RETECH Redes Territoriales de Especialización Tecnológica

RED ARGOS

Metodología para el análisis y gestión de riesgos de IA en los entornos de Salud y Smart Cities

Septiembre de 2025

Versión 1.0

Junta de Andalucía

Financiado por la Unión Europea NextGenerationEU

Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales

Plan de Recuperación, Transformación y Resiliencia

incibe_ Instituto Nacional de Ciberseguridad

Junta de Andalucía

Agencia Digital de Andalucía

Objeto del Expediente:

“ELABORACIÓN DE GUÍAS DE CIBERSEGURIDAD IA PARA ENTORNOS DE SMART CITIES Y SALUD” (CONTR 2024 829535).

Esta guía forma parte de la colección Guías de Ciberseguridad en IA para las Smart Cities y el sector Salud creada por la Agencia Digital de Andalucía como parte del Proyecto Red Argos, perteneciente al programa RETECH, de Redes Inteligentes de Especialización Tecnológica, en el marco del Plan de Recuperación, Transformación y Resiliencia, financiado por la Unión Europea-NextGenerationEU, a través de INCIBE.

Autor del documento: Agencia Digital de Andalucía

Tipo de documento: Guía

Fecha de elaboración: 08/09/2025

Fecha de última actualización: 25/09/2025

ÍNDICE

Financiado por la Unión Europea NextGenerationEU

ÍNDICE DE ILUSTRACIONES

ÍNDICE DE TABLAS

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

1. INTRODUCCIÓN

La integración de sistemas de Inteligencia Artificial (IA) en los sectores de la salud y las ciudades inteligentes (Smart Cities) constituye uno de los ejes de transformación más significativos de la sociedad digital contemporánea. Estas tecnologías permiten optimizar la gestión de recursos, personalizar la atención sanitaria, mejorar la eficiencia de los servicios públicos y reforzar la resiliencia de infraestructuras críticas.

No obstante, la creciente autonomía y complejidad de los sistemas de IA introducen nuevas categorías de riesgo que trascienden la ciberseguridad tradicional, afectando directamente a la seguridad de las personas, la privacidad, la equidad y la confianza pública. En entornos sanitarios, dichos riesgos pueden manifestarse como errores diagnósticos o terapéuticos automatizados, exposición de datos clínicos sensibles o sesgos algorítmicos que comprometan la equidad asistencial. En contextos urbanos inteligentes, pueden traducirse en vulnerabilidades de infraestructuras críticas, decisiones automatizadas que generen discriminación territorial o social, y afectaciones a derechos fundamentales derivados del uso de tecnologías de vigilancia y control.

La gestión de tales riesgos requiere un enfoque integral que combine la evaluación técnica, la gobernanza ética y el cumplimiento normativo. Este documento se alinea con el Reglamento (UE) 2024/1689 relativo a las normas armonizadas en materia de inteligencia artificial (AI Act), el Reglamento (UE) 2016/679 sobre protección de datos personales (RGPD) y los Reglamentos (UE) 2017/745 (MDR) y 2017/746 (IVDR) sobre productos sanitarios. Además, se fundamenta en los estándares internacionales ISO/IEC 23894:2023 y ISO/IEC 42001:2023, e incorpora la perspectiva de derechos humanos, democracia y Estado de derecho establecida en el marco HUDERIA del Consejo de Europa., garantizando que el desarrollo y uso de la IA en sectores críticos se realice de forma responsable, segura y conforme a los valores institucionales.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

2. OBJETIVO Y ALCANCE

2.1. Objetivo

Este documento tiene como objetivo establecer una metodología integral para el análisis y la gestión de riesgos asociados a los sistemas de Inteligencia Artificial (IA) en los ámbitos de salud y Smart Cities. La metodología está orientada a proporcionar un marco de referencia común que permita identificar, evaluar y tratar riesgos de forma sistemática y verificable dentro de estos entornos, considerando tanto los aspectos técnicos como los regulatorios, éticos y sociales vinculados al uso de la IA en sectores críticos.

El enfoque propuesto se fundamenta en la integración de dos pilares complementarios. En primer lugar, la evaluación técnica del riesgo, basada en la probabilidad de materialización de amenazas específicas de los sistemas de IA y el impacto sobre las dimensiones de fiabilidad, seguridad, privacidad, disponibilidad, equidad y norma social. En segundo lugar, la evaluación de impacto en derechos fundamentales, democracia y Estado de derecho, mediante la adopción del marco HUDERIA (Human Rights, Democracy and the Rule of Law Impact Assessment), desarrollado por el Consejo de Europa, que permite valorar la escala, alcance, reversibilidad y probabilidad de los efectos sobre las personas y colectivos afectados.

El enfoque propuesto facilita la medición objetiva del riesgo, la aplicación de controles adecuados y proporcionales, y la determinación de umbrales de riesgo aceptable que respalden la toma de decisiones. De esta forma, se contribuye así a garantizar que los sistemas de IA se desarrollen, desplieguen y utilicen de manera segura, ética y conforme a las normativas aplicables, al tiempo que se promueve su adopción responsable y sostenible en entornos con impacto directo sobre la ciudadanía.

2.2. Alcance

La presente metodología se aplica a todos los sistemas de Inteligencia Artificial (IA) empleados en los sectores de salud y Smart Cities, con independencia de su nivel de madurez o estado de implantación. Se incluyen en este ámbito los modelos de Machine Learning, Deep Learning e IA generativa, así como los activos asociados que resultan esenciales para su funcionamiento: datos, algoritmos, infraestructuras, interfaces y servicios vinculados.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

El alcance comprende el conjunto del ciclo de vida de los sistemas de IA, desde las fases de diseño y entrenamiento hasta la validación, despliegue, operación, monitorización y retirada, de forma que la gestión de riesgos se aborde de manera integral y continua.

Asimismo, la metodología está dirigida a los actores responsables de la construcción, implantación, operación y supervisión de sistemas de IA, garantizando la integración de los procesos de análisis y gestión de riesgos en las prácticas habituales de seguridad, cumplimiento y gestión tecnológica en los entornos de salud y Smart Cities.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

3. MARCO REGULATORIO

A continuación, se expone el entorno normativo que regula la materia a tratar en la metodología proporcionando una base legal al contenido desarrollado.

3.1 Normativas y marcos de referencia

El marco metodológico está basado en regulación aplicable, estándares, guías y metodologías reconocidas en el ámbito de la gestión de riesgos tecnológicos, tales como:

- ✓ **ISO 31000:2018:** Norma reconocida internacionalmente que ofrece una metodología de análisis y gestión del riesgo de carácter general, proporcionando principios, estructura y proceso para el diseño, implementación, evaluación y mejora continua de la gestión del riesgo en organizaciones.
- ✓ **ISO/IEC 27001:2022:** Estándar para la seguridad de la información que especifica los requisitos necesarios para establecer, implantar, mantener y mejorar un Sistema de Gestión de Seguridad de la Información (SGSI), aplicable a la protección de datos y sistemas críticos en entornos de IA.
- ✓ **ISO/IEC 23894:2023:** Norma reconocida internacionalmente que ofrece un enfoque estructurado para identificar, evaluar y gestionar los riesgos específicos de los sistemas de IA.
- ✓ **ISO/IEC 42001:2023:** Establece requisitos para la creación, implementación, mantenimiento y mejora continua de un Sistema de Gestión de Inteligencia Artificial. La norma ayuda a las organizaciones a desarrollar, implementar y usar sistemas de IA de manera responsable. Enfatiza la integración de sistemas de gestión de IA en estructuras organizacionales existentes y promueve la adopción de prácticas de IA transparentes, seguras y responsables.
- ✓ **Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 (RGPD):** Reglamento relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, relevante para la gestión de datos en sistemas de IA.
- ✓ **NIST AI Risk Management Framework (AI RMF):** Marco de referencia que proporciona directrices específicas para la gestión de riesgos en el desarrollo, implementación y uso de sistemas de IA.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
Plan de Recuperación, Transformación y Resiliencia
incibe_ INSTITUTO NACIONAL DE CIBERSEGURIDAD
Junta de Andalucía
Agencia Digital de Andalucía

- ✓ **Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 (Reglamento de IA):** Reglamento que establece normas armonizadas para el desarrollo y uso de la IA en la UE, incluyendo la gestión de sistemas y riesgos asociados, con un enfoque en la transparencia, supervisión y la protección de derechos fundamentales.
- ✓ **HUDERIA Framework (del inglés, Human Rights, Democracy and the Rule of Law Impact Assessement for AI):** Marco metodológico desarrollado por el Consejo de Europa que proporciona un enfoque estructurado para evaluar el impacto de los sistemas de IA en los derechos humanos, la democracia y el estado de derecho. El marco HUDERIA establece criterios prácticos como escala, alcance, reversibilidad y probabilidad, que permiten identificar riesgos sociales y regulatorios, complementando los marcos técnicos y normativos tradicionales.
- ✓ **Reglamento (UE) 2017/745 sobre productos sanitarios (MDR):** Regulación europeo que establece requisitos para dispositivos médicos, incluyendo software con función médica basado en IA.
- ✓ **Reglamento (UE) 2017/746 sobre productos sanitarios para diagnóstico in vitro (IVDR):** Regula dispositivos de diagnóstico in vitro, incluyendo software con funciones diagnósticas basadas en IA, estableciendo requisitos de validación clínica, evaluación de conformidad y trazabilidad.
- ✓ **MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems):** Marco de conocimiento desarrollado por MITRE Corporation que documenta tácticas, técnicas y procedimientos (TTPs) empleados por actores adversarios contra sistemas de IA. Proporciona una taxonomía estructurada de ataques específicos de IA, complementando MITRE ATT&CK con amenazas propias del aprendizaje automático.
- ✓ **Directiva (UE) 2022/2555 (NIS2):** Directiva relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión, aplicable a entidades esenciales e importantes, incluyendo infraestructuras críticas de Smart Cities y proveedores de servicios sanitarios digitales.

4. RIESGOS ASOCIADOS EN EL CICLO DE VIDA DEL SISTEMA DE IA

El ciclo de vida de los sistemas de Inteligencia Artificial (IA) en entornos de salud y ciudades inteligentes abarca desde su concepción y diseño hasta su despliegue, operación, monitorización y eventual retirada. En cada fase pueden materializarse riesgos específicos que deben identificarse y gestionarse de forma proactiva para garantizar la seguridad, la fiabilidad, la ética y el cumplimiento normativo.

En el sector sanitario, los riesgos pueden traducirse en errores diagnósticos, tratamientos inadecuados o exposición de información clínica sensible. En Smart Cities, pueden derivar en interrupciones de servicios esenciales, vigilancia desproporcionada o decisiones urbanas inequitativas que afecten a colectivos vulnerables.

Los principales riesgos se agrupan en las siguientes categorías:

4.1 Riesgos de cumplimiento normativo y regulatorio

- **Reglamento Europeo de IA (AI Act)**: riesgo de clasificación incorrecta del sistema en su nivel de riesgo (alto, limitado, mínimo), con incumplimiento de requisitos asociados.
- **Protección de datos personales (RGPD/LOPDGDD)**: tratamiento indebido de datos sensibles (ej. historiales clínicos en el sector sanitario o datos de movilidad en Smart Cities).
- **Reglamento (UE) 2017/745 sobre productos sanitarios (MDR)**:
- **Reglamento (UE) 2017/746 sobre productos sanitarios para diagnóstico in vitro (IVDR):**

4.2 Riesgos éticos y sociales

- **Sesgos algorítmicos**: riesgo de generación de resultados o decisiones discriminatorios que afectan de forma desigual a colectivos vulnerables (ej. personas mayores, pacientes con discapacidad o personas en situación de vulnerabilidad social o económica).
- **Transparencia y explicabilidad limitada**: sistemas de “caja negra” que dificultan justificar decisiones médicas o de gestión urbana.
- **Aceptación social y confianza**: pérdida de confianza ciudadana si los sistemas se perciben como invasivos o injustos.
- **Afectación de derechos fundamentales:** riesgo de vulneración del derecho a la privacidad, intimidad, no discriminación, autonomía personal, consentimiento informado (en salud) o libertad de circulación y reunión (en Smart Cities). Conforme al marco HUDERIA, estos impactos deben evaluarse considerando escala, alcance, reversibilidad y probabilidad de materialización.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

4.3 Riesgos operativos y tecnológicos

- **Fiabilidad y calidad de datos**: datos incompletos, ruidosos o mal etiquetados que afectan a la precisión del modelo.
- **Disponibilidad del sistema**: fallos técnicos que interrumpen servicios críticos (ej. IA en urgencias hospitalarias o gestión de semáforos inteligentes).
- **Dependencia de terceros**: vulnerabilidades ligadas a proveedores de software, hardware o servicios en la nube.
- **Actualizaciones y mantenimiento**: cambios no validados que introducen errores o degradan el rendimiento de sistemas en infraestructuras críticas.
- **Deriva del modelo**. Perdida progresiva de rendimiento efectiva del modelo debido a cambios en las distribuciones de datos o en el entorno operativo, afectando la fiabilidad y precisión de predicciones o decisiones automatizadas.
- **Interoperabilidad deficiente:** falta de integración segura y efectiva con sistemas existentes**,** generando riesgos de inconsistencia, autonomía de usuario o errores en la toma de decisiones.

