RETECH
Redes Territoriales de Especialización Tecnológica

RED ARGOS

Uso responsable y seguro de sistemas de IA en Salud para personas usuarias finales

Febrero de 2026

Versión 1.1

Junta de Andalucía

Financiado por la Unión Europea
NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA
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: 24/12/2025

Fecha de última actualización: 12/01/2026

ÍNDICE DE CONTENIDOS

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

Financiado por la Unión Europea NextGenerationEU

ÍNDICE DE TABLAS

1. Introducción

La utilización de sistemas de Inteligencia Artificial (IA) en el ámbito sanitario público se ha extendido de forma progresiva en los últimos años, incorporándose a procesos asistenciales, organizativos y de interacción con la ciudadanía. Estos sistemas se emplean, entre otros ámbitos, en el apoyo a la toma de decisiones clínicas, la estratificación y priorización de pacientes, la planificación de recursos, el análisis poblacional y la prestación de servicios digitales sanitarios.

A diferencia de otras tecnologías de la información sanitaria, los sistemas de IA se caracterizan por generar resultados inferenciales, tales como predicciones, recomendaciones, clasificaciones o contenidos generados de forma automatizada, a partir del tratamiento de datos. Estos resultados no se limitan a representar información existente, sino que influyen de manera directa o indirecta en decisiones humanas que pueden tener impacto sobre la seguridad del paciente, la calidad asistencial, el acceso a los servicios sanitarios y el ejercicio de derechos fundamentales.

En el sector público sanitario, este impacto se ve reforzado por la especial naturaleza de la actuación administrativa y asistencial, que exige decisiones motivadas, trazables y revisables, así como un nivel reforzado de protección de los datos de salud y de garantía de equidad y no discriminación. En este contexto, el riesgo asociado a los sistemas de IA no se limita a su diseño o desarrollo, sino que se materializa de forma especialmente relevante en la fase de uso, cuando los resultados generados por el sistema son interpretados y aplicados por personas concretas en situaciones reales.

La presente guía se elabora con el objetivo de abordar específicamente esta fase de uso, estableciendo criterios operativos y verificables para un uso responsable y seguro de los sistemas de IA en salud por parte de las personas usuarias finales. El documento se concibe como una guía operativa, orientada a reducir riesgos, reforzar la supervisión humana efectiva y promover una utilización de la IA alineada con los principios del sistema público de salud y con el marco normativo vigente.

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

El objetivo de esta guía es establecer criterios, principios y condiciones de uso que permitan garantizar un uso responsable, seguro y conforme a derecho de los sistemas de Inteligencia Artificial en el ámbito sanitario, desde la perspectiva de las personas usuarias finales que interactúan con dichos sistemas en el ejercicio de sus funciones.

En particular, la guía persigue:

- Promover un uso de la Inteligencia Artificial que priorice la seguridad del paciente, la calidad asistencial y la protección efectiva de los derechos de las personas.
- Prevenir riesgos derivados de la interpretación acrítica de los resultados generados por sistemas de IA, de la automatización de facto de decisiones y del uso de los sistemas fuera de la finalidad para la que han sido autorizados.
- Facilitar a las personas usuarias una comprensión adecuada del papel, las capacidades, las limitaciones y las responsabilidades asociadas al uso de sistemas de IA en contextos sanitarios reales.
- Reforzar la supervisión humana efectiva, la trazabilidad de las decisiones y la rendición de cuentas en los procesos asistenciales, organizativos y administrativos en los que interviene la IA.
- Contribuir al desarrollo de una cultura organizativa de uso diligente, responsable y trazable de la Inteligencia Artificial, alineada con los principios del sector público sanitario.

La guía no tiene por finalidad capacitar técnicamente en el funcionamiento interno de los sistemas de IA ni sustituir el criterio clínico, técnico o administrativo del personal profesional. Su propósito es establecer condiciones de uso que permitan integrar la IA de forma segura, controlada y coherente en los procesos sanitarios, complementando los mecanismos técnicos, organizativos y de gobernanza definidos por las entidades responsables.

2.2. Alcance

Esta guía se aplica al uso de sistemas de Inteligencia Artificial en el ámbito sanitario, cuando dichos sistemas se encuentran ya desplegados e integrados en procesos asistenciales, organizativos o de prestación de servicios sanitarios.

El alcance incluye:

- Sistemas de IA utilizados como apoyo a decisiones clínicas, diagnósticas o terapéuticas.
- Sistemas de IA aplicados a procesos de gestión sanitaria, planificación, priorización o asignación de recursos.
- Sistemas de IA utilizados en la interacción con pacientes o ciudadanía, cuando influyen en información, orientación o acceso a servicios sanitarios.
- Sistemas de IA generativa empleados en contextos sanitarios, incluidos aquellos utilizados para apoyo de manera informativo, documental o administrativo, siempre que su uso tenga impacto sobre procesos reales o sobre derechos de las personas.

La guía está dirigida a las siguientes personas usuarias finales:

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

- Personal sanitario que utiliza resultados generados por sistemas de IA en su práctica profesional.
- Personal de Tecnologías de la Información u otros perfiles técnicos que interactúan con sistemas de IA en su operación y supervisión en fase de uso.
- Pacientes, personas cuidadoras y ciudadanía, cuando interactúan directamente con sistemas de IA proporcionados por el sistema sanitario.

Quedan expresamente fuera del alcance de esta guía:

- El diseño, desarrollo, entrenamiento, validación o certificación de sistemas de IA.
- Los procesos de adquisición, contratación o evaluación técnica previa al despliegue.
- Las obligaciones de fabricantes, desarrolladores o proveedores de sistemas de IA, que se abordan en otras guías del proyecto.

3. Referencias normativas

Las referencias incluidas en este apartado se presentan con carácter informativo y orientativo, con el fin de ofrecer a la ciudadanía un contexto general sobre los marcos normativos y técnicos que regulan el uso de la inteligencia artificial en el ámbito sanitario público.

- **Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, Reglamento de Inteligencia Artificial (RIA):** Establece el marco jurídico común para el uso de sistemas de IA en la Unión Europea, introduciendo un enfoque basado en el riesgo. Clasifica como sistemas de alto riesgo determinados sistemas utilizados en el ámbito sanitario y exige garantías específicas durante la fase de uso, especialmente en materia de supervisión humana efectiva, gestión de riesgos, trazabilidad, control del uso y rendición de cuentas.

