera escalada, los siguientes procesos:

1. **Evaluación y diseño de los registros:** consiste en analizar y determinar la necesidad de generar el registro, definir los objetivos específicos para la generación del registro y establecer su alcance. En esta fase se diseña el registro a través de la identificación de campos y categorías para recopilar información. Dentro de este proceso de evaluación y diseño de los registros se deberá: identificar la necesidad, identificar los objetivos, definir el alcance, diseñar el registro e identificar a los responsables del registro. Este proceso debe apoyarse en:
   - Las medidas de la **Guía 5** sobre **gestión de riesgos**, a partir de cuyo inventario se determinará qué sucesos deben evidenciarse mediante registros.
   - La implantación del sistema de **vigilancia poscomercialización** previsto en la **Guía 13**, que permitirá identificar la información necesaria tras la comercialización del sistema.
   - Las medidas de **vigilancia humana** de la **Guía 6**, para identificar la información necesaria que el sistema debe proporcionar a tal efecto.

Este proceso debe revisarse de forma continua, especialmente si se producen cambios en el sistema que afecten al análisis de riesgos.

2. **Captura, almacenamiento y control de acceso:** consiste en capturar, almacenar y conservar los registros definidos en la fase (i) para garantizar su protección frente a cualquier acceso no autorizado, modificaciones no deseadas, pérdidas o destrucción. Para ello, se deberá guardar los registros de manera que se pueda garantizar una protección frente a lo anterior por lo que, entre otras cosas se debe: recabar información en los registros de acuerdo con los criterios establecidos en la fase (i); seleccionar los medios de almacenamiento y materiales de protección adecuados; implementar medidas de ciberseguridad y control de acceso adecuadas; desarrollar y definir roles y responsabilidades sobre la gestión de los riesgos; etc.
3. **Retención y eliminación de los registros:** consiste en el proceso que establece y recoge las necesidades de conservación y destrucción de los registros que han sido creados, capturados y almacenados. Por ello, se debe tener en cuenta que éstos están determinados por dos factores: por un lado, las necesidades de conservar los registros identificados en el proceso (i); y, por otro lado, se debe considerar los requisitos regulatorios o normativos aplicables (por ejemplo, la LOPDGDD y el RGPD en caso de que los registros incluyan datos personales).

   La eliminación de un registro deberá estar autorizada y documentada, debiendo, en todo momento, cumplir con las medidas de seguridad y acceso implementadas. No obstante, aquellos registros que estén en algún proceso legal no podrán ser destruidos hasta que se autorice su destrucción.
4. **Seguimiento y mejora continua:** el objetivo es asegurarse y mejorar la calidad y eficacia del sistema de gestión de riesgos, siendo necesario hacer un seguimiento y establecer unos periodos determinados de revisión y actualización del sistema de gestión de registros. Para ello se deberán establecer las siguientes fases como parte del proceso: monitorización e identificación de posibles errores, análisis de los datos registrados, implementación de mejoras, evaluación de las mejoras, y ciclo de mejora continua.

Por último, la Guía recalca la importancia de establecer las responsabilidades y autorizaciones correspondientes a cada uno de los procesos anteriores. Las responsabilidades deben estar asignadas a todo el personal involucrado en alguno de los procesos y se tiene que reflejar y documentar en descripciones de puestos de trabajo y declaraciones similares cuando corresponda. Todas las responsabilidades tienen que estar documentadas y reflejadas en los documentos.

Guía 13

La **Guía 13**, titulada **"Plan de vigilancia poscomercialización"**, se centra en explicar en qué consisten los sistemas de vigilancia poscomercialización y la importancia de los mismos en los sistemas de IA de alto riesgo. Estos sistemas son un conjunto de procesos y herramientas cuya finalidad es recabar los datos de un sistema para convertirlos en una serie de indicadores sobre su actividad con el objetivo de poder supervisar los sistemas de IA después de que sean lanzados al mercado. Con estos sistemas, el proveedor podrá evaluar si los sistemas de IA cumplen adecuadamente con los requisitos de sistemas de alto riesgo. Estos sistemas funcionan mediante subsistemas: sistemas de captación de indicadores, sistemas de registro de dichos indicadores, sistema de alertas automatizadas, y diferentes interfaces de análisis para los encargados de vigilancia.