4.4 Riesgos de ciberseguridad

- **Ataques adversariales**: manipulación intencional de datos de entrada para forzar resultados erróneos. En salud, pueden alterar imágenes diagnósticas o parámetros clínicos. En Smart Cities, pueden comprometer sistemas de reconocimiento, vigilancia o control automatizado.
- **Robo o fuga de datos sensibles**: especialmente crítico en salud y movilidad ciudadana.
- **Manipulación maliciosa del modelo**: inyección de datos corruptos que alteren el aprendizaje del sistema.
- **Extracción o ingeniería inversa del modelo**: obtención no autorizada de parámetros, arquitecturas o salidas del modelo mediante consultas reiteradas o análisis adversario, comprometiendo la propiedad intelectual y facilitando ataques posteriores.
- **Ataques de evasión** (adversarial evasión): uso de entradas diseñadas específicamente para engañar al modelo y provocar clasificaciones erróneas, con posibles consecuencias en diagnóstico clínico o detección automatizada en entornos urbanos.
- **Inyección de prompt**: manipulación de interfaces conversacionales o asistentes virtuales basados en IA generativa para obtener información sensible, alterar respuestas o ejecutar acciones no autorizadas.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- **Compromiso de la cadena de suministro**: vulnerabilidades introducidas a través de componentes de terceros (frameworks, modelos preentrenados, servicios cloud) que pueden ser explotadas para comprometer la seguridad del sistema.

4.5 Riesgos de gobernanza y responsabilidad

- **Falta de claridad en roles y obligaciones**: incertidumbre sobre si la responsabilidad recae en el proveedor, el integrador, el hospital/municipio o el operador.
- **Deficiencia en auditorías y trazabilidad**: imposibilidad de reconstruir decisiones automatizadas en caso de reclamaciones legales.
- **Desalineación con objetivos estratégicos**: sistemas de IA que no aportan valor real o generan costes no previstos.
- **Ausencia de comités de ética o supervisión multidisciplinar: falta** de estructuras organizativas que aseguren la revisión ética, técnica, clínica y jurídica de los sistemas de IA antes y durante su operación, comprometiendo la adopción responsable y conforme a valores institucionales.
- **Incumplimiento de obligaciones de transparencia institucional y comunicación:** no informar adecuadamente a pacientes, ciudadanía o partes interesadas sobre el uso de IA, sus finalidades, limitaciones y mecanismos de supervisión, vulnerando principios éticos y normativos de transparencia y consentimiento informado.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

5. ENFOQUE METODOLÓGICO

La metodología de análisis y gestión del riesgo permite obtener como resultado el valor del riesgo asociado a un determinado sistema de IA. A continuación, se identifican y detallan las diferentes fases:

- ✓ **Gestión de inventario de sistemas de IA.** Se identifican y caracterizan los sistemas de IA de los que disponga la Organización, además se les asigna una valoración de su criticidad en base a diferentes dimensiones, así como una clasificación en base al Reglamento de IA. Se incluye también la identificación preliminar de finalidad, grupos afectados y contexto de uso, conforme al marco HUDERIA, para anticipar impactos en derechos fundamentales.

- ✓ **Análisis de amenazas:** Se identifican y caracterizan las diferentes amenazas a las que se enfrentan los sistemas de IA, elaborando un catálogo de amenazas y estableciendo su aplicabilidad en función de la tipología del sistema. Los catálogos de amenazas se han elaborado de forma específica para los entornos de salud y Smart Cities (Anexo I y II), recogiendo tanto amenazas técnicas como aquellas derivadas de fallos organizativos, dependencias externas o factores humanos.

- ✓ **Cálculo del riesgo inherente.** Una vez asociadas las amenazas a los sistemas, se calcula el riesgo inherente (riesgo sin la aplicación de medidas de control) en base al potencial impacto sobre el sistema y la probabilidad de materialización de las amenazas. El cálculo del impacto integra los criterios HUDERIA (escala, alcance, reversibilidad), asegurando la consideración de las afectaciones a derechos fundamentales y confianza social.

- ✓ **Cálculo del riesgo residual:** Tras el cálculo del riesgo inherente, se procede a calcular el riesgo residual, es decir, teniendo en cuenta las medidas de control implementadas. El catálogo de controles aplicables abarca dominios técnicos, organizativos, de cumplimiento normativo, ético y de protección de derechos, adaptados a las especificidades de los entornos sanitario y urbano.

- ✓ **Definición del umbral de riesgo aceptable.** Se debe definir un umbral de riesgo aceptable que permitirá clasificar los riesgos identificados por nivel de tolerancia y obtener una percepción global del riesgo del sistema.

- ✓ **Gestión del riesgo.** Finalmente se selecciona una estrategia de tratamiento del riesgo, en base a la que se definirán los diferentes planes de acción.

Así mismo, se muestra un esquema con el proceso general de análisis y gestión del riesgo de sistemas de IA:

Ilustración 1: Metodología de análisis y gestión de riesgos.

6. FASES DE LA METODOLOGÍA

A continuación, se detallan las diferentes fases del análisis y gestión del riesgo de sistemas de IA identificadas en la metodología.

6.1 Gestión de inventario de sistemas de IA

La primera fase de la metodología de análisis y gestión del riesgo tiene por objetivo la identificación de los sistemas de IA de las diferentes áreas de negocio, así como una caracterización y una valoración de los mismos.

6.1.1 Identificación de sistemas de IA

En primer lugar, se requiere la recopilación de información detallada sobre cada sistema de IA que se está ideando dentro de las distintas áreas de negocio. En caso de existir sistemas de IA en fases posteriores (desarrollo, implementación o uso productivo), se sebe llevar a cabo cuanto antes su identificación para analizar y gestionar los potenciales riesgos asociados. El proceso de identificación puede incluir, entre otros, los siguientes pasos:

- ✓ **Reuniones y cuestionarios:** Definir reuniones (incluso con proveedores) y distribuir cuestionarios a los responsables de diferentes áreas de negocio para obtener información sobre los sistemas de IA. En función del sistema IA también podrá ser necesario implicar a distintas funciones u órganos (p.ej. propietarios del sistema, mantenimiento, responsables clínicos o sanitarios, gestores de infraestructuras urbanas, responsables de protección de datos, funciones de cumplimiento normativo y ética, etc.).
- ✓ **Revisión de planes y proyectos**: Analizar los planes a futuro y proyectos en curso para identificar sistemas de IA.
- ✓ **Colaboración interna**: Fomentar la colaboración entre diferentes departamentos para asegurarse de que se identifiquen y reporten todos los sistemas de IA.

Adicionalmente, se recomienda incluir desde esta fase información sobre:

- ✓ La **finalidad específica del sistema** en el contexto sanitario o urbano (ej. diagnóstico por imagen, gestión de tráfico, optimización energética).
- ✓ **Los grupos sociales**, colectivos o personas que podrían verse directa o indirectamente afectados por el uso del sistema (ej. pacientes, profesionales sanitarios, ciudadanía, colectivos vulnerables).

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- ✓ El **contexto operativo y las condiciones de uso previstas** (ej. entorno hospitalario, atención primaria, vía pública, infraestructuras críticas).
- ✓ Los posibles **impactos en derechos fundamentales**, conforme al marco HUDERIA: derecho a la privacidad, no discriminación, autonomía, consentimiento informado (en salud), libertad de circulación (en Smart Cities), entre otros.

Esta práctica responde a las directrices del marco metodológico **HUDERIA**, desarrollado por el Consejo de Europa, que proporciona criterios de referencia para anticipar impactos en derechos fundamentales, en la calidad democrática y en el Estado de derecho. De esta manera, la metodología incorpora consideraciones de impacto social y legal desde la identificación inicial de los sistemas de IA.

6.1.2 Caracterización de sistemas de IA

Para cada sistema de IA identificado, es crucial recopilar información detallada de modo que sea posible su caracterización. Algunos de los elementos clave a documentar son:

- ✓ **Nombre:** Un nombre único y una breve descripción del sistema de IA.
- ✓ **Objetivo y funcionalidad:** Describir el objetivo principal del sistema de IA y sus funcionalidades clave. En salud, especificar si se trata de apoyo diagnóstico, prescripción terapéutica, monitorización de pacientes, gestión clínica u otro uso asistencial. En Smart Cities, indicar si se orienta a movilidad, energía, seguridad, medio ambiente, servicios ciudadanos u otra función urbana.
- ✓ **Estado de implementación**: Indica el estado de implementación del sistema de IA, especificando si se trata de una prueba de concepto o PoC (implementación acotada y sencilla para comprobar la viabilidad de una idea y que funciona como se espera), un piloto (versión reducida del producto final para evaluar la usabilidad, funcionalidad y diseño), un producto mínimo viable o MVP (versión de prueba con una inversión y tiempo mínimos, capaz de ofrecer valor en producción), o si ya está en producción, operando en un entorno real y a escala completa dentro de la Organización.
- ✓ **Responsables del sistema:** Identifica a los responsables de la gestión y mantenimiento del sistema de IA, así como las personas de contacto clave.
- ✓ **Rol de la entidad con respecto al Reglamento de IA:** Define el papel de la entidad sobre el sistema de IA identificado, clasificándolo en:

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- **Proveedor:** Desarrolla y suministra el sistema de IA, incluyendo soporte técnico y actualizaciones.
- **Responsable del despliegue:** Responsable de la implementación y correcta configuración del sistema en el entorno operativo designado.
- **Representante autorizado:** Asegura la conformidad del sistema de IA con las regulaciones, actuando como enlace entre el proveedor y los organismos reguladores o usuarios finales.
- **Importador:** Introduce sistemas de IA de otros países en el mercado europeo, gestionando la adaptación a las normativas y estándares propios.
- **Distribuidor:** Extiende la distribución del sistema de IA a diversos usuarios y sectores, ofreciendo servicios de mantenimiento y soporte técnico.

✓ **Clasificación con respecto al Reglamento IA:** Categoriza el sistema de IA en base a los niveles de riesgo que el Reglamento IA establece (riesgo inaceptable, alto riesgo, riesgo limitado, riesgo mínimo), lo que determina los requisitos regulatorios y obligaciones específicas, así como medidas que deben aplicarse.

✓ **Exposición:** Indica si el sistema de IA es de uso interno, lo cual significa que solo se emplea dentro de la Organización; o de uso externo, implicando que está disponible para usuarios fuera de la organización. Esto ayuda a determinar el alcance de las medidas de control necesarias para proteger tanto los datos como el propio sistema.

✓ **Tipo de sistema de IA:** Clasifica el sistema según la tecnología subyacente que utiliza: Machine Learning (ML), Deep Learning (DeepL) o Generative AI (Gen AI).

✓ **Modelo de IA:** Especifica el modelo que utiliza en el sistema de IA, en términos de solución/producto y versión asociada, de cara a tener un mejor entendimiento de las amenazas y riesgos existentes.

✓ **Tipo de usuarios:** Tipifica a los usuarios del sistema de IA según su especialización y experiencia, asegurando que las interfaces, funcionalidades y capacitaciones estén adecuadamente adaptadas a sus niveles de conocimiento y necesidades. Concretamente, diferenciando entre usuarios de IT, usuarios de negocio o clientes finales que son externos a la Organización. En salud, debe especificarse si los usuarios son profesionales clínicos (médicos, enfermeros, técnicos), personal administrativo o pacientes. En Smart Cities, si son técnicos municipales, operadores de infraestructuras o ciudadanía en general.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- ✓ **Datos personales y sensibles tratados:** Describe el tipo, volumen y sensibilidad de los datos personales tratados por el sistema, con especial énfasis en datos de salud (historiales clínicos, imágenes diagnósticas, información genética), datos de movilidad, datos biométricos o datos de menores.

6.1.3 Análisis de factores de riesgos

El análisis de factores tiene como objetivo identificar las condiciones internas y externas que pueden aumentar la probabilidad o el impacto de amenazas. En este análisis se emplean los criterios del marco **HUDERIA**, que estructura la evaluación de riesgos a partir de cuatro dimensiones clave:

- **Escala del impacto**: gravedad de las consecuencias que puede generar el sistema sobre individuos, colectivos o servicios esenciales.
- **Alcance**: número de personas afectadas y extensión territorial.
- **Reversibilidad**: grado en que los efectos pueden corregirse o compensarse o restaurarse.
- **Probabilidad**: posibilidad de que el riesgo se materialice, considerando factores técnicos, organizativos y contextuales.

La aplicación de estos criterios permite integrar consideraciones sobre derechos fundamentales, equidad y legitimidad social en la evaluación de riesgos técnicos, alineando la metodología con los principios del Consejo de Europa y el enfoque basado en derechos humanos.

6.1.3.1 Factores internos

- **Datos**: calidad, diversidad, actualización y sensibilidad de los datos empleados. En salud, incluye historiales clínicos, imágenes diagnósticas, resultados de laboratorio y datos de monitorización de pacientes. En Smart Cities, abarca datos de sensores urbanos, cámaras, plataformas de movilidad, consumos energéticos y registros administrativos. La falta de calidad o representatividad de los datos puede introducir sesgos, errores diagnósticos o decisiones urbanas inequitativas.
- **Modelos**: explicabilidad, robustez, sesgos técnicos y resiliencia a ataques adversariales.
- **Procesos organizativos**: existencia y madurez de auditorías, trazabilidad, validación técnica y clínica, supervisión humana, gestión de cambios, protocolos de respuesta a incidentes y mecanismos de mejora continua.
- **Capacidades internas**: nivel de formación del personal, madurez en gobernanza y recursos disponibles.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- **Infraestructuras tecnológicas:** seguridad, disponibilidad, escalabilidad y resiliencia de las infraestructuras que soportan el sistema de IA.
- **Dependencias internas:** grado de acoplamiento del sistema de IA con otros sistemas organizativos esenciales (ej. historiales clínicos electrónicos, plataformas de gestión municipal).

6.1.3.2 Factores externos

Detallar los factores de riesgos externos que pueden presentarse.