- **Reglamento (UE) 2016/679, Reglamento General de Protección de Datos (RGPD):** Resulta plenamente aplicable al uso de sistemas de IA en salud, dado que estos tratan habitualmente datos de salud, considerados categorías especiales de datos personales y sujetos a un nivel reforzado de protección, con especial incidencia en los principios de licitud, minimización, limitación de la finalidad, exactitud, seguridad y responsabilidad proactiva.

- **Reglamento (UE) 2025/327, por el que se establece el Espacio Europeo de Datos de Salud (European Health Data Space – EHDS):** Define el marco jurídico específico para el uso primario y secundario de los datos de salud en la Unión Europea, con impacto directo en los sistemas de IA utilizados en contextos asistenciales, de gestión sanitaria, investigación e innovación, reforzando los principios de control por parte de las personas, trazabilidad del acceso y protección de los datos de salud.

- **Ley Orgánica 3/2018, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD):** Desarrolla y complementa el RGPD en el ordenamiento español, incorporando derechos digitales y estableciendo obligaciones de transparencia, supervisión y rendición de cuentas aplicables también a sistemas basados en inteligencia artificial.

- **Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, Directiva NIS2:** Establece obligaciones reforzadas de ciberseguridad para entidades esenciales e importantes, incluyendo los servicios sanitarios públicos y los sistemas digitales críticos que soportan el uso de sistemas de IA en salud.

- **Esquema Nacional de Seguridad (ENS):** Define los principios básicos y los requisitos mínimos de seguridad aplicables a los sistemas de información del sector público español, incluidos aquellos que incorporan componentes de Inteligencia Artificial y tratan información sanitaria sensible, garantizando la confidencialidad, integridad, disponibilidad, autenticidad y trazabilidad de la información.

- **Reglamento (UE) 2017/745, relativo a los productos sanitarios (MDR**): Resulta de aplicación a aquellos sistemas de Inteligencia Artificial que, por su finalidad prevista, estén certificados

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

como productos sanitarios o como software como producto sanitario. En el contexto de esta guía, el MDR constituye el marco regulador que condiciona el uso seguro de dichos sistemas en entornos clínicos, en particular en lo relativo a su finalidad autorizada, las instrucciones de uso, las limitaciones clínicas y la necesidad de supervisión por parte de profesionales sanitarios cualificados.

- **Reglamento (UE) 2017/746, relativo a los productos sanitarios para diagnóstico in vitro (IVDR):** Es aplicable a los sistemas de IA utilizados como apoyo al diagnóstico in vitro, a la interpretación de resultados analíticos o a la toma de decisiones clínicas basadas en datos de laboratorio. En la fase de uso, el IVDR establece condiciones y límites que deben ser respetados por las personas usuarias finales, especialmente en relación con la interpretación de resultados, la validación clínica y la supervisión humana.

- **Data Governance Act, Reglamento (UE) 2022/868:** Regula la gobernanza y la reutilización segura de datos en la Unión Europea, fomentando el intercambio seguro y confiable de información y la protección de los datos sensibles, incluidos los de carácter sanitario.

- **Digital Services Act, Reglamento (UE) 2022/2065:** Establece obligaciones de transparencia, trazabilidad y diligencia debida para prestadores de servicios digitales y plataformas en línea, aplicables a la gestión de contenidos sintéticos generados mediante IA.

- **Ley 41/2002, de 14 de noviembre**, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. Establece el marco legal de referencia en materia de derechos del paciente, información clínica, consentimiento y toma de decisiones, elementos que deben ser respetados en todo uso de sistemas de IA que influyan en procesos asistenciales.

- **ISO/IEC 42001:2023 – Sistemas de gestión de la inteligencia artificial**: Establece requisitos para implantar y mantener un sistema de gestión de IA que garantice un uso seguro, responsable, trazable y alineado con principios éticos y legales en organizaciones públicas y sanitarias.

- **ISO/IEC 23894:2023 – Gestión del riesgo en la inteligencia artificial:** Proporciona un marco estructurado para la identificación, análisis y mitigación de los riesgos derivados del uso de sistemas de IA, incluidos los riesgos clínicos, organizativos, legales y de impacto sobre derechos fundamentales.

- **ISO/IEC 38507:2022 – Gobernanza de la inteligencia artificial:** Define directrices para integrar la IA en los sistemas de gobernanza organizativa, asignando responsabilidades claras, mecanismos de supervisión y control en la toma de decisiones asistidas por sistemas automatizados.

- **ISO/IEC 27001:2022 – Seguridad de la información, ciberseguridad y protección de la privacidad, junto con ISO 27799:2016 – Seguridad de la información en salud:** Constituyen el

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

marco de referencia para la protección de la información sanitaria y la gestión de la seguridad en entornos asistenciales y administrativos que incorporan sistemas de IA.

- **ISO 31000:2018 – Gestión del riesgo**: Proporciona principios y directrices generales para la gestión del riesgo organizativo, aplicables a los riesgos tecnológicos y algorítmicos asociados al uso de sistemas de IA en el ámbito sanitario.

- **NIST AI Risk Management Framework 1.0 (2023):** Marco de referencia internacional para la identificación, evaluación y mitigación de riesgos asociados al uso de la Inteligencia Artificial a lo largo de su ciclo de vida, utilizado como referencia técnica complementaria.

- **Guías nacionales sobre gobernanza, gestión de riesgos y uso responsable de la Inteligencia Artificial en el sector público, por AESIA** y los organismos competentes de la Administración General del Estado: las directrices, guías y criterios técnicos emitidos por AESIA constituyen el marco de referencia nacional para la gobernanza, el uso responsable y la supervisión efectiva de los sistemas de Inteligencia Artificial en el sector público, incluido el ámbito sanitario. Las dieciséis guías publicadas establecen principios, modelos organizativos y criterios metodológicos que deben ser considerados de forma complementaria a la presente guía, sin perjuicio de las especificidades propias del ámbito sanitario.

4. Definiciones

Este capítulo recopilará los conceptos clave y acrónimos utilizados a lo largo de la guía. Incluirá una tabla con definiciones de términos como inteligencia artificial, sistema automatizado, contenido sintético, datos personales, supervisión humana, ciberseguridad, accesibilidad digital, sesgo algorítmico, entre otros.