Para desarrollar un adecuado sistema de vigilancia poscomercialización, la Guía presenta un resumen sobre las fases de desarrollo de estos sistemas y los componentes que deben tener:

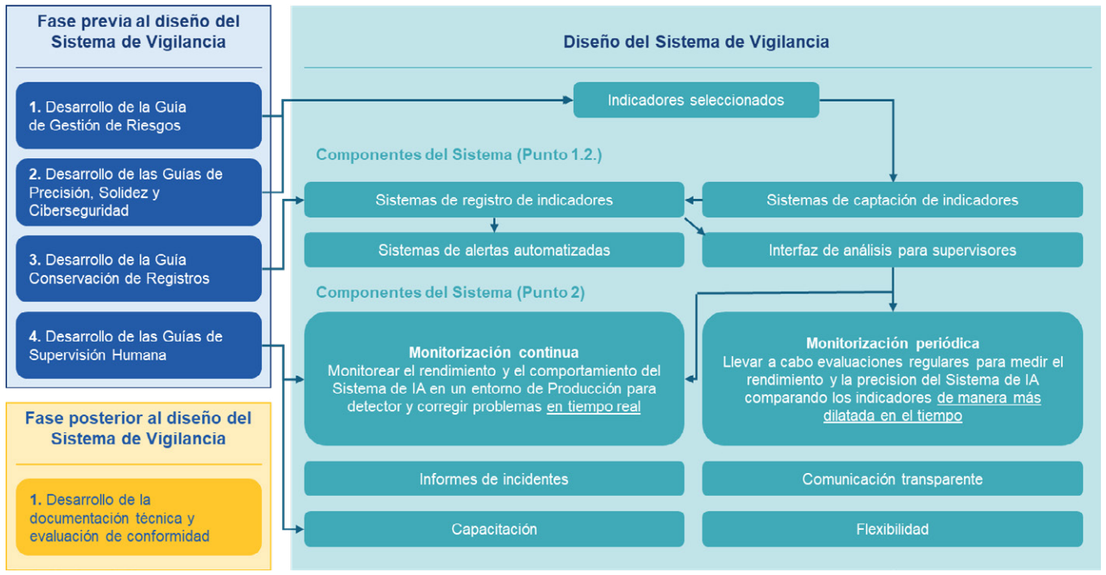


Ilustración. *'Plan de vigilancia poscomercialización'. Fuente: Guía n.º 13.*

Así, para implementar un adecuado plan de vigilancia poscomercialización, se deben implementar las siguientes medidas/actuaciones:

- **Vigilancia continua de los sistemas de IA de alto riesgo** para poder garantizar que el sistema continúe funcionando de manera segura y eficaz una vez esté ya en el mercado. Además, la vigilancia continuada (a través del control de los indicadores del sistema, los indicadores de seguridad, supervisión de los cambios en los indicadores mediante alertas, etc.) es primordial para estar preparados ante cualquier cambio abrupto en el comportamiento del sistema y ante cualquier problema de rendimiento del sistema cono consecuencia de una variedad de factores como, por ejemplo, el envejecimiento de los datos de entrenamiento o la falta de capacitación adecuada.
- **Evaluaciones regulares** (vigilancia periódica) para medir el rendimiento y la precisión del sistema de IA, las cuales permitirán detectar rápidamente problemas y tomar medidas para corregirlos. Para evaluar el rendimiento y la precisión de un sistema, algunas de las medidas para llevarlo a cabo son: las pruebas de rendimiento, que miden el tiempo de respuesta del sistema y su capacidad para manejar grandes cantidades de datos; y las pruebas de precisión, las cuales miden la precisión de los sistemas al realizar tareas específicas, como el reconocimiento de objetos en imágenes o la traducción de idiomas.

- **Comunicación transparente** al receptor de dicha información (eg, proveedor, responsable del despliegue, etc.) sobre las características del sistema, el rendimiento de este y las consecuencias de su uso en producción, para así facilitar una comprensión correcta de todas las implicaciones de su empleo.
- **Capacitación** a los supervisores dándoles formaciones básicas sobre el funcionamiento del sistema de IA y cómo se utiliza.
- **Flexibilidad** a través de un plan flexible y escalable para mejorar la vigilancia del sistema, es decir, adaptar el sistema a los cambios internos y externos que puedan afectar a su funcionamiento. Entre las medidas para llevar a cabo lo anterior están: identificar las regulaciones aplicables, evaluar el rendimiento y la seguridad del sistema, identificar los riesgos de rendimiento y seguridad, vigilar el cumplimiento de la regulación existente, establecer un plan de contingencia, etc.

Por último, la Guía considera fundamental que para que se dé un adecuado sistema de vigilancia poscomercializadora debe tenerse en cuenta el resto de las Guías AESIA. Es por esto por lo que, en su último apartado de esta Guía, se correlaciona los sistemas de vigilancia poscomercialización con el resto de las Guías AESIA.

Guía 14

La **Guía 14**, denominada **"Notificación de incidentes graves"**, detalla el marco procedimental y las medidas operativas que deben implementar los proveedores y, en algunos supuestos, los responsables del despliegue para cumplir con el artículo 73 del RIA.