- **Regulación aplicable:** cumplimiento del AI Act, RGPD, MDR Y IVDR.
- **Dependencia de terceros:** proveedores cloud, modelos pre entrenados.
- **Entorno social:** confianza ciudadana, percepción pública sobre el uso de IA en los entornos de salud y Smart Cities y legitimidad del uso.
- **Ciberseguridad:** amenazas dirigidas a infraestructuras críticas urbanas o sanitarias.
- **Aspectos económicos y ambientales:** sostenibilidad financiera, costes de desarrollo, despliegue, mantenimiento y actualización, consumo energético e impacto ambiental.

6.1.4 Valoración de sistemas de IA

La valoración de sistemas de IA se enriquece con los criterios **HUDERIA**, que permiten mapear los impactos internos de negocio con riesgos sociales y regulatorios externos. En particular:

- La **Escala** del impacto se refleja en la severidad asignada a cada dimensión (Alta, Media, Baja), evaluando la gravedad de las consecuencias sobre individuos, colectivos o servicios.
- El **Alcance** amplifica la clasificación cuando el impacto afecta a un elevado número de usuarios, pacientes, ciudadanos o servicios esenciales.
- La **Reversibilidad** determina si los daños identificados pueden corregirse, compensarse o restaurase, o si tienen carácter permanente, elevando el nivel de riesgo en este último caso.
- La **Probabilidad** se integra posteriormente en el cálculo del riesgo inherente (apartado 5.2).

Con esta integración, las dimensiones de análisis adoptadas en la presente metodología, siendo **Fiabilidad (F), Seguridad/Privacidad (S/P), Disponibilidad (D), Equidad (E) y Norma Social (NS)** se alinean con el marco HUDERIA, asegurando una coherencia entre la evaluación técnica, el cumplimiento regulatorio y la protección de derechos fundamentales.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

Una vez identificados y caracterizados los sistemas de IA, se debe llevar a cabo su valoración en función de un conjunto de dimensiones, de cara a establecer el nivel de importancia que tiene cada una de ellas para el negocio (en términos económicos, reputacionales, perjuicios a los usuarios, pacientes o ciudadanía, y sanciones regulatorias o contractuales). Para ello, se debe realizar una clasificación sencilla en tres niveles (Alto, Medio o Bajo).

- ✓ **Fiabilidad (F):** Impacto para el negocio (en términos económicos, reputacionales, perjuicios a los usuarios, pacientes o ciudadanía, y sanciones regulatorias o contractuales) en el caso de que las salidas generadas por el sistema de IA no fueran correctas.

Fiabilidad	
Alta (A)	Las salidas incorrectas del sistema pueden causar un impacto significativo a los usuarios, pacientes o ciudadanía. Con pérdidas económicas relevantes, daño reputacional grave, incumplimiento de regulaciones o pérdida de confianza de los clientes. En salud, pueden comprometer la seguridad del paciente o derivar en responsabilidad profesional o institucional. En Smart Cities, pueden afectar a infraestructuras críticas o servicios esenciales.
Media (M)	Las salidas incorrectas del sistema pueden causar un impacto moderado al negocio, afectando a la operación y con pérdidas económicas no relevantes o daño reputacional no grave. En salud, pueden requerir revisión clínica adicional o generar desconfianza puntual. En Smart Cities, pueden ocasionar interrupciones temporales o molestias ciudadanas.
Baja (B)	Las salidas incorrectas del sistema pueden causar un impacto leve al negocio, siendo fácilmente manejable y sin pérdidas económicas (o insignificantes) ni daño reputacional. En salud, no afectan a la seguridad del paciente ni requieren intervención clínica. En Smart Cities, no comprometen servicios esenciales.

Tabla 1: Escala de fiabilidad.

- ✓ **Seguridad/Privacidad (S/P):** Impacto para el negocio (en términos económicos, reputacionales, perjuicios a los usuarios, pacientes o ciudadanía, y sanciones regulatorias o contractuales) debido a una filtración de datos del sistema de IA.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

Seguridad/Privacidad	
Alta (A)	Un incidente de seguridad o brecha de datos tiene consecuencias graves para el negocio, con pérdidas económicas relevantes, daño reputacional grave, sanciones regulatorias o pérdida de confianza de los usuarios, pacientes o ciudadanía, afectando a un gran volumen de datos sensibles. En salud, puede derivar en reclamaciones legales, sanciones de autoridades de protección de datos o pérdida de certificaciones. En Smart Cities, puede comprometer la confianza ciudadana y la legitimidad institucional.
Media (M)	Un incidente de seguridad o brecha de datos tiene consecuencias moderadas para el negocio, con pérdidas económicas no relevantes, daño reputacional no grave, afectando a datos sensibles. En salud, puede requerir notificación a autoridades y a afectados, pero sin comprometer gravemente la seguridad del paciente. En Smart Cities, puede generar quejas ciudadanas y revisión de medidas de seguridad.
Baja (B)	Un incidente de seguridad o brecha de datos tiene un impacto leve al negocio, siendo fácilmente manejable y sin pérdidas económicas (o insignificantes) ni daño reputacional, afectando a datos no sensibles. En salud, afecta a datos administrativos sin relevancia clínica. En Smart Cities, afecta a datos agregados o anonimizados sin riesgo de reidentificación.

Tabla 2: Escala de seguridad/privacidad.

- ✓ **Disponibilidad (D):** Impacto para el negocio (en términos económicos, reputacionales, perjuicios a los clientes y sanciones regulatorias o contractuales) en el caso de una indisponibilidad del sistema de IA.

Disponibilidad	
Alta (A)	La indisponibilidad del sistema tiene un impacto crítico en las operaciones del negocio, afectando a la continuidad de servicios esenciales y generando pérdidas económicas relevantes, daño reputacional grave o penalizaciones significativas de los usuarios, pacientes o ciudadanía. En salud, puede comprometer la atención urgente o crítica. En Smart Cities, puede afectar a infraestructuras críticas o servicios públicos esenciales.
Media (M)	La indisponibilidad del sistema tiene un impacto moderado en las operaciones del negocio, sin afectar a la continuidad de servicios esenciales y generando pérdidas económicas no relevantes, daño reputacional moderado o penalizaciones asumibles de los clientes.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

	En salud, puede requerir procedimientos alternativos manuales. En Smart Cities, puede ocasionar interrupciones temporales de servicios.
Baja (B)	La indisponibilidad del sistema tiene un impacto leve en las operaciones del negocio, sin afectar a la continuidad de servicios esenciales, siendo fácilmente manejable y sin pérdidas económicas (o insignificantes) ni daño reputacional o penalizaciones de los clientes. En salud, no afecta a la atención clínica. En Smart Cities, no compromete servicios esenciales.

Tabla 3: Escala de disponibilidad.

- ✓ **Equidad (E):** Impacto para el negocio (en términos económicos, reputacionales, perjuicios a los clientes y sanciones regulatorias o contractuales) de la materialización de sesgos y comportamientos desiguales por parte del sistema de IA.

Equidad	
Alta (A)	Los sesgos en el sistema pueden tener consecuencias graves para el negocio, con perjuicios graves para ciertos grupos, con pérdidas económicas relevantes, daño reputacional grave y pérdida de confianza de los usuarios, pacientes o ciudadanía. En salud, pueden comprometer la equidad asistencial y derivar en reclamaciones o responsabilidad institucional. En Smart Cities, pueden generar denuncias por discriminación o vulneración de derechos fundamentales.
Media (M)	Los sesgos en el sistema pueden tener un impacto moderado, con consecuencias perceptibles, con pérdidas económicas asumibles, pero sin un daño reputacional crítico. En salud, pueden requerir revisión de protocolos y recalibración del modelo. En Smart Cities, pueden generar quejas ciudadanas y revisión de políticas de despliegue.
Baja (B)	Los sesgos en el sistema pueden tener un impacto leve o no perceptible, con consecuencias mínimas o manejables, sin pérdidas económicas (o insignificantes) ni daño reputacional. En salud, no afectan a la calidad asistencial. En Smart Cities, no generan desigualdades significativas.

Tabla 4: Escala de equidad.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- ✓ **Norma social (NS):** Impacto para el negocio (en términos económicos, reputacionales, perjuicios a los clientes y sanciones regulatorias o contractuales) debido a la generación de respuestas o comportamientos inadecuados por parte del sistema de IA.

Norma social	
Alta (A)	La generación de respuestas inadecuadas puede tener consecuencias graves para el negocio, generando un daño reputacional relevante o denuncias/quejas masivas de usuarios, pacientes o ciudadanía, con posibilidad de generar una pérdida significativa de estos. En salud, puede derivar en pérdida de confianza profesional o reclamaciones éticas. En Smart Cities, puede comprometer la legitimidad de decisiones públicas basadas en IA.
Media (M)	La generación de respuestas inadecuadas puede tener consecuencias moderadas para el negocio, generando un daño reputacional controlado o denuncias/quejas con alcance limitado de clientes, con posibilidad de generar una pérdida acotada de usuarios, pacientes o ciudadanía. En salud, puede requerir revisión de protocolos de comunicación. En Smart Cities, puede generar quejas puntuales y revisión de contenidos.
Baja (B)	La generación de respuestas inadecuadas puede tener consecuencias leves para el negocio, siendo fácilmente manejable, con daño reputacional y denuncias/quejas de clientes nulas o insignificantes, sin posibilidad de generar una pérdida de usuarios, pacientes o ciudadanía. En salud, no compromete la relación clínico-paciente. En Smart Cities, no afecta la percepción ciudadana.

Tabla 5: Escala de Norma social.

6.2 Análisis de amenazas

La segunda fase tiene por objetivo la identificación y caracterización del conjunto de amenazas que afectan a los sistemas de IA.

6.2.1 Identificación de amenazas

En primer lugar, se deben identificar las potenciales amenazas que pueden materializarse en los sistemas de IA.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

La identificación de amenazas tiene como objetivo determinar los eventos que pueden afectar a los sistemas de Inteligencia Artificial (IA) a lo largo de su ciclo de vida. Esta fase constituye la base para la caracterización y evaluación del riesgo, ya que establece el conjunto de situaciones adversas que deben ser consideradas en el análisis.

6.2.1.1 Fuentes de identificación

Para la identificación de amenazas, la metodología se apoya en dos elementos principales:

1. **Catálogos sectoriales incluidos en los anexos de la guía:** Los anexos incorporan catálogos de amenazas específicamente elaborados para los entornos de salud y Smart Cities. Estos catálogos sirven como punto de partida en el proceso de identificación, proporcionando un repertorio adaptado al contexto sectorial y facilitando la cobertura inicial de amenazas relevantes. Los catálogos han sido desarrollados considerando las particularidades operativas, regulatorias, éticas y de ciberseguridad propias de cada sector, asegurando que las amenazas identificadas reflejen los riesgos reales a los que se enfrentan los sistemas de IA en entornos sanitarios y urbanos en base al marco regulatorio el apartado 3.
2. **Marcos y estándares internacionales:** Además de los catálogos sectoriales proporcionado en los anexos, la identificación puede complementarse con referencias internacionales que permiten ampliar el análisis:
    - ✓ MITRE ATLAS / MITRE ATT&CK for AI: repositorios que documentan tácticas, técnicas y procedimientos empleados contra sistemas de IA, especialmente útiles para el análisis de amenazas de ciberseguridad avanzadas.
    - ✓ ISO/IEC 23894:2023: norma internacional que establece un marco de riesgos para sistemas de IA y proporciona una taxonomía de amenazas a lo largo de su ciclo de vida.
    - ✓ ISO/IEC 27032:2012: estándar que ofrece un catálogo de amenazas comunes en ciberseguridad (phishing, malware, ransomware, denegación de servicio, fuga de información), aplicables también a los entornos donde se integran sistemas de IA.
    - ✓ HUDERIA Framework: enfoque desarrollado por el Consejo de Europa que permite identificar riesgos de impacto en derechos fundamentales, democracia y Estado de derecho, complementando el análisis técnico con una perspectiva social y regulatoria.

6.2.1.2 Metodología de identificación

La metodología de identificación de amenazas comprende las siguientes etapas:

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

1. Crear un Inventario de activos para determinar los activos que forman parte del sistema de IA, incluyendo:
    - **Datos** (clínicos, de movilidad, de sensores, etc.).
    - **Modelos y algoritmos** (machine learning, deep learning, IA generativa).
    - **Infraestructuras tecnológicas** (servidores, entornos cloud, redes críticas).
    - **Interfaces y APIs** (conexiones internas y externas con otros sistemas).
    - **Usuarios y operadores** (personal sanitario, gestores municipales, ciudadanos, técnicos especializados).
    - **Documentación y evidencias:** registros de validación clínica, certificaciones regulatorias, documentación de cumplimiento normativo.
2. Elaboración de la matriz Amenaza–Activo:
    - A partir del inventario, se elabora una matriz en la que se cruzan los activos identificados con las amenazas extraídas de los catálogos sectoriales (Anexos I y II) y, de forma complementaria, de marcos internacionales como MITRE ATLAS o ISO/IEC 23894.
    - La matriz permite establecer de forma explícita qué amenazas afectan a cada activo y asegura que el análisis cubra todos los elementos de la solución.

6.2.1.3 Consolidación de amenazas identificadas

- El resultado de la matriz se traduce en un listado consolidado de amenazas aplicables al sistema de IA. Este listado debe integrar tanto las amenazas provenientes de los catálogos sectoriales de los anexos I y II, como las identificadas mediante marcos internacionales.
- La consolidación debe ser validada con las partes interesadas relevantes (responsables de seguridad, cumplimiento, operación y usuarios especializados), garantizando que el conjunto de amenazas refleje adecuadamente el contexto real de operación del sistema.
- La validación debe documentarse formalmente, incluyendo actas de reunión, criterios de priorización y justificación de amenazas excluidas o incorporadas para su trazabilidad durante las auditorias regulatorias.