Término/Acrónimo	Definición
Inteligencia Artificial (IA)	Conjunto de sistemas informáticos que, a partir del análisis de datos y con distintos niveles de autonomía, generan resultados como predicciones, recomendaciones, clasificaciones, contenidos o decisiones que influyen en entornos físicos o digitales.
Algoritmo	Conjunto finito de reglas o instrucciones para resolver un problema o ejecutar un cálculo. En IA, incluye algoritmos de entrenamiento, inferencia, preprocesado y decisión.
Aprendizaje automático (Machine Learning, ML)	Rama de la IA en la que los modelos aprenden patrones a partir de datos para generalizar a casos nuevos, sin reglas explícitas para cada situación.
Aprendizaje profundo (Deep Learning, DL)	Subtipo de ML basado en redes neuronales profundas (múltiples capas) especialmente útil en imagen médica, lenguaje natural y señales.
IA generativa (GenAI)	Sistemas (normalmente basados en modelos de lenguaje o difusión) que generan contenido nuevo (texto, imagen, audio, código) coherente con patrones aprendidos. El contenido que producen puede parecer coherente y convincente, pero aun así ser incorrecto o inexacto.
Modelo fundacional (Foundation Model)	Modelo de propósito general entrenado con grandes volúmenes de datos (a menudo de múltiples dominios) que puede adaptarse a muchas tareas.
Contenido sintético	Contenido (texto, imagen, audio, vídeo) generado o alterado mediante técnicas automatizadas (incluida IA) de manera que no procede directamente de una captura o registro "natural" del hecho representado.
Resultado/Salida del sistema	Predicción, recomendación, clasificación, puntuación de riesgo, priorización o contenido generado por el sistema. Es lo que ve la persona usuaria final y lo que debe someterse a interpretación crítica.
Finalidad prevista / finalidad autorizada	Uso para el que un sistema (o, si aplica, un producto sanitario) ha sido diseñado, validado y autorizado. Usarlo fuera de esa finalidad aumenta riesgo y puede incumplir obligaciones regulatorias.
Uso fuera de finalidad	Aplicar el sistema a poblaciones, contextos, patologías, datos de entrada o decisiones distintas de las contempladas en su validación/instrucciones. Es una fuente frecuente de error operativo.

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

Término/Acrónimo	Definición
Supervisión humana efectiva	Capacidad real (no solo formal) de una persona competente para comprender el papel del sistema, revisar críticamente sus salidas, detectar incoherencias, intervenir y apartarse del resultado cuando proceda, dejando trazabilidad de la decisión.
Sesgo algorítmico (bias)	Desviación sistemática en el rendimiento o en los resultados del sistema que perjudica o favorece injustificadamente a ciertos grupos. Puede originarse por datos, diseño, contexto de uso o interacción humano-sistema.
Discriminación	Trato desfavorable o impacto perjudicial injustificado sobre una persona o colectivo por características protegidas o relevantes (p. ej., sexo, edad, origen). En IA puede aparecer como disparate impact (impacto desproporcionado) aunque no exista intención.
Equidad	Propiedad del sistema/proceso por la cual el acceso, la priorización y los resultados no generan desigualdades injustificadas entre grupos, teniendo en cuenta necesidades clínicas y contexto social.
Auditabilidad/ interpretabilidad	Grado en que se pueden entender las razones del resultado del sistema. En uso clínico, suele traducirse en "auditabilidad práctica": conocer qué datos influyeron, límites de aplicabilidad, incertidumbre y cómo actuar ante discrepancias.
Transparencia	Deber de informar de forma clara sobre: que se usa IA, para qué, qué hace y qué no hace, limitaciones, fuentes de datos relevantes, y cómo se garantiza supervisión, trazabilidad y revisión.
Trazabilidad	Capacidad de reconstruir a posteriori qué ocurrió: qué datos entraron, qué versión del sistema se usó, qué salida se mostró, a quién, cuándo, y qué decisión se tomó. En el sector público es clave para revisión y rendición de cuentas.
Registro (log)	Evidencia técnica (eventos, accesos, cambios, consultas, salidas) almacenada de forma segura para auditoría, análisis de incidentes y trazabilidad.
Gestión del riesgo	Proceso continuo de identificar, analizar, evaluar, tratar y monitorizar riesgos (clínicos, técnicos, legales, éticos) asociados al uso del sistema, con controles y responsables definidos.
Gobernanza de IA	Marco organizativo de roles, políticas, controles y decisiones para dirigir y supervisar el uso de IA (inventario, autorizaciones, cambios, formación, auditorías, incidencias, etc.).
Deriva (drift)	Pérdida de rendimiento del sistema por cambios en datos o contexto (población, protocolos, codificación, dispositivos). Puede ser gradual y pasar inadvertida si no se monitoriza.
Validación (clínica/operativa)	Evidencia de que el sistema cumple su finalidad con desempeño aceptable en condiciones específicas. La validación "en laboratorio" no garantiza desempeño "en producción" si cambia el contexto.

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

Término/Acrónimo	Definición
Agencia Española de Supervisión de la Inteligencia Artificial (AESIA)	Organismo público encargado de garantizar el uso ético y seguro de la inteligencia artificial en España.
Iatrogenia	Daño no deseado, no buscado y, en muchos casos, inevitable, que una persona sufre como consecuencia de una intervención sanitaria correcta, ya sea un tratamiento, medicamento, cirugía o procedimiento, y que aun siendo adecuado produce un perjuicio no intencionado.
Autenticación multifactor (MFA)	Método de identificación que requiere más de una prueba de verificación para acceder a una cuenta o sistema (por ejemplo, contraseña + código SMS/ huella).
Off-label	Uso de un medicamento en condiciones diferentes a las aprobadas en su ficha técnica, ya sea para otra enfermedad, dosis, población o vía de administración.
SIEM	Herramienta de ciberseguridad que recopila, correlaciona y analiza en tiempo real los registros de actividad de sistemas, redes y aplicaciones para detectar amenazas y anticiparse a potenciales incidentes.

TABLA 1: DEFINICIONES EMPLEADAS.

5. La inteligencia artificial en el entorno de Salud

La Inteligencia Artificial se ha incorporado al ámbito sanitario público como una tecnología transversal que apoya, complementa o condiciona procesos asistenciales, organizativos y de relación con la ciudadanía. Su utilización se produce en un contexto caracterizado por una elevada complejidad clínica, una fuerte presión sobre los recursos disponibles y la necesidad de garantizar decisiones seguras, equitativas y basadas en criterios objetivos, sin menoscabar la calidad asistencial ni los derechos de las personas.