La Guía subraya la importancia de identificar qué constituye un incidente grave, definido por el artículo 3(49) RIA como aquel defecto de funcionamiento que provoque: el fallecimiento o daños graves para la salud; la alteración grave e irreversible de infraestructuras críticas; la vulneración de derechos fundamentales; o daños graves al medio ambiente o la propiedad.

Destaca los siguientes aspectos operativos clave:

- **Sujetos obligados:** El proveedor es el sujeto principal, con independencia de su origen geográfico, siempre que el sistema opere en el mercado de la UE. El responsable del despliegue asume la obligación de notificar a las autoridades si detecta el incidente y no logra contactar con el proveedor.
- **Jerarquía de plazos:** La notificación debe realizarse inmediatamente después de establecer un vínculo causal entre el sistema y el incidente, respetando los siguientes plazos:
  - **2 días:** En caso de infracción generalizada o incidente relativo a infraestructuras críticas.
  - **10 días:** En caso de fallecimiento.
  - **15 días:** Para el resto de incidentes graves.
- **Notificaciones incrementales:** Se permite presentar inicialmente una notificación incompleta para garantizar la puntualidad, seguida de una completa una vez recabada toda la información.
- **Excepciones por regímenes equivalentes:** Para sistemas sujetos a legislaciones sectoriales de la UE con obligaciones de notificación equivalentes (incluidos los componentes de seguridad de productos sanitarios regulados por los Reglamentos 2017/745 y 2017/746), la notificación se limitará exclusivamente a los incidentes que afecten a derechos fundamentales. Si el sistema opera en varios Estados miembros, la notificación deberá dirigirse a todas las Autoridades de Vigilancia de Mercado (AVM) de los Estados afectados.

Asimismo, la Guía identifica los siguientes procesos para una adecuada gestión:

1. **Evaluación e investigación técnica:** Tras la notificación, el proveedor realizará sin demora una evaluación de riesgos y adoptará medidas correctoras. No podrá modificar el sistema de un modo que pueda repercutir en la evaluación de las causas sin informar previamente a las autoridades. La Autoridad de Vigilancia de Mercado (**AVM**) dispondrá de 7 días para adoptar las medidas adecuadas, que pueden llegar a la retirada o prohibición del sistema, y deberá notificar inmediatamente a la Comisión Europea. Si el incidente afecta a derechos fundamentales, la AVM informará también a los organismos nacionales competentes en la materia.

2. **Integración en la gobernanza y el SGC:** El procedimiento debe formalizarse dentro del Sistema de Gestión de la Calidad (**SGC**) del proveedor. Las medidas operativas clave incluyen: disponer del contacto con la AVM; establecer un canal de comunicación con el responsable del despliegue (articulo 13.3.a); conocer la categorización del sistema para determinar si aplican las excepciones; y conocer los derechos fundamentales de la Unión para identificar cuándo una desviación constituye un incumplimiento notificable.

Guía 15

La **Guía 15**, denominada "**Documentación técnica**", está dirigida exclusivamente al proveedor del sistema de IA por ser el responsable de la elaboración de la documentación técnica del sistema de IA.

La documentación técnica en sí misma es toda la información necesaria que evalúa si el sistema de IA cumple con los requisitos pertinentes y facilitar la vigilancia poscomercialización. Esta información deberá incluir: las características generales, las capacidades y las limitaciones del sistema y los algoritmos, datos y procesos de entrenamiento, prueba y validación empleados, así como la documentación sobre el sistema de gestión de riesgos pertinente, elaborada de manera clara y completa.

La Guía resume, a través de un esquema, los requisitos impuestos por el RIA sobre la documentación técnica:

Durante todo el proceso de la documentación técnica, se tiene que tener en cuenta que:

- La documentación técnica debe ser completa antes de que el sistema de IA se comercialice y se ponga en servicio.
- Para que la documentación esté actualizada durante toda la vida útil del sistema de IA, el proveedor debe tomar medidas como, por ejemplo: dentro de los procesos de gestión del sistema de IA, establecer un procedimiento de seguimiento de los cambios que tenga su reflejo en la actualización de la documentación; generar una cadena de responsabilidad (o designar un responsable de la gestión del cambio sobre el sistema que se encargue de actualizar la documentación de acuerdo con los cambios); o establecer, definir y dimensionar un sistema de gestión documental o solución técnica equivalente, que le permita garantizar la conservación y cambios de esta.
- Se deberá conservar la documentación durante un período de diez años desde la introducción en el mercado del sistema de IA. Por tanto, el proveedor deberá disponer de las medidas técnicas necesarias para preservar dicha documentación y que no exista riesgo alguno de perderse.