6.2.2 Caracterización de amenazas

Una vez identificadas las amenazas aplicables a sistemas de IA, se debe llevar a cabo una caracterización estructurada de las mismas, estableciendo:

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- ✓ **Degradación producida:** Nivel de perjuicio sobre el sistema de IA, es decir, daño causado por una amenaza en el supuesto de que se materialice la amenaza. Esta degradación se caracteriza como una fracción del valor del activo:
  - Cuando las amenazas no son intencionales, es suficiente con conocer la fracción perjudicada de un activo para calcular la pérdida proporcional de valor que se pierde.
  - Cuando la amenaza es intencional, no se puede pensar en proporcionalidad alguna pues el atacante puede causar daño de forma selectiva.

Así mismo, los posibles valores de la degradación de la amenaza sobre el activo pueden ser:

Degradación	
Muy alta (MA)	Si la amenaza se materializa, el sistema resulta totalmente inservible.
Alta (A)	Si la amenaza se materializa, el sistema resulta prácticamente inservible.
Media (M)	Si la amenaza se materializa, el sistema resulta funcionalmente degradado, con un rendimiento bajo.
Baja (B)	Si la amenaza se materializa, el sistema dispone de una ligera degradación que no impide el funcionamiento.
Muy baja (MB)	Si la amenaza se materializa, el sistema sigue en perfecto estado tras la materialización de la amenaza.

Tabla 6: Escala de degradación.

- ✓ **Probabilidad de materialización:** Se refiere a la posibilidad de que una amenaza específica se convierta en un evento, afectando así al sistema. Para evaluar la probabilidad de que una amenaza afecte a un sistema de IA, se utiliza la siguiente escala, que clasifica la frecuencia esperada de materialización basándose en observaciones y análisis del entorno de amenazas:

Probabilidad		
Muy alta (MA)	5	Al menos una vez al semestre. La naturaleza de la amenaza hace que la materialización de la misma pueda ser muy posible y repetible.
Alta (A)	4	Al menos una vez al año. La naturaleza de la amenaza hace que la materialización de la misma sea posible y repetible.
Media (M)	3	Al menos una vez cada dos años. La naturaleza de la amenaza hace que la materialización de la misma sea posible.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

Baja (B)	2	Al menos una vez cada cinco años. La naturaleza de la amenaza hace que la materialización de la misma sea poco frecuente.
Muy baja (MB)	1	Menos de una vez cada cinco años. La naturaleza de la amenaza hace que la materialización de la misma sea inusual.

Tabla 7: Escala de probabilidad.

6.2.3 Asociación de amenazas con sistemas de IA

Es necesario determinar cuáles son los sistemas que pueden verse afectados por cada amenaza. Para ello se deben caracterizar las amenazas según la tecnología subyacente de los sistemas a los que aplicaría cada amenaza:

- ✓ Machine Learning (ML)
- ✓ Deep Learning (DeepL)
- ✓ IA generativa

Asimismo, debe especificarse el sector de aplicación (salud, Smart Cities) y, cuando sea relevante, el subsector concreto (ej. diagnóstico por imagen, urgencias hospitalarias, movilidad urbana, gestión energética) con el fin de aplicar los catálogos de amenazas específicos de los Anexos I y II.

6.3 Cálculo de riesgo inherente

Tras llevar a cabo el análisis de las amenazas que afectan a cada uno de los sistemas de IA, se calcula el riesgo inherente, el cual es el resultado del cálculo de la probabilidad de la materialización de la amenaza y el impacto producido sobre los sistemas, sin tener en cuenta las medidas de control implementadas por la Organización.

6.3.1 Determinación del impacto

El potencial impacto sobre un sistema de IA se calcula a partir de dos parámetros:

- ✓ **Valor del sistema**. Ver el apartado 6.1.4 de Valoración de sistemas de IA. En esta determinación, el "valor del sistema" debe interpretarse conforme a la clasificación de dimensiones realizada en el apartado anterior, la cual se ha construido aplicando los criterios HUDERIA (escala, alcance, reversibilidad). Esto asegura que el impacto no solo se mida en términos de negocio, sino también considerando sus efectos sobre derechos fundamentales y confianza social. Para el cálculo del impacto se emplea la valoración más restrictiva (negativa) de todas las dimensiones evaluadas para el sistema de IA analizado.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
Plan de Recuperación, Transformación y Resiliencia
incibe_ INSTITUTO NACIONAL DE CIBERSEGURIDAD
Junta de Andalucía
Agencia Digital de Andalucía

En el sector sanitario, la valoración debe reflejar el impacto potencial sobre la seguridad del paciente, la calidad asistencial, la privacidad de datos clínicos y el cumplimiento de obligaciones regulatorias según el MDR y IVDR. En Smart Cities, debe considerar el impacto sobre infraestructuras críticas, servicios públicos esenciales, seguridad ciudadana y equidad territorial.

- ✓ **Degradación producida**. Ver el apartado 6.2.2 de Caracterización de amenazas.

A continuación, se muestra la fórmula empleada para el cálculo del impacto:

$$Impacto = Valor\ del\ sistema\ x\ Degradación$$

A continuación, se muestra la matriz de representación del nivel de impacto:

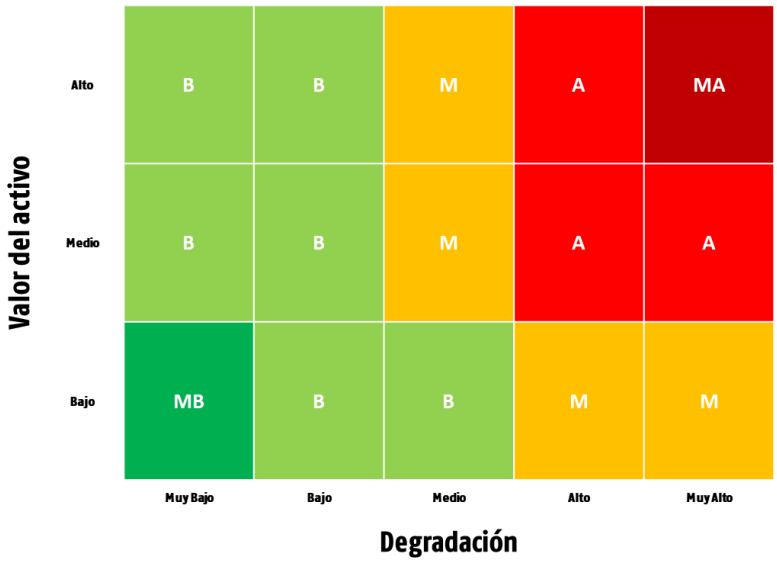

Ilustración 2: Nivel de impacto (valor del sistema x Degradación).

Así mismo, como resultado de la matriz anterior, se detalla la escala de impacto:

Impacto		
Muy alta (MA)	5	Si el sistema es considerado de categoría alta y tiene una degradación sobre el mismo que resultaría totalmente inservible.
Alta (A)	4	Si el sistema es considerado de categoría alta y/o media, y tiene una degradación sobre el mismo que resultaría prácticamente inservible.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

Media (M)	3	Si el sistema es considerado de categoría alta y/o media y la materialización de la amenaza sobre el mismo resultaría funcionalmente degradado, o si el sistema es considerado de categoría baja, y tiene una degradación sobre el mismo que resultaría prácticamente inservible.
Baja (B)	2	Si el sistema es considerado de categoría alta, medio y/o baja, teniendo una ligera degradación sobre el mismo que no impide su funcionamiento.
Muy baja (MB)	1	Si el sistema es considerado de categoría baja y sigue en perfecto estado tras la materialización de la amenaza.

Tabla 8: Escala de impacto.

6.4 Evaluación del riesgo inherente

Finalmente, conociendo el impacto y la probabilidad de las amenazas sobre los sistemas, se debe calcular el nivel de riesgo inherente de manera individual de cada una de las amenazas, seleccionando como referencia la amenaza que genere un mayor nivel de riesgo inherente.

A continuación, se muestra la fórmula empleada para el cálculo del riesgo:

$$Riesgo\ Inherente = \frac{Impacto\ \ x\ Probabilidad}{Nº\ total\ de\ niveles\ de\ riesgo}$$

A continuación, se muestra la matriz de representación del nivel de riesgo:

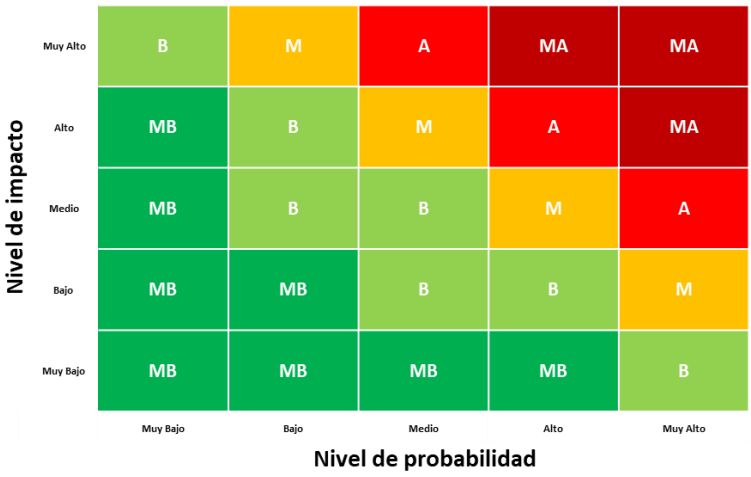

Ilustración 3: Nivel de riesgo (Nivel de impacto x Probabilidad).

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

A continuación, se detalla la escala del nivel de riesgo:

NIVEL DE RIESGO		
MA	Muy alto	4<= R <=5
A	Alto	3<= R <4
M	Medio	2<= R <3
B	Bajo	1<= R <2
MB	Muy bajo	0<= R <1

Tabla 9: Nivel de riesgo.

Conforme al marco HUDERIA, los riesgos clasificados como ***Muy alto*** o ***Alto*** que afecten a derechos fundamentales, equidad o transparencia deben ser objeto de atención prioritaria, contemplándose estrategias de mitigación reforzada antes de proceder al despliegue o continuidad del sistema de IA en entornos de salud o Smart Cities.

6.5 Cálculo de riesgo residual

Tras calcular el riesgo inherente, se da paso a la identificación y evaluación de medidas de control sobre los diferentes sistemas de IA, permitiendo así obtener una reducción del riesgo inherente, mediante la reducción de la probabilidad de que ocurra el riesgo y/o del impacto de las amenazas, dando lugar a lo que se denomina riesgo residual.

6.5.1 Identificación de medidas de control

El primer paso para calcular el riesgo residual consiste en la identificación del conjunto de medidas de control que pueden mitigar las amenazas identificadas.

Dichas medidas de control están definidas a partir de los requerimientos o controles establecidos en las referencias metodológicas del apartado 2. El Anexo III de la presente metodología proporciona un catálogo consolidado de controles aplicables a sistemas de IA en salud y Smart Cities, estructurado en dominios técnicos, organizativos, de cumplimiento normativo, ético y de protección de derechos. Los controles han sido diseñados para abordar las amenazas específicas identificadas en los Anexos I y II, garantizando coherencia entre amenazas, riesgos y medidas de mitigación

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

La identificación de controles debe realizarse de forma sistemática, cruzando cada amenaza con los controles del catálogo que resulten aplicables. En el proceso deben participar responsables técnicos, clínicos (en salud), de infraestructuras urbanas (en Smart Cities), de seguridad, protección de datos, cumplimiento y ética, asegurando una cobertura integral de los riesgos.

6.5.2 Evaluación de medidas de control

Una vez identificadas las medidas de control aplicables, se da paso a la evaluación de su madurez y eficacia. La evaluación de madurez debe realizarse de forma objetiva, basándose en evidencias documentales, auditorías técnicas, registros de operación o certificaciones externas.

Dicha evaluación se realiza por cada medida de control aplicable a un sistema de IA. A continuación, se muestra la escala de valores en términos de madurez y eficacia:

Nivel de madurez	Eficacia	Detalle
N.A.	-	✓ La medida de control no aplica, por lo que no se tiene en cuenta a la hora de realizar el cálculo del riesgo residual.
L0 Inexistente	0%	✓ Se considera que la medida de control es necesaria, pero no se lleva a cabo dentro de la Organización.
L1 Inicial/ad-hoc	10%	✓ Existe evidencia de la necesidad del control y es necesario tratarlo. Sin embargo, no existen iniciativas estándar y/o centralizadas. En su lugar, enfoques ad hoc que tienden a ser aplicados de forma individual o caso por caso. Existen carencias respecto a un enfoque homogéneo. ✓ Estado inicial donde el éxito de la medida de control se basa, en la mayoría de las ocasiones, en el esfuerzo del personal. ✓ Los procedimientos son inexistentes o están localizados únicamente en áreas concretas.
L2 Repetible, pero intuitivo	25%	✓ Las medidas de control se establecen de una forma similar en diferentes sistemas o procesos, pero no se encuentran definidas ni documentadas. ✓ Se normalizan las buenas prácticas sobre la base de la experiencia.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

Nivel de madurez	Eficacia	Detalle
L3 Proceso definido	60%	✓ La realización de las actividades se ha estandarizado, documentado y se han difundido. Sin embargo, se deja que el personal decida cómo realizarla. La realización de las actividades en sí no es sofisticada, pero formalizan las prácticas existentes.
L4 Gestionado y Medible	90%	✓ Cuando se dispone de un sistema de medidas y métricas para conocer el desempeño (eficacia y eficiencia) de los procesos. Las áreas de negocio son capaces de establecer objetivos cualitativos a alcanzar y disponen de medios para valorar si se han alcanzado los objetivos y en qué medida.
L5 Optimizado	95%	✓ La realización de las actividades de control se ha refinado hasta un nivel de mejor práctica, se basan en los resultados de mejoras continuas y en un modelo de madurez en el que se tienen en cuenta otras organizaciones. Las actividades de control están integradas y automatizadas en el flujo de trabajo, brindando herramientas para mejorar la calidad y la efectividad. ✓ En base a criterios cuantitativos se determinan las desviaciones más comunes para poder optimizar la efectividad de las medidas de control.