A diferencia de los sistemas de información sanitaria tradicionales, los sistemas de Inteligencia Artificial no se limitan a almacenar, transmitir o presentar información clínica, sino que generan resultados inferenciales a partir del tratamiento automatizado de datos. Estos resultados pueden adoptar la forma de predicciones, recomendaciones, clasificaciones, priorizaciones o contenidos generados automáticamente, y son utilizados por personas concretas como profesionales sanitarios, personal técnico o pacientes, en situaciones reales de atención sanitaria.

Desde la perspectiva del sector público sanitario, la introducción de la IA no puede entenderse únicamente como una mejora tecnológica. Su uso se integra en procesos regulados, sometidos a principios de legalidad, seguridad del paciente, transparencia, trazabilidad, igualdad de acceso y rendición de cuentas. En consecuencia, el valor de la IA en salud depende no solo de su capacidad técnica, sino de las condiciones en las que se utiliza y de la forma en que sus resultados son interpretados, supervisados y aplicados por las personas usuarias finales.

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

5.1. Tipologías de sistemas de IA utilizados en salud

En el ámbito sanitario público, los sistemas de Inteligencia Artificial pueden agruparse en distintas tipologías funcionales en función de su finalidad y del tipo de apoyo que proporcionan a los procesos existentes. Esta diversidad funcional es relevante porque condiciona directamente los riesgos asociados al uso, el tipo de supervisión necesaria y el impacto potencial sobre pacientes y ciudadanía.

De forma no exhaustiva, pueden distinguirse las siguientes tipologías principales:

- **Sistemas de apoyo a la toma de decisiones clínicas**, que generan recomendaciones, alertas o clasificaciones a partir de datos clínicos, imágenes médicas, resultados analíticos o historiales asistenciales. Estos sistemas influyen de manera significativa en la valoración clínica y en la priorización de actuaciones, aun cuando no sustituyen el juicio profesional ni la decisión clínica final.
- **Sistemas de IA orientados a la gestión sanitaria y a la planificación de recursos**, utilizados para prever demandas asistenciales, optimizar agendas, priorizar listas de espera o asignar recursos humanos y materiales. Aunque su impacto no siempre es clínico de forma directa, sí puede afectar de manera relevante al acceso efectivo a los servicios sanitarios, a la equidad del sistema y a la organización de la atención.
- **Sistemas de IA utilizados en la interacción con pacientes y ciudadanía**, tales como asistentes virtuales, sistemas de triaje automatizado, herramientas de orientación sanitaria o generación de contenidos personalizados. En estos casos, el riesgo se vincula especialmente a la calidad, claridad y adecuación de la información proporcionada, así como a la posibilidad de inducir decisiones incorrectas, retrasos en la atención o una falsa sensación de seguridad.
- **Sistemas de IA generativa empleados en contextos sanitarios para apoyo documental, administrativo o informativo**, siempre que su uso tenga impacto sobre procesos reales, decisiones humanas o derechos de las personas. Aunque estos sistemas no actúan directamente sobre el paciente, pueden influir de manera relevante en la elaboración de informes, comunicaciones clínicas o documentación administrativa, lo que exige un control específico de su uso.

5.2. La fase de uso como punto crítico de riesgo

El riesgo asociado a los sistemas de Inteligencia Artificial en salud no se materializa exclusivamente en su diseño o desarrollo, sino de forma especialmente relevante en la fase de uso operativo. Es en este momento cuando los resultados generados por el sistema se integran en procesos reales, son interpretados por personas concretas y se traducen en decisiones clínicas, organizativas o informativas con efectos directos sobre pacientes y ciudadanía.

En la práctica, la persona usuaria final no interactúa con el modelo de IA en sí, sino con sus salidas. Esta circunstancia puede generar una percepción de objetividad o neutralidad que favorezca una confianza excesiva en el sistema, especialmente cuando los resultados se presentan de forma automatizada, con apariencia de precisión, consistencia o respaldo estadístico.

La fase de uso es, por tanto, el punto en el que pueden producirse fenómenos críticos como:

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 automatización de facto de decisiones.
- La utilización de resultados fuera de su finalidad autorizada.
- La aplicación acrítica de recomendaciones generadas por el sistema.
- La pérdida de una supervisión humana efectiva.

Estos riesgos se ven incrementados en contextos de alta carga asistencial, presión temporal o escasez de recursos, habituales en el sistema sanitario público. Por este motivo, el uso de la IA en salud exige condiciones específicas que garanticen que los resultados generados por el sistema se interpreten de forma contextualizada, se contrasten con el juicio profesional y se utilicen dentro de los límites establecidos por su finalidad, su validación y el marco normativo aplicable.

5.3. Supervisión humana y responsabilidad en el uso de la IA

Uno de los elementos estructurales del uso responsable de la Inteligencia Artificial en salud es la existencia de una supervisión humana efectiva. Esta supervisión no puede entenderse como una mera posibilidad formal de intervención, sino como una capacidad real y ejercida por personas competentes para comprender, cuestionar y, en su caso, apartarse de los resultados generados por el sistema.

La supervisión humana implica que la persona usuaria final conozca las condiciones de uso del sistema, sus limitaciones, los supuestos para los que ha sido diseñado y los márgenes de error razonables. Asimismo, exige que exista la posibilidad organizativa y técnica de revisar decisiones, documentar discrepancias y justificar la adopción o no de las recomendaciones de la IA.

Desde el punto de vista de la responsabilidad, el uso de sistemas de IA no desplaza la responsabilidad profesional, técnica o administrativa de las personas que intervienen en los procesos sanitarios. Los resultados generados por la IA constituyen un elemento de apoyo, pero no una decisión autónoma. En consecuencia, la responsabilidad última de las decisiones adoptadas recae en el personal profesional y en la organización sanitaria, conforme a sus respectivos ámbitos de competencia.

5.4. Relación con la gobernanza y la gestión de riesgos de la IA

El uso de sistemas de Inteligencia Artificial en salud se inscribe en un marco organizativo más amplio de gobernanza y gestión de riesgos, definido por las entidades responsables del sistema sanitario. La presente guía se centra en la fase de uso por parte de las personas usuarias finales, pero debe entenderse como complementaria a los modelos de gobierno de la IA y a las metodologías de gestión de riesgos establecidas a nivel institucional.