Ilustración. *'Documentación técnica'. Fuente: Guía n.º 15.*

Así, la Guía presenta, para ayudar a los proveedores de los sistemas de IA, la estructura que deberá tener la documentación técnica para cumplir con el contenido mínimo dispuesto en el Anexo IV del RIA. Entre otras cosas, se deberá incluir: una descripción general del sistema de IA donde se incluya la finalidad prevista (descripción del uso para el que ha sido diseñado, contexto de utilización del sistema, y las condiciones de uso), el nombre del proveedor y la versión del sistema; la manera en que el sistema de IA interactúa o puede utilizarse para interactuar con *hardware* o *software*, así como con otros sistemas de IA, que no formen parte del propio sistema de IA; o introducciones de uso general dirigidas al responsable del despliegue e instrucciones de instalación.

Manual de *checklist* (apoyo procedimental)

Esta guía proporciona una metodología estructurada en forma de lista de verificación que permite a las organizaciones evaluar fácilmente su nivel de cumplimiento, identificar deficiencias y diseñar un plan de adecuación estructurado.

Guía 16

La **Guía 16**, denominada **"Manual de *checklist* de guías de requisitos"**, tiene como finalidad que las empresas sepan cómo realizar un autodiagnóstico del cumplimiento de todos los requisitos establecidos en el RIA por parte de los sistemas de IA de alto riesgo y poder diseñar un plan de adaptación de sus sistemas a los requisitos establecidos por el mismo.

Esta herramienta es un documento Excel, que se compone de 9 pestañas de las cuales: cinco de ellas son informativas con instrucciones de uso e información de contexto, y cuatro de ellas son pestañas operativas en las que se completa la información requerida.

Las cinco pestañas informativas están formadas por:

- **"Portada":** contiene un recordatorio de confidencialidad obligatorio.
- **"Introducción":** describe de manera resumida los pasos y lo que la herramienta puede hacer.
- **"Artículo RIA":** muestra los apartados del artículo del RIA sobre los cuales la entidad va a hacer autodiagnóstico.
- **"Medidas Guías":** expone las medidas explicativas con detalle recogidas en cada una de las guías. Además, se incluye por cada medida unas cuestiones orientativas con el objetivo de contextualizarla para que la respuesta a las mismas pueda dar una idea de si el sistema ya cumple o no la medida en cuestión.
- **"Relación MG-Apart.":** hace un resumen de la potencial aplicación de las medidas expuestas en la pestaña sobre "Medidas Guías" a cada uno de los apartados del artículo.

Las cuatro pestañas operativas están formadas por:

- **"Autoeval MG.":** permite identificar tanto el nivel de madurez de implantación de la medida propuesta en el sistema, como el nivel de dificultad percibido para llevarla a cabo. Una vez cumplimentada la pestaña, surge el diseño del Plan de Adaptación. Además, si la empresa considera que alguna medida puede servir para ser aplicada a algún apartado adicional, se podrá indicar al final de las filas pre-informadas, debiéndose cumplimentar dos columnas en relación con el nivel de dificultad percibido y el nivel de madurez.
- **"Medidas Adicionales (MA)":** el diagnóstico incluido en esta pestaña, junto con las dos siguientes, consiste en la información de las entidades acerca de las medidas que ellas mismas proponen como posibles para el cumplimiento del RIA, y cuya idoneidad evalúa la SEDIA. La empresa deberá indicar las medidas que, en su experiencia, permiten el cumplimiento de apartados del artículo. Por ello, deberá agregar una fila por cada una de las medidas indicando una descripción breve de la medida y nombre del archivo.
- **"Relación MA-Apart.":** resume de manera ejecutiva la potencial aplicación de las medidas informadas en la pestaña anterior a cada uno de los apartados del artículo.
- **"Autoeval MA.":** es una pestaña que aparece automáticamente pre-informada con una fila por cada una de las relaciones descritas en la pestaña anterior.

Contacto

Pablo García Mexía
Director de derecho digital
T +34 91 423 4010
pablo.garciamexia@hsfkramer.com

Iria Calviño
Socia, Sectores regulados
T +34 91 423 4022
iria.calvino@hsfkramer.com

Jaime Bofill
Socio,
Seguros/ FSR/ Fintech
T +34 91 423 4008
jaime.bofill@hsfkramer.com

Autores

Elena Valín
Asociada
T +34 91 423 4181
elena.valin@hsfkramer.com

Rebeca de Oriol
Asociada júnior
T +34 91 423 4152
rebeca.oriol@hsfkramer.com

For a full list of our global offices visit **HSFKRAMER.COM**