Tabla 10: Nivel de madurez y eficacia.

6.5.3 Evaluación del riesgo residual

Finalmente, a partir del riesgo inherente, una vez evaluadas las medidas de control aplicables y conociendo sus niveles de madurez, se obtiene el nivel de riesgo residual. A continuación, se muestra la fórmula empleada para el cálculo del riesgo:

$$Riesgo\ Residual = Riesgo\ Inherente \times \left(1 - \frac{\sum Nº\ de\ medidas\ de\ nvl.de\ madurez\ Li\ \times Eficacia\ Li}{Nº\ de\ medidas\ de\ control}\right)$$

A continuación, se detalla la escala del nivel de riesgo:

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

Nivel de riesgo		
MA	Muy alto	4<= R <=5
A	Alto	3<= R <4
M	Medio	2<= R <3
B	Bajo	1<= R <2
MB	Muy bajo	0<= R <1

Tabla 11: Nivel de riesgo.

6.6 Definición del umbral de riesgo aceptable

Posteriormente, se debe definir un umbral como el nivel máximo de riesgo que la organización está dispuesta a asumir. Este umbral puede variar según la Organización, su contexto operativo, su tolerancia al riesgo y factores regulatorios o de mercado.

Una vez establecido dicho umbral, se llevará a cabo un estudio de los resultados obtenidos a partir del análisis, comparando los riesgos residuales con el umbral de tolerancia definido. Dicho estudio, permitirá clasificar los riesgos identificados por regiones de tolerancia y obtener una percepción global del riesgo de la Organización.

Los riesgos identificados pueden ser clasificados en las siguientes categorías:

- ✓ **Riesgos aceptables:** Riesgos que se consideran tolerables y que no precisan un plan de actuación.
- ✓ **Riesgos inaceptables:** Riesgos que no se pueden tolerar salvo en circunstancias extraordinarias. Es necesario, por tanto, aplicar alguna estrategia de tratamiento de manera inmediata. Estos riesgos superan el umbral de riesgo aceptable que ha definido la organización.

El umbral de riesgo aceptable debe ser definido anualmente, en función de los resultados obtenidos y teniendo en consideración los resultados obtenidos en análisis de riesgos anteriores. En el ámbito sanitario, el umbral debe alinearse con las exigencias de seguridad del paciente, cumplimiento de MDR/IVDR, y con los principios éticos. En Smart Cities, debe considerar la criticidad de infraestructuras, impacto en servicios esenciales, cumplimiento de la Directiva NIS2 y aceptabilidad social del riesgo.

6.7 Gestión del riesgo

Una vez clasificados en las diferentes categorías según el umbral definido (aceptables e inaceptables), se debe seleccionar la estrategia de gestión del riesgo que se pretende seguir para tratar los riesgos que la Organización no quiere asumir. En caso de que aplique, se definirán planes de tratamiento del riesgo.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

6.7.1 Selección de la estrategia de gestión del riesgo

Finalmente, se debe decidir la estrategia de tratamiento, es decir, cómo se abordan los riesgos identificados, especialmente aquellos que exceden el umbral de riesgo aceptable:

- ✓ **Evitar:** Eliminar el riesgo eliminando la causa o activo en cuestión. Conforme al marco HUDERIA, la evitación constituye la estrategia prioritaria cuando el riesgo afecte gravemente a derechos fundamentales, seguridad del paciente o infraestructuras críticas, y no existan medidas de mitigación efectivas o proporcionadas. En salud, puede implicar la retirada del sistema de IA, la sustitución por alternativas no automatizadas o la limitación del ámbito de uso. En Smart Cities, puede derivar en la no implementación de sistemas de vigilancia o control automatizado cuando el impacto en privacidad o libertades ciudadanas resulte desproporcionado.
- ✓ **Mitigar:** Reducir la probabilidad de que un riesgo ocurra y/o su impacto mediante la puesta en marcha de líneas de actividad o planes de mitigación. Generalmente, estas acciones deben tener un coste que debe ser inferior al beneficio obtenido (reducción del riesgo). En sistemas de IA en salud, las medidas de mitigación pueden incluir validación clínica reforzada, supervisión humana obligatoria, calibración de modelos o revisión de protocolos asistenciales. En Smart Cities, pueden incluir redundancia de infraestructuras, monitorización continua, segregación de redes o actualización de sistemas de control.
- ✓ **Transferir:** Transferir, total o parcialmente, la propiedad de los riesgos y sus consecuencias a una tercera parte. Esto podrá realizarse a través de un seguro, fianza, garantía, externalización, etc. A menudo conlleva el pago de una prima de riesgo. Se recomienda transferir la propiedad de un riesgo a una tercera parte cuando dicha parte presente mayor capacidad para absorberlo que la propia Organización. En entornos sanitarios y Smart Cities, la transferencia de riesgos debe formalizarse contractualmente, especificando responsabilidades, obligaciones de notificación de incidentes, cumplimiento normativo y derechos de auditoría. No obstante, la transferencia contractual no exime a la Organización de sus obligaciones regulatorias y de rendición de cuentas ante autoridades competentes y ciudadanía.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- ✓ **Aceptar:** En este caso ha de considerarse una reserva de tiempo, dinero y recursos para estos riesgos. La contingencia ha de definirse en base al posible impacto. El responsable debe chequear periódicamente que el riesgo no se desplaza hacia niveles no tolerables. La aceptación de riesgos solo debe aplicarse cuando el riesgo residual se encuentre por debajo del umbral de tolerancia definido, y nunca debe aceptarse un riesgo alto o muy alto que afecte a derechos fundamentales, seguridad del paciente o infraestructuras críticas sin aprobación expresa de la alta dirección y evaluación documentada de proporcionalidad y necesidad. La aceptación debe formalizarse mediante documento firmado por el responsable competente, indicando las razones, alternativas consideradas y medidas de monitorización continua.
- ✓ **Restaurar:** Establecer mecanismos y procedimientos para restablecer los derechos, servicios o condiciones afectadas cuando el riesgo se haya materializado. Conforme al marco HUDERIA, la restauración constituye una estrategia complementaria que busca revertir o corregir los efectos adversos producidos por el sistema de IA. En salud, puede incluir la corrección de diagnósticos erróneos, revisión de historiales clínicos afectados, notificación a pacientes y profesionales, y adopción de medidas correctivas en protocolos asistenciales. En Smart Cities, puede implicar la restauración de servicios interrumpidos, rectificación de decisiones automatizadas incorrectas, reparación de infraestructuras comprometidas o restablecimiento de condiciones de acceso equitativo a servicios públicos.
- ✓ **Compensar:** Proporcionar reparación o indemnización a las personas o colectivos afectados cuando los efectos adversos del sistema de IA no puedan ser completamente evitados, mitigados o restaurados. Conforme al marco HUDERIA, la compensación constituye la estrategia de último recurso, aplicable cuando el daño es irreversible o de difícil corrección. En salud, puede incluir indemnizaciones a pacientes por errores diagnósticos o terapéuticos, cobertura de tratamientos adicionales, o reconocimiento institucional del daño causado. En Smart Cities, puede implicar compensaciones económicas, medidas de apoyo a colectivos discriminados, acciones de reparación simbólica o implementación de garantías reforzadas para prevenir la recurrencia.

La selección de la estrategia adecuada dependerá de varios factores como la tolerancia de riesgo de la Organización, los recursos disponibles, el impacto potencial del riesgo, y el entorno operativo.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
Plan de Recuperación, Transformación y Resiliencia
incibe_ INSTITUTO NACIONAL DE CIBERSEGURIDAD
Junta de Andalucía
Agencia Digital de Andalucía

Por regla general, aquellos cuyo nivel de riesgo sea superior al riesgo objetivo propuesto por la Organización, requerirán del establecimiento de líneas de actividad orientadas a tratar dichos riesgos. Así mismo, el análisis se deberá llevar a cabo por cada uno de los riesgos para definir la decisión más apropiada.

En coherencia con HUDERIA, las estrategias de gestión de riesgos deben **priorizar la siguiente jerarquía de actuación**: evitar, mitigar, restaurar y, en último término, compensar. Este enfoque garantiza que, siempre que sea posible, los riesgos se eliminen en origen y que se preserve la protección de derechos fundamentales, la calidad democrática y la confianza ciudadana en entornos de salud y Smart Cities.

En sistemas de alto riesgo conforme al Reglamento de IA, o en dispositivos médicos con IA sujetos a MDR/IVDR, la estrategia de gestión debe alinearse con las obligaciones regulatorias de documentación, trazabilidad, vigilancia poscomercialización y notificación de incidentes adversos.

6.7.2 Definición del plan de tratamiento del riesgo

Una vez seleccionadas las estrategias de gestión del riesgo, el siguiente paso es definir un plan de tratamiento del riesgo, donde se identifique para cada acción:

- Situación actual o necesidad
- Principales riesgos
- Objetivo
- Fases y actividades
- Duración estimada de las actividades
- Controles asociados
- Responsables/Involucrados
- Criticidad (a nivel subjetivo: Alta/Media/Baja)
- Esfuerzo (a nivel subjetivo: Alto/Medio/Baja)
- Estimación económica

Todas las acciones integradas en el plan de tratamiento de riesgos deberán tener asociado un responsable para su ejecución.

De cara a obtener unos resultados adecuados durante el proceso de ejecución del plan de tratamiento del riesgo, se deben llevar a cabo revisiones periódicas del avance de las acciones proporcional al nivel de riesgo. Las revisiones deben documentarse formalmente, registrando el estado de ejecución, desviaciones detectadas, acciones correctivas adoptadas y decisiones de cierre o prórroga.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

7. DEFINICIONES

A continuación, se detallan los conceptos más relevantes vinculados con el análisis y gestión del riesgo de Inteligencia Artificial:

- ✓ **Análisis de riesgos:** Estudio de las consecuencias previsibles de un posible evento, considerando su impacto en una Organización y la probabilidad de que ocurra.
- ✓ **Amenaza:** Acciones que explotan una vulnerabilidad para comprometer el sistema, potencialmente causando efectos negativos en sus componentes. Para sistemas de IA se incluyen ataques específicos como la inyección de prompts, el envenenamiento de modelos o las alucinaciones. Adicionalmente, las amenazas pueden surgir de eventos físicos, como incendios o inundaciones, o de negligencias y decisiones inapropiadas dentro de la Organización, como la mala gestión de contraseñas o la falta de uso de cifrado. Las amenazas pueden ser tanto internas, originadas por personal o procesos internos de la Organización, como externas, provenientes de actores o circunstancias fuera de la misma.
- ✓ **Deep Learning (DeepL):** Subdisciplina de Machine Learning que emplea redes neuronales con múltiples capas para modelar datos complejos. Es especialmente eficaz en el reconocimiento de imágenes y voz, así como en el procesamiento del lenguaje natural.
- ✓ **Deriva del modelo (Model Drift):** Degradación progresiva del rendimiento de un modelo de IA debido a cambios en distribuciones de datos, patrones poblacionales o condiciones operativas. Requiere recalibración, reentrenamiento o revisión del modelo.
- ✓ **Evaluación de Impacto en Derechos Fundamentales (EIDF):** Proceso estructurado de identificación, análisis y valoración de los efectos de un sistema de IA sobre derechos humanos, democracia y Estado de derecho, conforme al marco HUDERIA. Determina escala, alcance, reversibilidad y probabilidad del impacto, definiendo medidas de evitación, mitigación o compensación.
- ✓ **Gestión del riesgo:** Actividades coordinadas para dirigir y controlar una Organización con respecto al riesgo.
- ✓ **IA Generativa:** Modelos de inteligencia artificial que pueden crear nuevo contenido, como texto, imágenes y música, a partir de los datos con los que han sido entrenados. Ejemplos: GPT para texto o GANs para imágenes.
- ✓ **Impacto:** Magnitud del daño que podría causar una amenaza cuando se materializa sobre un sistema.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

- ✓ **Medidas de control:** Conjunto de disposiciones encaminadas a proteger al sistema de los riesgos a los que estuviese sometido, con el fin de asegurar sus objetivos. Puede tratarse de medidas de prevención, de disuasión, de protección, de detección y reacción, o de recuperación.
- ✓ **Machine Learning (ML):** Rama de la IA que utiliza datos y algoritmos para aprender de forma gradual. Los algoritmos de ML pueden adaptarse y mejorar su rendimiento a medida que se exponen a más datos.
- ✓ **Probabilidad:** Tendencia a que una amenaza se produzca sobre un sistema.
- ✓ **Riesgo:** Depende de la probabilidad y el impacto de que una amenaza se materialice y la información, los datos personales, el acceso a cuentas bancarias, entre otras, queden expuestas o sean modificadas.
- ✓ **Riesgo inherente:** El riesgo inherente es el resultado del cálculo de la probabilidad de la materialización de una amenaza y el impacto producido sobre los sistemas, sin tener en cuenta medidas de control que puedan reducir el riesgo.
- ✓ **Riesgo residual:** Es el resultado de la identificación, evaluación e implementación de medidas de control sobre los diferentes sistemas, permitiendo así obtener una reducción del riesgo inherente.
- ✓ **Sistema de IA:** Sistema diseñado para operar con diversos grados de autonomía, capaz de adaptarse después de su implementación. Estos sistemas, ya sea de manera explícita o implícita, procesan entradas para generar salidas como predicciones, contenido, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Incluyen modelos de aprendizaje automático (Machine Learning), aprendizaje profundo (Deep Learning) e IA generativa.
- ✓ **Sesgo algorítmico:** Distorsión sistemática en los resultados de un modelo de IA derivada de datos no representativos, diseño inadecuado o factores contextuales. Puede generar discriminación injustificada contra colectivos vulnerables.
- ✓ **Supervisión humana:** Capacidad de las personas para supervisar, intervenir, corregir o detener un sistema de IA, especialmente cuando las decisiones automatizadas puedan afectar a la seguridad, derechos o bienestar de individuos. Requisito fundamental del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024 (Reglamento de IA) para sistemas de alto riesgo.
- ✓ **Trazabilidad:** Capacidad de rastrear y documentar el origen, transformación, procesamiento y uso de datos, modelos, decisiones y acciones a lo largo del ciclo de vida del sistema de IA.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

8. ANEXOS

Este apartado incluye información complementaría según el tema de la guía, que respalda y amplía el contenido principal del documento.