Los criterios de uso responsable y seguro definidos en esta guía se apoyan en la identificación previa del nivel de riesgo de los sistemas de IA, en la asignación de responsabilidades organizativas y en la existencia de mecanismos de control y supervisión definidos por la entidad sanitaria. La coherencia entre estos elementos es esencial para garantizar que el uso cotidiano de la IA no se aparte de los objetivos de seguridad, legalidad y protección de derechos establecidos a nivel estratégico.

De este modo, la correcta utilización de la IA en salud no depende únicamente del comportamiento individual de la persona usuaria final, sino de la integración efectiva entre gobernanza, gestión de riesgos

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

y prácticas de uso, asegurando que los sistemas de IA aporten valor al sistema sanitario sin generar riesgos inaceptables para las personas ni para la organización.

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. Riesgos asociados al uso de sistemas de IA en el ámbito sanitario

El uso de sistemas de Inteligencia Artificial en el ámbito sanitario introduce un conjunto específico de riesgos que se materializan, de forma principal, durante la fase de uso operativo. Estos riesgos no derivan únicamente de fallos técnicos del sistema, sino de la interacción entre los resultados generados por la IA, los procesos asistenciales u organizativos en los que se integran y las decisiones adoptadas por las personas usuarias finales.

En el contexto del sector público sanitario, la materialización de estos riesgos puede tener consecuencias directas sobre la seguridad del paciente, la calidad asistencial, el ejercicio de derechos fundamentales, la validez de las decisiones administrativas y la confianza de la ciudadanía en el sistema sanitario. Por ello, resulta imprescindible identificar y comprender los principales tipos de riesgo asociados al uso de la IA, como paso previo a la definición de condiciones de uso responsable y seguro.

6.1. Riesgos técnicos

Los riesgos técnicos se derivan de vulnerabilidades, fallos o limitaciones en los componentes tecnológicos que conforman los sistemas de IA (modelos, datos, infraestructuras, integraciones, controles de seguridad y mecanismos de registro). Su materialización puede comprometer tanto la fiabilidad de los resultados como la confidencialidad, integridad y disponibilidad de la información sanitaria.

I. Calidad, representatividad y actualización de los datos

Los sistemas de IA dependen críticamente de los datos con los que fueron entrenados y de los datos que procesan durante su operación. En la práctica asistencial, se producen riesgos cuando los datos están incompletos, mal codificados, no son representativos de la población atendida o han quedado desactualizados respecto del contexto clínico. Esto puede llevar a que el sistema presente un rendimiento inferior y sistemáticamente erróneo para determinados grupos demográficos, contraviniendo los principios de equidad y no discriminación.

Caso documentado (Epic Sepsis Model, 2021): Una evaluación externa de este modelo predictivo de sepsis, ampliamente utilizado, mostró un rendimiento sustancialmente inferior en un entorno real al reportado por el desarrollador. La falta de generalización del modelo a nuevas poblaciones de pacientes evidenció el riesgo de trasladar sin cautelas resultados de validaciones internas a otros escenarios asistenciales, con el potencial de generar falsos negativos en diagnósticos críticos.

II. Degradación del rendimiento en producción (drift) y cambios del contexto clínico

Incluso cuando un sistema fue validado, su rendimiento puede degradarse por cambios en la población atendida (case-mix), en los protocolos clínicos, en los dispositivos/formatos de captura o en la calidad del dato en producción. Este riesgo es especialmente crítico cuando el sistema se percibe como “estable” y no existen señales claras para la persona usuaria final de que el rendimiento está cambiando.

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 términos operativos, la deriva no siempre es visible “caso a caso”, y puede consolidarse durante semanas o meses si no existe monitorización institucional y alertas de degradación.

III. Vulnerabilidades en modelos, software e integraciones

Además de las vulnerabilidades comunes del software, los sistemas de IA incorporan riesgos específicos que atacan directamente la lógica del modelo:

- **Ataques Adversarios:** Manipulación de los datos de entrada mediante la adición de perturbaciones sutiles e imperceptibles para un humano (p.ej., un patrón de ruido en una imagen médica) con el fin de inducir un error de clasificación deliberado en el modelo.
- **Envenenamiento de Datos (Data Poisoning):** Corrupción de los datos de entrenamiento para introducir "puertas traseras" o sesgos que un atacante puede explotar posteriormente, degradando la fiabilidad del modelo de forma sigilosa.
- **Errores de integración:** Fallos en el mapeo de variables, conversiones de unidades o normalizaciones incorrectas que cambian el significado clínico de los datos de entrada, llevando al modelo a operar sobre premisas falsas.

Caso documentado (Vulnerabilidad de sistemas de diagnóstico por imagen): Investigaciones académicas han demostrado de forma consistente que los sistemas de aprendizaje profundo para diagnóstico por imagen son vulnerables a perturbaciones adversarias. Estas pueden inducir errores de clasificación (p.ej., diagnosticar una lesión maligna como benigna) con un alto grado de confianza por parte del modelo, lo que fundamenta la necesidad de controles de ciberseguridad y validación técnica reforzada cuando la IA se integra en circuitos clínicos.

IV. Falta de trazabilidad técnica y registros suficientes

Cuando no existen registros adecuados, se dificulta reconstruir qué mostró el sistema, a qué persona usuaria, con qué versión del modelo y en qué condiciones. Esto afecta directamente a la capacidad de auditoría, análisis de incidentes, aprendizaje organizativo y rendición de cuentas. En sistemas con impacto clínico o administrativo, la trazabilidad no es un “extra”: condiciona la posibilidad real de supervisión humana efectiva y de revisión de decisiones.

6.2. Riesgos operativos

Los riesgos operativos afectan al funcionamiento cotidiano de los servicios sanitarios y a la interacción real entre personas y sistemas. Suelen emerger por la combinación de presión asistencial, integración deficiente en los flujos de trabajo, diseño inadecuado de procesos y expectativas erróneas sobre el rol de la IA.

I. Automatización de facto y sesgo de automatización (automation bias)

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

Se produce cuando el resultado de la IA se acepta de manera rutinaria sin revisión crítica, o cuando actúa como criterio dominante por carga de trabajo, urgencia o cultura organizativa. Se manifiesta, típicamente, como:

- Aceptación sistemática del resultado,
- Validaciones mecánicas sin contraste con el juicio profesional,
- Disminución del pensamiento crítico en casos atípicos o complejos.

Caso documentado (Algoritmo de gestión de riesgos de Optum, 2019): Un análisis publicado en la revista Science mostró cómo un algoritmo ampliamente utilizado en EE. UU. para la gestión poblacional infravaloraba sistemáticamente las necesidades clínicas de los pacientes de raza negra al utilizar el gasto sanitario previo como un proxy de la necesidad de salud. Esto evidencia que un sistema puede "funcionar" operacionalmente y, aun así, consolidar decisiones con sesgos sistémicos si no existen supervisión crítica y controles de equidad.

II. Integración deficiente en flujos de trabajo y fatiga por alertas (alert fatigue)

La IA aporta valor cuando está integrada en un flujo de trabajo con roles, tiempos y responsabilidades definidos. Si no, puede generar duplicidades, retrasos, contradicciones con protocolos, y sobrecarga cognitiva por exceso de alertas, aumentando el riesgo de que señales relevantes sean ignoradas de forma sistemática. En términos de uso real, este riesgo no es "del modelo" sino del encaje proceso-persona-tecnología, donde un diseño no centrado en la persona usuaria final puede tener consecuencias clínicas directas.

Caso documentado (Fatiga por alarmas en sistemas de monitorización, Alemania): Un análisis de incidentes reportados al Instituto Federal Alemán de Medicamentos y Productos Sanitarios (BfArM) entre 2009 y 2015 sobre monitores de pacientes reveló un problema análogo y directamente aplicable a la IA. Casi la mitad de los incidentes reportados como "fallos del dispositivo" por el personal clínico correspondían, en realidad, al funcionamiento previsto por el fabricante. La complejidad de las funciones y la sobrecarga de alarmas llevaban al personal a malinterpretar el comportamiento del sistema o a desactivar alertas, con el consiguiente riesgo de omitir eventos clínicos críticos. Este fenómeno, bien documentado como "fatiga por alarmas", se exacerba con sistemas de IA que pueden generar un volumen aún mayor de alertas predictivas.

III. Uso fuera de finalidad o fuera de condiciones de validez

Incluye aplicar el sistema a poblaciones, patologías o escenarios no contemplados en su validación, o emplear salidas clínicas para decisiones administrativas. En sistemas sujetos al Reglamento de Productos Sanitarios (MDR/IVDR), el uso fuera de la finalidad prevista puede comprometer la seguridad del paciente y generar responsabilidades organizativas, tensionando la validez de las decisiones. La eficacia y seguridad de un modelo de IA están estrictamente ligadas a los datos con los que fue entrenado y validado.

Caso documentado (Riesgo de extrapolación en modelos de IA): El principio está claramente establecido en múltiples dominios clínicos. Por ejemplo, un modelo de IA validado para la detección de retinopatía diabética en adultos no puede ser utilizado "off-label" en pacientes pediátricos. La

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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

fisiopatología, la progresión de la enfermedad y los artefactos de imagen son diferentes, invalidando las garantías del modelo. Hacerlo constituiría una extrapolación sin evidencia, con un alto riesgo de diagnósticos erróneos (falsos positivos o negativos). Este principio se aplica a cualquier extrapolación a demografías, comorbilidades o equipamiento no incluidos en la validación original del sistema.

IV. "Shadow AI" y herramientas no autorizadas en el trabajo diario

En entornos reales, es frecuente que profesionales utilicen herramientas no aprobadas (especialmente IA generativa) para resumir historia clínica, redactar informes o preparar comunicaciones. Esto introduce riesgos simultáneos como:

- **Pérdida de control de datos:** La introducción de cualquier dato de un paciente en una plataforma pública constituye una cesión de datos personales de categoría especial a un tercero, a menudo fuera del Espacio Económico Europeo, sin una base legal, sin un acuerdo de tratamiento de datos (DPA) y en violación directa del RGPD.
- **Falta de trazabilidad:** La acción queda fuera de cualquier registro de auditoría de la organización.
- **Riesgo de contenido incorrecto (alucinaciones):** Los modelos de IA generativa pueden inventar datos ("alucinar") con un estilo de redacción verosímil, introduciendo información falsa en un resumen clínico que puede llevar a decisiones peligrosas.
- **Ausencia de garantías organizativas:** La organización pierde toda capacidad de gobernar, securizar y responder por el tratamiento de esos datos.

Caso documentado (Bloqueo de ChatGPT por el Garante italiano, 2023): La autoridad italiana de protección de datos adoptó medidas temporales contra el servicio ChatGPT por riesgos vinculados al tratamiento ilícito de datos personales, falta de transparencia y ausencia de base legal. Este caso ilustra el riesgo regulatorio y operativo de usar plataformas externas sin control institucional, que resulta aún más crítico cuando se introducen datos de salud protegidos por el RGPD.

6.3. Riesgos estratégicos, éticos y sociales

Estos riesgos trascienden el plano técnico-operativo. Afectan a derechos fundamentales, legitimidad institucional, confianza pública y cumplimiento de principios del sector público sanitario (equidad, no discriminación, motivación y auditabilidad).

I. Sesgos, discriminación y afectación de la equidad

Un sistema de IA puede reproducir, amplificar y consolidar sesgos existentes en los datos sanitarios históricos, llevando a un comportamiento sistemáticamente diferente y perjudicial para subpoblaciones específicas (definidas por sexo, edad, origen étnico, comorbilidades o nivel socioeconómico). El riesgo se materializa de forma crítica cuando la IA influye en la priorización de listas de espera, la derivación a especialistas, el seguimiento de crónicos, la asignación de recursos o el acceso efectivo a servicios, pudiendo perpetuar o incluso agravar desigualdades sanitarias preexistentes.

Caso documentado (Algoritmo de gestión de riesgos de Optum, 2019): Un análisis publicado en la revista Science reveló que un algoritmo comercial, utilizado por grandes aseguradoras y hospitales en

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

EE.UU. para identificar pacientes de *alto riesgo* que necesitaban atención adicional, discriminaba sistemáticamente a los pacientes de raza negra. El modelo utilizaba el gasto sanitario previo como un indicador (proxy) de las necesidades futuras de salud. Debido a desigualdades socioeconómicas históricas, los pacientes negros con el mismo nivel de enfermedad generaban, en promedio, un menor gasto sanitario. Como resultado, el algoritmo les asignaba una puntuación de riesgo mucho menor que a los pacientes blancos con un estado de salud equivalente, negándoles el acceso a programas de atención reforzada. Este caso es un ejemplo paradigmático de cómo un modelo, sin una intención discriminatoria explícita, puede generar un impacto adverso desproporcionado (discriminación indirecta) con consecuencias directas sobre la equidad en salud.