8.1 Anexo 1: Catálogo de amenazas – Smart Cities

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.SC.1	Seguridad de datos	Violación de privacidad o tratamiento ilícito de datos personales en sistemas inteligentes	Riesgo de exposición o tratamiento indebido de datos personales recopilados mediante sensores, cámaras, micrófonos o plataformas de IA integradas en servicios urbanos. Ejemplo: divulgación de imágenes o patrones de movilidad de personas. Puede implicar infracción del RGPD y pérdida de confianza ciudadana.
A.SC.2	Seguridad de datos	Acceso no autorizado a repositorios de datos urbanos	Acceso ilícito o manipulación de bases de datos que contienen información sobre infraestructuras críticas, tráfico, energía o servicios públicos. Un atacante podría alterar la información, comprometiendo la integridad de decisiones automatizadas.
A.SC.3	Seguridad de datos	Uso indebido o no autorizado de información o propiedad intelectual	Utilización de datos o modelos de IA con licencias restringidas o derechos de autor sin autorización. La reutilización no controlada de información puede acarrear sanciones legales y afectar la reputación institucional.
A.SC.4	Seguridad de datos	Eliminación o conservación no conforme de modelos o conjuntos de datos	Conservación de modelos o datos de entrenamiento obsoletos que contienen información sensible. La falta de procedimientos de eliminación segura puede derivar en brechas de seguridad o incumplimiento normativo.
A.SC.5	Seguridad de datos	Ausencia de cifrado o medidas de protección de la información	Transmisión o almacenamiento de datos sin cifrado ni autenticación. Ejemplo: comunicación entre sensores y servidores municipales vulnerable a interceptación o manipulación.
A.SC.6	Seguridad de red	Intercepción o análisis de tráfico	Captura o análisis del tráfico de red entre dispositivos IoT, sistemas de control y plataformas de IA, lo que permite inferir información sensible sobre la operación de servicios urbanos.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.SC.7	Seguridad de red	Vulnerabilidad o compromiso de interfaces o APIs abiertas	Explotación de vulnerabilidades en interfaces de programación o servicios abiertos utilizados por sistemas de IA municipales. Puede permitir manipulación de datos o ejecución remota no autorizada.
A.SC.8	Intrusión	Acceso no autorizado a sistemas de control o supervisión urbana	Explotación de vulnerabilidades o robo de credenciales para obtener acceso a plataformas que gobiernan infraestructuras críticas (energía, transporte, agua). Riesgo de alteración de parámetros operativos.
A.SC.9	Intrusión	Suplantación de identidad o elevación indebida de privilegios	Actores maliciosos adoptan la identidad de usuarios o administradores autorizados, obteniendo acceso a sistemas de IA o paneles de control urbano.
A.SC.10	Intrusión	Creación o uso de servicios o sistemas de IA falsificados	Creación de aplicaciones o portales falsos que imitan servicios oficiales basados en IA (phishing institucional). Riesgo de fraude o filtración de datos.
A.SC.11	Intrusión	Configuración incorrecta o no conforme	Establecimiento inadecuado de permisos o políticas de acceso en entornos de IA, lo que puede exponer datos o funcionalidades críticas.
A.SC.12	Vulnerabilidades de IA	Generación de código inseguro por IA	Código producido mediante herramientas de IA sin revisión técnica ni validación de seguridad, susceptible de contener vulnerabilidades o errores.
A.SC.13	Vulnerabilidades de IA	Generación de resultados falsos o alucinaciones del modelo	Producción de información incorrecta o no fundamentada por modelos de IA que respaldan decisiones en movilidad, energía o seguridad pública. Puede inducir acciones operativas erróneas.
A.SC.14	Vulnerabilidades de IA	Inyección de prompts o manipulación de entrada	Inserción de instrucciones adversarias en asistentes o interfaces de IA que provocan respuestas no previstas o exposición de información sensible.
A.SC.15	Vulnerabilidades de IA	Manipulación o envenenamiento de datos de entrenamiento	Introducción deliberada de datos falsos en conjuntos de entrenamiento que distorsionan los modelos de predicción o clasificación utilizados en servicios urbanos.
A.SC.16	Vulnerabilidades de IA	Robo o extracción de modelos	Obtención no autorizada de parámetros, arquitecturas o salidas de modelos predictivos mediante técnicas de ingeniería inversa o consultas reiteradas.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.SC.17	Vulnerabilidades de IA	Evasión del modelo	Uso de entradas diseñadas para engañar a un modelo de IA, provocando errores en la clasificación o reconocimiento (por ejemplo, alteración de señales de tráfico detectadas por sensores de inteligencia artificial).
A.SC.18	Vulnerabilidades de IA	Sesgo algorítmico y discriminación	Modelos entrenados con datos no representativos pueden producir decisiones inequitativas entre grupos poblacionales o zonas geográficas, afectando la igualdad en la prestación de servicios en Smart Cities.
A.SC.19	Vulnerabilidades de IA	Deriva del modelo o degradación del rendimiento	Degradación progresiva del rendimiento de la modelo debida a cambios en los patrones de datos o condiciones operativas, afectando la fiabilidad del sistema.
A.SC.20	Gobernanza y transparencia	Falta de explicabilidad o transparencia del sistema	Dificultad para comprender o justificar las decisiones tomadas por sistemas de IA empleados en la gestión pública, afectando la rendición de cuentas y la confianza ciudadana.
A.SC.21	Contenido dañino	Generación o difusión automatizada de desinformación	Uso indebido de herramientas de IA para generar o difundir contenidos falsos o sesgados en plataformas institucionales, con impacto en la confianza social.
A.SC.22	Contenido dañino	Software malicioso (malware)	Inserción o ejecución de código malicioso que compromete el funcionamiento de los sistemas de IA o dispositivos IoT conectados.
A.SC.23	Disponibilidad	Ataques de denegación de servicio (DDoS)	Saturación deliberada de recursos o servicios IA, provocando indisponibilidad en infraestructuras críticas (semáforos, gestión de tráfico, energía).
A.SC.24	Disponibilidad	Interrupción de operaciones automatizadas	Dependencia excesiva de soluciones de IA sin planes de contingencia manual, lo que puede detener servicios esenciales en caso de fallo.
A.SC.25	Disponibilidad	Deficiencia en respaldo o recuperación	Ausencia o ineficacia de mecanismos de copia y restauración de datos, configuraciones o modelos, dificultando la recuperación tras incidentes.
A.SC.26	Personal	Errores de operadores o administradores	Fallos no intencionados del personal con privilegios sobre sistemas de IA, ocasionando interrupciones o exposición de datos.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.SC.27	Personal	Uso malicioso o sabotaje interno	Empleados o contratistas con acceso autorizado que emplean deliberadamente sistemas de IA para causar daño, filtrar información o alterar resultados.
A.SC.28	Cadena de suministro	Dependencia de software o componentes inseguros	Uso de librerías, modelos o servicios de terceros sin controles de seguridad adecuados, aumentando la superficie de ataque.
A.SC.29	Cadena de suministro	Incidentes en proveedores de servicios IA	Fallos, fugas de datos o indisponibilidad en proveedores externos que impactan la continuidad de los servicios municipales.
A.SC.30	Cumplimiento y auditoría	Supervisión o monitorización insuficiente	Falta de vigilancia continua del comportamiento de los sistemas de IA integrados, lo que puede permitir fallos no detectados o fugas de información.
A.SC.31	Cumplimiento y auditoría	Falta de transparencia institucional	No comunicación clara al ciudadano del uso de IA en los servicios públicos, vulnerando principios éticos y normativos.
A.SC.32	Cumplimiento y auditoría	Incumplimiento normativo o contractual	Inobservancia de marcos legales o contractuales en el uso de IA (por ejemplo, RGPD, ISO 42001), lo que puede conllevar sanciones o pérdida de certificación

8.2 Anexo II: Catálogo de amenazas – Salud

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.H.1	Seguridad de datos	Violación de privacidad de datos clínicos	Exposición o uso no autorizado de información médica (imágenes, historiales, resultados) procesada por sistemas de IA. Supone infracción del RGPD, del principio de confidencialidad profesional y riesgo de pérdida de confianza institucional.
A.H.2	Seguridad de datos	Acceso no autorizado a información sanitaria	Acceso ilícito a repositorios o sistemas que contienen datos clínicos o parámetros de diagnóstico generados por IA. Posible alteración de resultados o robo de información sensible.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.H.3	Seguridad de datos	Reutilización indebida de datos médicos	Empleo de información sanitaria para el entrenamiento o validación de modelos IA sin consentimiento o sin propósito clínico legítimo, infringiendo el RGPD y MDR/IVDR.
A.H.4	Seguridad de datos	Eliminación inadecuada de datos o modelos obsoletos	Conservación de conjuntos de datos clínicos o versiones de modelos que deberían ser eliminados, incumpliendo políticas de retención o el principio de minimización del tratamiento de datos.
A.H.5	Seguridad de datos	Falta de cifrado en datos clínicos en tránsito o reposo	Ausencia de cifrado o autenticación en sistemas que transmiten información médica (por ejemplo, entre dispositivos médicos e infraestructura hospitalaria). Riesgo de interceptación o manipulación.
A.H.6	Seguridad de red	Intercepción o manipulación de comunicaciones médicas	Captura de tráfico de red entre sistemas clínicos o plataformas IA que puede alterar resultados o exponer datos de pacientes.
A.H.7	Intrusión	Acceso no autorizado a sistemas clínicos	Explotación de vulnerabilidades o robo de credenciales que permite acceder a sistemas IA implicados en diagnóstico o soporte clínico.
A.H.8	Intrusión	Suplantación de identidad profesional	Uso fraudulento de credenciales de personal sanitario o técnico para alterar resultados, decisiones o auditorías generadas por IA.
A.H.9	Intrusión	Software o dispositivos de IA falsificados	Distribución de aplicaciones o dispositivos falsificados que imitan software sanitario legítimo. Riesgo de fraude y daño clínico.
A.H.10	Vulnerabilidades de IA	Envenenamiento de datos de entrenamiento clínico	Inserción deliberada de datos alterados o falsos en los conjuntos de entrenamiento que generan sesgos o diagnósticos erróneos.
A.H.11	Vulnerabilidades de IA	Sesgo algorítmico en diagnóstico o tratamiento	Modelos entrenados con datos no representativos que producen resultados injustos o inexactos según edad, género, etnia o patología, afectando la equidad clínica.
A.H.12	Vulnerabilidades de IA	Falta de explicabilidad o interpretabilidad del modelo	Incomprensibilidad de las decisiones del sistema de IA, lo que dificulta su validación, auditoría o aceptación por autoridades regulatorias.