II. Falta de transparencia y explicación en el punto de uso

En el ámbito sanitario, la opacidad de un sistema de IA afecta directamente a la relación médico-paciente, a la validez de las decisiones y a la confianza en el sistema. Si un profesional no puede comprender, al menos a un nivel funcional, por qué un modelo ha generado una recomendación, se ve incapacitado para:

- Explicar al paciente de forma adecuada las bases de una decisión clínica.
- Motivar debidamente un acto administrativo que se apoya en dicha recomendación.
- Realizar una supervisión humana verdaderamente efectiva, ya que se limita a aceptar o rechazar el resultado sin un entendimiento crítico.

Caso documentado (IBM Watson for Oncology): La adopción clínica de este sistema de apoyo a la decisión oncológica se vio limitada por su carácter de "caja negra". El sistema generaba recomendaciones de tratamiento que, en ocasiones, no se alineaban con la práctica clínica estándar o se basaban en evidencia considerada débil por los especialistas. La incapacidad de la plataforma para proporcionar un razonamiento clínico claro y comprensible que justificara sus conclusiones impidió la generación de confianza entre el personal médico, que debía optar entre aceptar una recomendación opaca o descartarla, dificultando su integración efectiva en el proceso de toma de decisiones.

III. Riesgos legales y de responsabilidad organizativa

La falta de trazabilidad en el uso de un sistema de IA o la incapacidad de documentar el proceso de supervisión humana crean un "vacío de responsabilidad" en caso de incidente adverso. Esta situación dificulta la determinación de la causa raíz del error y la atribución de responsabilidades entre el desarrollador, la institución sanitaria y el profesional. En el sector público, esto entra en conflicto directo con el principio de motivación y auditabilidad de los actos, y compromete la capacidad de la organización para responder ante reclamaciones o auditorías.

Escenario de riesgo (Atribución de responsabilidad en diagnóstico erróneo): Un paciente sufre un perjuicio derivado de un retraso diagnóstico en una patología que un sistema de IA no identificó. Ante una reclamación, la institución sanitaria, el profesional y el fabricante podrían atribuirse mutuamente la responsabilidad. Si los registros de auditoría (logs) son insuficientes para demostrar qué información exacta proporcionó la IA (incluido su grado de incertidumbre) y cómo interactuó el profesional con ella, se origina un vacío probatorio que compromete la defensa jurídica de la organización y los derechos del

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

paciente. El Reglamento de IA (RIA) aborda este riesgo al exigir capacidades de registro robustas para los sistemas de alto riesgo.

IV. Pérdida de confianza institucional

Los proyectos que implican el tratamiento de datos de salud son altamente sensibles a la percepción pública. Un incidente de seguridad, una filtración de datos, o una comunicación deficiente sobre el uso y la gobernanza de la IA pueden erosionar la confianza de la ciudadanía y del personal profesional de forma rápida e irreversible. Dicha pérdida de confianza puede afectar no solo al proyecto en cuestión, sino a la adopción futura de cualquier tecnología sanitaria digital.

Caso documentado (Proyecto "care.data" del NHS, Reino Unido, 2014): Esta iniciativa del Servicio Nacional de Salud británico pretendía centralizar los historiales clínicos de atención primaria para fines de investigación y planificación sanitaria. A pesar de sus potenciales beneficios, el proyecto fue cancelado debido a una reacción social adversa. La causa principal fue una estrategia de comunicación que no informó de manera transparente sobre los fines del tratamiento, las salvaguardas de privacidad y el procedimiento de exclusión voluntaria (opt-out). La percepción de falta de control sobre los datos personales generó una desconfianza generalizada. Este caso constituye un precedente sobre cómo la ausencia de una "licencia social para operar", basada en la transparencia y la confianza, puede provocar el fracaso estratégico de iniciativas de datos de salud, un riesgo que se intensifica con el uso de tecnologías de IA.

6.4. Riesgos de confidencialidad y protección de datos de salud

El uso de sistemas de IA en salud implica, por definición, el tratamiento de datos personales de categoría especial (datos de salud), sujetos al máximo nivel de protección según el Reglamento General de Protección de Datos (RGPD). Los riesgos en este dominio no se limitan a brechas de seguridad externas, sino que incluyen prácticas operativas que pueden derivar en la pérdida de control, el uso indebido de la información o transferencias no autorizadas.

I. Exposición indebida por uso de plataformas no autorizadas (especialmente IA generativa)

El uso por parte del personal de herramientas de IA no aprobadas por la organización, en particular plataformas de IA Generativa de acceso público, para procesar información clínica real, constituye un riesgo de alta gravedad. Esta práctica implica una cesión de datos personales de categoría especial a un tercero, a menudo localizado fuera del Espacio Económico Europeo, sin una base legal que lo legitime, sin un contrato de encargado de tratamiento (Art. 28 RGPD) y en violación directa de los principios de confidencialidad e integridad.

II. Reidentificación por combinación de fuentes y datos auxiliares

Incluso cuando los sistemas operan sobre datos pseudonimizados, la combinación de múltiples variables (clínicas, demográficas, geográficas, socioeconómicas) puede aumentar exponencialmente el

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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

riesgo de reidentificación de un individuo. Este riesgo es particularmente elevado en conjuntos de datos que incluyen patologías raras, poblaciones de áreas geográficas pequeñas o combinaciones de variables muy singulares, lo que puede permitir inferir la identidad del paciente.

Caso documentado (Reidentificación de datos genómicos, 2013): Investigadores de la Universidad de Cambridge demostraron la posibilidad de reidentificar a participantes de un proyecto público de datos genómicos (el "1000 Genomes Project") cruzando la información con bases de datos genealógicas de acceso público y otros metadatos como la edad y el lugar de residencia. Este caso, aunque no relacionado directamente con IA, establece un precedente fundamental sobre los riesgos de reidentificación en conjuntos de datos sanitarios complejos, riesgo que la capacidad de análisis de la IA puede amplificar.

III. Minimización y limitación de finalidad en el uso real

El principio de minimización de datos (Art. 5.1.c RGPD) exige que el tratamiento de datos sea adecuado, pertinente y limitado a lo necesario en relación con los fines. En la fase de uso de la IA, este riesgo se materializa cuando se accede, procesa o exporta un volumen de datos mayor del estrictamente indispensable para la tarea en cuestión, o cuando la información generada por el sistema se reutiliza para finalidades distintas de las previstas originalmente

6.5. Riesgos de disponibilidad, continuidad asistencial y dependencia tecnológica

Cuando la IA se integra en procesos clínicos o administrativos críticos, su indisponibilidad, degradación o retirada no planificada puede afectar de forma directa a la continuidad de la atención sanitaria y a la seguridad del paciente.