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.H.13	Vulnerabilidades de IA	Deriva del modelo y pérdida de precisión	Degradación del rendimiento clínico del modelo debido a cambios en datos poblacionales o contextos operativos. Puede generar diagnósticos incorrectos o pérdida de eficacia terapéutica.
A.H.14	Vulnerabilidades de IA	Evasión o manipulación del modelo	Ataques que buscan provocar resultados erróneos introduciendo entradas diseñadas (p. ej., imágenes médicas modificadas).
A.H.15	Validación y seguridad del producto	Validación clínica insuficiente	Implementación de sistemas de IA sin validación técnica o clínica adecuada conforme a MDR/IVDR y ISO 14971, lo que puede comprometer la seguridad del paciente.
A.H.16	Validación y seguridad del producto	Modificaciones o reentrenamientos no validados	Modificación del modelo o sus parámetros sin validación ni documentación posterior, contraviniendo IEC 81001-5-1 y los requisitos de control de cambios.
A.H.17	Validación y seguridad del producto	Falta de trazabilidad de versiones de modelo	Ausencia de registros sobre versiones, datasets o algoritmos empleados, impidiendo auditorías regulatorias o la reproducción de resultados.
A.H.18	Seguridad del paciente	Recomendaciones diagnósticas o terapéuticas incorrectas	Emisión de resultados erróneos o inapropiados por defectos de diseño, entrenamiento o uso del sistema IA. Riesgo directo para la seguridad del paciente.
A.H.19	Seguridad del paciente	Integración insegura con dispositivos médicos	Falta de validación en la interoperabilidad entre IA y equipos médicos conectados, pudiendo alterar mediciones o alarmas críticas.
A.H.20	Seguridad del paciente	Ausencia de supervisión humana adecuada	Decisiones críticas automatizadas sin revisión o aprobación por profesionales cualificados, incumpliendo el principio de supervisión humana del AI Act.
A.H.21	Seguridad del paciente	Uso no conforme con la indicación o finalidad prevista (‘off-label’)	Empleo de sistemas IA en contextos clínicos distintos a los validados o certificados. Riesgo de infracción regulatoria y daño clínico.
A.H.22	Seguridad del paciente	Actualizaciones automáticas no validadas	Implementación de parches o versiones de IA sin validación previa que puedan modificar el comportamiento clínico del sistema.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Grupo de la amenaza	Nombre de la amenaza	Descripción de la amenaza
A.H.23	Seguridad del paciente	Integración insegura o no conforme con sistemas de historia clínica electrónica (HCE	Conexión insegura o mal configurada con sistemas HCE (HL7, FHIR), generando exposición de datos o resultados incongruentes.
A.H.24	Cadena de suministro	Dependencia de software o servicios externos no certificados	Utilización de componentes de terceros sin evaluación de seguridad o conformidad regulatoria, incrementando la superficie de ataque.
A.H.25	Cadena de suministro	Incidentes en proveedores tecnológicos o de nube	Brechas, fallos o indisponibilidad en servicios de IA externos que afectan operaciones hospitalarias críticas.
A.H.26	Cumplimiento y auditoría	Ausencia de vigilancia poscomercialización	Falta de monitorización continua de desempeño, errores o incidentes adversos tras el despliegue clínico, contraviniendo MDR/IVDR y AI Act.
A.H.27	Cumplimiento y auditoría	Deficiencias en registro y trazabilidad	Ausencia de documentación completa del ciclo de vida del sistema IA, incluyendo datos, modelos, validaciones y auditorías.
A.H.28	Cumplimiento y auditoría	No conformidad con requisitos legales o regulatorios aplicables	Falta de adhesión a MDR, IVDR, ISO 42001, IEC 81001-5-1 o RIA. Puede resultar en sanciones, retirada de marcado CE o suspensión del sistema.
A.H.29	Ética y gobernanza	Ausencia de consentimiento informado específico para el uso de sistemas de IA	Empleo de sistemas IA sin comunicar su implicación al paciente o sin obtener su consentimiento explícito, vulnerando derechos fundamentales.
A.H.30	Ética y gobernanza	Pérdida de autonomía humana o ambigüedad en la responsabilidad clínica	Delegación excesiva de decisiones en IA sin mecanismos claros de rendición de cuentas ni salvaguardas humanas.
A.H.31	Ética y gobernanza	Uso no conforme o fuera del contexto clínico previsto	Aplicación del sistema fuera del entorno o finalidad clínica prevista por su certificación. Riesgo ético, legal y operativo.

8.3 Anexo III: Catálogo de controles – Smart Cities y Salud

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
Plan de Recuperación, Transformación y Resiliencia
incibe_ INSTITUTO NACIONAL DE CIBERSEGURIDAD
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.01	Gobernanza y gestión del riesgo	Marco organizativo para sistemas de inteligencia artificial	Establece un sistema de gestión formal que abarque el ciclo de vida completo de los sistemas de inteligencia artificial, con políticas, procedimientos, principios y responsabilidades aprobados por la alta dirección. Debe estar alineado con normas ISO/IEC 42001 y 31000, garantizando la integración con los marcos corporativos de seguridad, calidad, cumplimiento y ética.
C.02	Gobernanza y gestión del riesgo	Roles, responsabilidades y segregación de funciones	Define y documenta funciones específicas para todas las partes implicadas (técnicas, éticas, clínicas, de ciberseguridad, de protección de datos y de gobernanza). Asegura la segregación de funciones críticas —por ejemplo, desarrollo, validación, operación y supervisión— y establece mecanismos de rendición de cuentas y sustitución en caso de conflicto de intereses.
C.03	Gobernanza y gestión del riesgo	Proceso institucional de gestión del riesgo de inteligencia artificial	Integra la gestión del riesgo de IA en el sistema global de gestión de riesgos de la organización conforme a ISO/IEC 23894 e ISO 31000. Incluye procesos de identificación, análisis, evaluación, tratamiento, aceptación y seguimiento de riesgos específicos de IA, con criterios de probabilidad, impacto y evidencia documentada.
C.04	Gobernanza y gestión del riesgo	Comité de supervisión ética y de conformidad en inteligencia artificial	Crea un comité multidisciplinar con representación técnica, ética, clínica, jurídica y social, responsable de supervisar el cumplimiento normativo, la equidad, la transparencia y la seguridad de los sistemas de IA. Tiene autoridad para revisar decisiones, solicitar auditorías o suspender sistemas en caso de riesgo grave o incumplimiento.
C.05	Gobernanza y gestión del riesgo	Evaluación de impacto en derechos fundamentales	Obliga a realizar evaluaciones previas sobre los posibles efectos de los sistemas de IA en los derechos fundamentales, la igualdad, la no discriminación, la privacidad y otros valores constitucionales. Incluye identificación de riesgos, medidas compensatorias, consulta a grupos afectados y revisión continua durante la operación.

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.06	Gobernanza y gestión del riesgo	Registro institucional de sistemas de inteligencia artificial	Mantiene un inventario institucional de todos los sistemas de IA, clasificándolos por finalidad, nivel de riesgo, responsable, versión, contexto de uso y controles aplicados. Este registro constituye la base documental de auditorías y revisiones regulatorias.
C.07	Gobernanza y gestión del riesgo	Política de uso previsto y límites de operación	Define los supuestos de diseño, contexto y límites de uso de cada sistema de IA. Prohíbe explícitamente su aplicación fuera de dichos límites o en contextos no validados, y establece procedimientos de revisión ante usos indebidos o desviaciones detectadas.
C.08	Gobernanza y gestión del riesgo	Gestión del cambio en sistemas de inteligencia artificial	Implementa un proceso formal de evaluación, aprobación y documentación de cambios en modelos, datos o configuraciones. Incluye pruebas de impacto, validación previa, reversión segura y comunicación a las partes interesadas.
C.09	Gobernanza y gestión del riesgo	Plan de retirada, sustitución y fin de vida	Establece procedimientos documentados para la retirada temporal o definitiva de sistemas de IA, incluyendo preservación de evidencias, sustitución de versiones, eliminación o anonimización de datos y continuidad del servicio durante la transición.
C.10	Gobernanza y gestión del riesgo	Transparencia institucional y comunicación responsable	Garantiza la comunicación clara y accesible sobre el uso de sistemas de IA, sus finalidades, principales características y limitaciones. Asegura la existencia de canales de información, consulta y reclamación para usuarios, pacientes y ciudadanía.
C.11	Gestión de datos	Gobierno y trazabilidad de datos	Documenta el origen, licitud, condiciones de uso, formato y calidad de todos los datos empleados en el ciclo de vida del sistema. Garantiza trazabilidad integral mediante registros verificables de adquisición, transformación y utilización.
C.12	Gestión de datos	Control de calidad de datos	Implementa controles de completitud, coherencia, exactitud y vigencia de los datos. En salud, incluye validación de señales clínicas, imágenes o textos; en entornos urbanos, calibración de sensores y verificación de consistencia temporal.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
Plan de Recuperación, Transformación y Resiliencia
incibe_ INSTITUTO NACIONAL DE CIBERSEGURIDAD
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.13	Gestión de datos	Minimización y limitación de finalidad	Asegura que el tratamiento de datos se limite estrictamente a la finalidad declarada y que no se realicen usos secundarios no autorizados. Promueve la aplicación de técnicas de anonimización y seudonimización.
C.14	Gestión de datos	Clasificación y protección de información	Clasifica datos y artefactos según su nivel de sensibilidad y criticidad, aplicando controles de confidencialidad, integridad y disponibilidad proporcionales. Establece esquemas de cifrado, control de acceso y registro de actividad.
C.15	Gestión de datos	Catálogo de activos y metadatos normalizados	Mantiene un catálogo estructurado de todos los conjuntos de datos, con metadatos estandarizados que describen procedencia, contexto, calidad, limitaciones y restricciones de uso, asegurando la reproducibilidad y auditabilidad.
C.16	Gestión de datos	Control de versiones y huellas de integridad de datos	Versiona conjuntos de datos y conserva huellas criptográficas y descripciones de cambios. Facilita la comparación de resultados y garantiza integridad y trazabilidad.
C.17	Gestión de datos	Etiquetado y anotación con garantía de calidad	Define criterios, procesos y controles para el etiquetado de datos, con formación de anotadores, validación cruzada, revisión independiente y corrección de errores sistemáticos.
C.18	Diseño y desarrollo	Arquitectura segura y defensa en profundidad	Implementa principios de ingeniería segura, segmentación de redes, protección de componentes críticos, gestión de secretos y reducción de la superficie de exposición.
C.19	Diseño y desarrollo	Entornos segregados y datos no identificables fuera de producción	Mantiene separación entre entornos de desarrollo, prueba y producción, utilizando únicamente datos anonimizados o sintéticos fuera de producción.
C.20	Diseño y desarrollo	Seguridad del código y de las dependencias	Realiza revisiones de código, análisis estático y dinámico, control de dependencias externas y verificación de integridad antes del despliegue en producción.
C.21	Diseño y desarrollo	Gestión de configuración del sistema y del modelo	Establece un proceso formal de gestión de configuración que incluya versiones, parámetros, canalizaciones de entrenamiento e inferencia, y documentación justificativa.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.22	Validación y verificación	Validación técnica previa al despliegue	Asegura que el sistema de IA cumple los requisitos de rendimiento, fiabilidad, seguridad y estabilidad antes de su uso operativo, mediante pruebas reproducibles y documentadas.
C.23	Validación y verificación	Validación contextual (clínica u operativa)	Evalúa la adecuación del sistema a su contexto de uso: seguridad y utilidad clínica, o continuidad y seguridad del servicio público. Garantiza documentación técnica y validación independiente.
C.24	Validación y verificación	Evaluación de sesgo y equidad	Realiza análisis de equidad y sesgo en resultados y datos de entrenamiento, identificando disparidades injustificadas y aplicando medidas de mitigación y seguimiento continuo.
C.25	Validación y verificación	Explicabilidad y justificación de decisiones	Garantiza que las decisiones automatizadas sean comprensibles, justificables y trazables. Incluye métodos de interpretabilidad y documentación de supuestos, limitaciones y factores determinantes.
C.26	Validación y verificación	Pruebas de seguridad funcional y modos degradados	Verifica que el sistema responda de manera segura ante fallos, datos erróneos o picos de carga. Define mecanismos de paro seguro y funcionamiento degradado documentado.
C.27	Robustez y seguridad de IA	Protección frente a envenenamiento de datos	Implementa controles para detectar y mitigar intentos de manipulación de datos de entrenamiento mediante validaciones cruzadas, auditorías y revisión humana.
C.28	Robustez y seguridad de IA	Defensa ante evasión del modelo	Evalúa la capacidad del sistema para resistir ataques adversarios o entradas maliciosas que intenten alterar su salida. Incluye detección, alerta y respuesta automatizada.
C.29	Robustez y seguridad de IA	Prevención de extracción o ingeniería inversa del modelo	Aplica controles técnicos y de monitorización para reducir el riesgo de extracción de parámetros o reconstrucción del modelo a través de consultas reiteradas o patrones de explotación.
C.30	Robustez y seguridad de IA	Detección de entradas fuera de distribución	Identifica entradas anómalas o fuera del dominio del entrenamiento y activa medidas automáticas o revisión humana antes de emitir una decisión.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.31	Robustez y seguridad de IA	Monitorización de deriva y recalibración	Supervisa cambios en las distribuciones de datos, rendimiento o comportamiento del modelo durante la operación. Establece umbrales de alerta y procedimientos documentados de recalibración, reentrenamiento o retirada controlada cuando se detecte pérdida de rendimiento o desviación significativa.
C.32	Robustez y seguridad de IA	Combinación y comparación de modelos	Implementa estrategias de ensamblado, cotejo o validación cruzada entre modelos para mejorar la fiabilidad, robustez y detección de anomalías en los resultados. Documenta criterios de consenso, desempate y condiciones de selección de modelos activos.
C.33	Operación y supervisión	Supervisión humana efectiva en producción	Garantiza la supervisión humana efectiva sobre las decisiones o recomendaciones del sistema de IA, en especial aquellas con impacto clínico, ético o de seguridad pública. Establece puntos de control, mecanismos de revisión y capacidad de intervención o detención inmediata.
C.34	Operación y supervisión	Control de rendimiento, capacidad y latencia	Supervisa el comportamiento del sistema en tiempo real (rendimiento, latencia, uso de recursos y colas de inferencia). Define indicadores, umbrales y medidas correctivas, documentando resultados de los ajustes realizados.
C.35	Operación y supervisión	Filtrado y saneamiento de salidas	Implementa controles automáticos y manuales para evitar la exposición de datos personales o sensibles en las salidas del sistema. Aplica técnicas de desidentificación, revisión semántica y filtrado contextual según la finalidad.
C.36	Operación y supervisión	Paro seguro y conmutación a modos alternativos	Define mecanismos técnicos y procedimentales que permitan detener o conmutar el sistema a un modo alternativo seguro en caso de fallo, degradación o detección de comportamiento anómalo. Documenta responsabilidades, tiempos máximos de respuesta y protocolos de recuperación.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.37	Operación y supervisión	Registro integral y trazabilidad del ciclo de vida	Mantiene registros completos, verificables y sellados temporalmente de todas las fases del ciclo de vida del sistema (entrenamiento, validación, despliegue, inferencias y revisiones). Asegura trazabilidad de decisiones y responsables.
C.38	Operación y supervisión	Integridad y protección de registros	Protege los registros y evidencias mediante mecanismos de control de acceso, cifrado, almacenamiento inmutable o huellas criptográficas. Asegura la detección de alteraciones no autorizadas y la conservación de copias verificables.
C.39	Operación y supervisión	Sincronización temporal consistente	Asegura la sincronización horaria precisa entre componentes, servidores y dispositivos asociados al sistema, garantizando la correlación coherente de eventos, auditorías y análisis forense.
C.40	Privacidad y protección de datos	Base jurídica y finalidad determinadas	Asegura que todo tratamiento de datos personales se base en una causa legítima documentada y en una finalidad explícita, compatible y limitada. Incluye revisión de bases legales, consentimiento informado y registro de evaluaciones.
C.41	Privacidad y protección de datos	Evaluación de impacto en protección de datos	Realiza evaluaciones de impacto (EIPD) conforme a las exigencias normativas cuando el tratamiento pueda implicar alto riesgo. Identifica medidas de mitigación, responsables y procedimientos de revisión periódica.
C.42	Privacidad y protección de datos	Control de acceso a datos personales y sensibles	Aplica controles de acceso basados en privilegios mínimos, autenticación fuerte y segregación de funciones. Registra y revisa accesos a datos personales, clínicos u operativos sensibles.
C.43	Privacidad y protección de datos	Prevención de fugas de información	Implementa medidas preventivas, de detección y reactivas frente a exfiltración o divulgación no autorizada de datos en almacenamiento, transmisión o salidas del sistema. Incluye análisis de flujo, registro y bloqueo automatizado.
C.44	Privacidad y protección de datos	Conservación y supresión de datos	Define periodos de retención, políticas de supresión y anonimización verificable de datos y modelos al concluir su finalidad o plazo establecido. Registra las eliminaciones efectuadas y las evidencias de verificación.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.45	Privacidad y protección de datos	Evaluación y mitigación del riesgo de reidentificación	Evalúa sistemáticamente el riesgo de reidentificación de personas a partir de datos seudonimizados, anonimizados o agregados, conforme a metodologías reconocidas. Aplica técnicas de mitigación (generalización, agregación, perturbación o reducción de precisión) y conserva evidencias documentadas del nivel de riesgo residual aceptable, en línea con las directrices del Reglamento General de Protección de Datos.
C.46	Privacidad y protección de datos	Derechos de protección de datos de las personas interesadas	Establece procedimientos documentados para atender solicitudes de derechos de acceso, rectificación, supresión, oposición y limitación del tratamiento. Define plazos, responsables y registro de decisiones adoptadas.
C.47	Gestión de incidentes	Organización y preparación para incidentes	Define una estructura organizativa y funcional para la gestión de incidentes relacionados con la seguridad, la privacidad o la exactitud del sistema de IA. Incluye roles, herramientas, planes de respuesta y formación del personal.
C.48	Gestión de incidentes	Detección, correlación y alerta	Implanta mecanismos de detección temprana y correlación de eventos relevantes (logs, métricas, alertas del modelo, telemetría). Establece umbrales de alerta y procedimientos de escalado proporcionales al riesgo.
C.49	Gestión de incidentes	Clasificación, priorización y coordinación	Clasifica los incidentes en función de su impacto en seguridad, privacidad, disponibilidad o precisión. Define criterios de severidad, responsables de coordinación y protocolos de comunicación interna y externa.
C.50	Gestión de incidentes	Respuesta, recuperación y comunicación	Ejecuta acciones de contención, erradicación y recuperación del sistema ante incidentes. Gestiona la comunicación con partes interesadas, autoridades y usuarios, documentando causas, impacto y medidas correctivas.
C.51	Gestión de incidentes	Notificación a autoridades competentes	Establece los procedimientos para notificar incidentes graves a autoridades reguladoras o competentes, conforme a los plazos, formatos y requisitos normativos aplicables. Mantiene registro verificable de las comunicaciones realizadas.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.52	Gestión de incidentes	Lecciones aprendidas y actualización del riesgo	Incorpora los aprendizajes derivados de incidentes o cuasi-incidentes en el registro institucional de riesgos y en la revisión de políticas y controles. Evalúa la eficacia de las medidas implantadas y promueve la mejora continua.
C.53	Continuidad y resiliencia	Diseño para alta disponibilidad	Diseña arquitecturas redundantes y tolerantes a fallos proporcionales a la criticidad del sistema. Garantiza disponibilidad continua en sistemas esenciales (p. ej., clínicos o de infraestructuras urbanas).
C.54	Continuidad y resiliencia	Copias de seguridad y recuperación verificadas	Mantiene copias actualizadas y cifradas de datos, configuraciones, modelos y registros. Realiza ejercicios periódicos de restauración y documenta resultados, tiempos de recuperación y evidencias de verificación.
C.55	Continuidad y resiliencia	Plan de continuidad y recuperación ante desastres	Define escenarios de contingencia (pérdida de servicios, indisponibilidad de proveedores, degradación del modelo) y procedimientos de recuperación escalonada. Incluye designación de responsables y comunicación a partes interesadas.
C.56	Continuidad y resiliencia	Pruebas de resiliencia operativa	Realiza ejercicios de carga, fallo de componentes, interrupciones de red y simulaciones adversarias para validar la capacidad de respuesta y recuperación del sistema. Actualiza planes y capacidades con base en resultados.
C.57	Cadena de suministro y terceros	Evaluación de seguridad y madurez de proveedores	Evalúa los controles de seguridad, continuidad y madurez de los proveedores que aportan datos, modelos, servicios o componentes críticos. Mantiene evidencias documentadas de las evaluaciones y planes de mejora.
C.58	Cadena de suministro y terceros	Contratos con obligaciones de seguridad y continuidad	Incorpora en los contratos cláusulas específicas de confidencialidad, seguridad, notificación de incidentes, gestión de cambios, niveles de servicio, continuidad y derechos de auditoría técnica.
C.59	Cadena de suministro y terceros	Gestión de cambios del proveedor y subcontratistas	Exige notificación y aprobación previa de cambios significativos por parte de proveedores o subcontratistas que puedan afectar a la seguridad, privacidad o disponibilidad del sistema.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA
Plan de Recuperación, Transformación y Resiliencia
incibe_ INSTITUTO NACIONAL DE CIBERSEGURIDAD
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.60	Cadena de suministro y terceros	Inventario de materiales de software	Mantiene un inventario actualizado de todos los componentes de software utilizados (bibliotecas, frameworks, dependencias), incluyendo versiones, licencias y vulnerabilidades conocidas.
C.61	Cadena de suministro y terceros	Inventario de materiales del modelo	Documenta de forma estructurada los componentes del modelo (arquitectura, parámetros, datasets asociados, métricas y versiones). Facilita auditorías, verificaciones y trazabilidad completa.
C.62	Dispositivos y entornos físicos/OT	Arranque confiable y verificación de software	Garantiza que los dispositivos ejecuten únicamente software firmado y verificado criptográficamente. Implementa arranque seguro y detección de alteraciones o manipulaciones.
C.63	Dispositivos y entornos físicos/OT	Endurecimiento y control físico de equipos	Aplica medidas de seguridad física y lógica en los equipos: desactivación de servicios innecesarios, protección de gabinetes, inventario, control de acceso y registro de visitas a emplazamientos.
C.64	Dispositivos y entornos físicos/OT	Actualización segura en campo	Implementa canales autenticados, cifrados y verificados para la actualización remota de software o modelos. Controla versiones, mantiene integridad y prevé mecanismos de reversión segura.
C.65	Interoperabilidad e integración	Integración segura con sistemas existentes	Garantiza autenticación, autorización e integridad de los flujos de intercambio con sistemas externos. Incluye pruebas de compatibilidad, gestión de dependencias y validación de interfaces.
C.66	Interoperabilidad e integración	Gestión de versiones y cambios en interfaces	Controla versiones de interfaces y esquemas de datos. Realiza pruebas de regresión, documentación técnica y comunicación temprana de cambios a las partes interesadas.
C.67	Transparencia y participación	Información a usuarios, pacientes y ciudadanía	Proporciona información clara, comprensible y accesible sobre el uso de IA, su finalidad, limitaciones, grado de autonomía y medidas de supervisión humana. Establece canales de consulta y queja accesibles.
C.68	Transparencia y participación	Gestión de contenidos generados o manipulados por IA	Identifica, etiqueta y registra contenidos sintéticos o modificados por IA. Controla su difusión para evitar desinformación o confusión en contextos institucionales, clínicos o públicos.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.69	Auditoría y mejora continua	Auditoría técnica y de conformidad	Programa auditorías internas y externas periódicas para evaluar el diseño, datos, modelos y operación de los sistemas de IA. Documenta hallazgos, recomendaciones y acciones de mejora.
C.70	Auditoría y mejora continua	Documentación integral del ciclo de vida	Mantiene documentación completa y actualizada sobre el propósito, contexto, riesgos, datos, modelos, validaciones, operación y cambios del sistema, disponible para auditoría o revisión externa.
C.71	Auditoría y mejora continua	Revisión por la dirección y mejora del programa	La alta dirección revisa periódicamente el desempeño del marco de control de IA, los resultados de auditorías, incidentes y cambios regulatorios, estableciendo acciones de mejora y reasignación de recursos.
C.72	Formación y competencia	Competencia técnica y alfabetización en IA	Garantiza que el personal implicado posee la competencia técnica, ética y operativa necesaria en IA, sus riesgos y limitaciones. Mantiene registros de formación y evaluación periódica de capacidades.
C.73	Formación y competencia	Concienciación sobre amenazas específicas de IA	Desarrolla programas de sensibilización sobre amenazas relevantes (inyección de instrucciones, manipulación de datos, sesgos, ataques adversariales), buenas prácticas de uso y reporte de incidentes.
C.74	Uso previsto y seguridad del paciente/servicio	Control de uso conforme al contexto	Supervisa que el sistema de IA se utilice exclusivamente en los contextos, condiciones y propósitos aprobados. Detecta desviaciones operativas y establece mecanismos de corrección inmediata.
C.75	Uso previsto y seguridad del paciente/servicio	Integración segura con equipos y plataformas críticas	Verifica la interoperabilidad segura entre el sistema de IA y equipos o plataformas críticas (p. ej., dispositivos médicos, sistemas de tráfico o energía), previniendo efectos colaterales en seguridad o continuidad.
C.76	Uso previsto y seguridad del paciente/servicio	Revisión humana de resultados de alto impacto	Exige revisión y validación humana previa a la ejecución de decisiones automatizadas con alto impacto en seguridad, salud o derechos. Documenta la intervención y los criterios de decisión.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

ID	Dominio	Nombre del control	Descripción del control
C.77	Seguridad técnica	Gestión de vulnerabilidades y ciclo de parches	Identifica, evalúa y prioriza vulnerabilidades técnicas en componentes, librerías o configuraciones. Aplica parches y mitigaciones con control de pruebas, impacto y trazabilidad.
C.78	Seguridad técnica	Criptografía y gestión de claves	Implementa políticas y procedimientos de gestión de claves y cifrado en tránsito y reposo. Asegura generación segura, rotación periódica, almacenamiento protegido y registro de uso.
C.79	Seguridad técnica	Pruebas de seguridad y ejercicios controlados	Realiza pruebas de penetración, auditorías técnicas y ejercicios controlados sobre las superficies de exposición del sistema. Registra resultados, prioriza corrección y verifica cierre.
C.80	Supervisión del desempeño	Indicadores, umbrales y acuerdos de nivel de servicio	Define indicadores clave de desempeño (KPI) y umbrales operativos. Establece acuerdos de nivel de servicio y acciones automáticas o manuales ante desviaciones detectadas.
C.81	Supervisión del desempeño	Control de versiones de inferencia y liberaciones controladas	Gestiona la publicación gradual de nuevas versiones de modelos con validación comparativa antes de su adopción completa. Documenta riesgos, métricas y aprobación.
C.82	Supervisión del desempeño	Gestión de retroalimentación y reclamaciones	Recoge, analiza y responde a la retroalimentación y reclamaciones de usuarios, pacientes o ciudadanía. Prioriza mejoras, documenta respuestas y verifica eficacia de las acciones adoptadas.

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

9. REFERENCIAS

1. International Organization for Standardization / International Electrotechnical Commission. (2023). ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system.
https://www.iso.org/standard/81230.html

2. International Organization for Standardization / International Electrotechnical Commission. (2023). ISO/IEC 23894:2023 Information technology — Artificial intelligence — Guidance on risk management.
https://www.iso.org/standard/77304.html

3. International Organization for Standardization. (2018). ISO 31000:2018 Risk management — Guidelines.
https://www.iso.org/standard/65694.html

4. International Organization for Standardization / International Electrotechnical Commission. (2022). ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection — Information security management systems — Requirements.
https://www.iso.org/standard/82875.html

5. European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR).
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

6. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0).
https://www.nist.gov/itl/ai-risk-management-framework

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

7. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

8. Council of Europe. (2022). HUDERIA Framework: Human Rights, Democracy and the Rule of Law Impact Assessment for Artificial Intelligence. https://www.coe.int/en/web/artificial-intelligence/huderia

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía

Metodología para el análisis y gestión de riesgos de IA en los entornos de salud y Smart Cities

Septiembre de 2025

Versión 1.0

Financiado por la Unión Europea NextGenerationEU
Gobierno de España · Ministerio para la Transformación Digital y de la Función Pública · Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales
Plan de Recuperación, Transformación y Resiliencia
incibe_ Instituto Nacional de Ciberseguridad
Junta de Andalucía
Agencia Digital de Andalucía