I. Indisponibilidad del sistema y ausencia de procedimientos alternativos

La integración de la IA en procesos de alta criticidad, como el triaje en servicios de urgencias, la priorización de listas de espera quirúrgicas o el soporte al diagnóstico en tiempo real, crea una fuerte dependencia operativa. La interrupción del servicio, si no se acompaña de procedimientos alternativos manuales claramente definidos, comunicados y ensayados, puede provocar un aumento de los tiempos de respuesta, errores en la gestión de pacientes y, en última instancia, un deterioro de la calidad asistencial.

Caso documentado (Ataque de ransomware al Hospital Universitario de Düsseldorf, 2020): Un ciberataque paralizó 30 servidores del hospital, afectando a la totalidad de sus sistemas digitales. Como consecuencia, el servicio de urgencias tuvo que cerrar y los pacientes críticos fueron desviados a otros centros. Trágicamente, una paciente en situación de emergencia vital falleció durante su traslado a otro hospital, a 30 km de distancia. Aunque el ataque no fue dirigido a un sistema de IA, este incidente demuestra de forma inequívoca el impacto directo sobre la seguridad del paciente que tiene la indisponibilidad de sistemas digitales críticos y la ausencia de planes de contingencia eficaces.

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

II. Dependencia de proveedores y efectos sistémicos

La utilización de sistemas de IA, especialmente aquellos basados en plataformas en la nube (cloud) o proporcionados como un servicio gestionado (SaaS), introduce una dependencia estratégica del proveedor. Un incidente de seguridad, un fallo técnico o una interrupción del servicio en la infraestructura del proveedor pueden tener un impacto sistémico, afectando simultáneamente a todas las entidades sanitarias que utilizan dicho servicio.

III. Degradación silenciosa del servicio

A diferencia de una interrupción total del servicio, la degradación silenciosa es una disminución gradual y a menudo no perceptible del rendimiento del sistema. Puede manifestarse como un aumento de la latencia en las respuestas, fallos intermitentes en la integración con otros sistemas o una pérdida de precisión del modelo (deriva) que no es monitorizada. Este riesgo es particularmente insidioso, ya que el sistema mantiene una apariencia de normalidad operativa, mientras que los profesionales basan sus decisiones en resultados que han perdido progresivamente su fiabilidad y validez.

6.6. Riesgos para la seguridad del paciente

Esta constituye la categoría de riesgo de mayor criticidad, ya que su materialización tiene consecuencias directas sobre la salud y la integridad física de las personas. Estos riesgos se producen cuando un sistema de IA influye, de forma incorrecta o no fiable, en decisiones diagnósticas, terapéuticas, de priorización asistencial o de seguimiento. El riesgo se agrava si el resultado del sistema es interpretado por personas usuarias finales como una verdad objetiva, en lugar de como una inferencia probabilística con un grado de incertidumbre y unas condiciones de validez específicas.

I. Errores diagnósticos inducidos o no detectados (falsos negativos / falsos positivos)

Un sistema de IA puede inducir errores diagnósticos de dos tipos principales, ambos con graves consecuencias potenciales:

- **Falsos negativos:** El sistema no detecta una patología o un hallazgo clínico que está presente. Esto puede provocar un retraso en el diagnóstico y, consecuentemente, en el inicio del tratamiento, con un posible empeoramiento del pronóstico del paciente.
- **Falsos positivos:** El sistema identifica erróneamente una patología o un hallazgo que no existe. Esto puede conducir a la realización de pruebas diagnósticas invasivas e innecesarias, generar ansiedad en el paciente y derivar en iatrogenia. El riesgo de ambos tipos de error aumenta significativamente cuando el sistema se utiliza fuera de su contexto de validación, en casos con presentaciones clínicas atípicas o cuando el profesional delega su criterio en el sistema debido al sesgo de automatización.

II. Recomendaciones terapéuticas inadecuadas o no aplicables

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

Cuando un sistema de IA genera recomendaciones de tratamiento, dosificación de fármacos o planes terapéuticos, el riesgo para la seguridad del paciente es directo e inmediato. Una recomendación puede ser inadecuada por múltiples razones:

- El sistema no tiene acceso al contexto clínico completo del paciente (ej. comorbilidades no registradas, alergias, interacciones con otros fármacos).
- El modelo fue entrenado con guías clínicas o evidencia científica desactualizadas.
- El sistema genera una recomendación que, aunque teóricamente plausible, es incorrecta o peligrosa para un paciente específico.

III. Priorización inadecuada y riesgos en triaje/urgencias

Los sistemas de IA utilizados para el triaje en servicios de urgencias o para la priorización de listas de espera gestionan un recurso crítico: el tiempo de acceso a la atención sanitaria. Un error de priorización puede tener consecuencias graves.

- **Infravaloración del riesgo:** El sistema asigna una prioridad baja a un paciente con una condición grave, pero con una presentación atípica, retrasando su atención.
- **Sobrevaloración del riesgo:** El sistema asigna una prioridad alta a casos de baja gravedad, consumiendo recursos que deberían destinarse a pacientes más urgentes. En la fase de uso, este riesgo se maximiza cuando el flujo de trabajo convierte la puntuación generada por el sistema en una decisión "de facto", y el profesional pierde la capacidad real (o el tiempo disponible) para evaluar al paciente de forma integral y contradecir la recomendación algorítmica.

IV. Riesgo acumulativo por encadenamiento de decisiones asistidas por IA

En un entorno sanitario digitalizado, un mismo paciente puede verse afectado por una cadena de decisiones, cada una de ellas asistida por un sistema de IA diferente (p. ej., un modelo predice el riesgo de reingreso, otro prioriza la cita con el especialista, un tercero apoya el diagnóstico por imagen y un cuarto sugiere el tratamiento). Aunque cada sistema, de forma aislada, tenga un rendimiento aceptable y un riesgo gestionado, la interacción y el encadenamiento de sus salidas pueden amplificar errores y sesgos de formas imprevistas.

Este riesgo sistémico es especialmente difícil de detectar y gestionar si la organización no dispone de un marco de gobernanza de la IA que contemple una supervisión transversal, así como de la trazabilidad completa de la secuencia de intervenciones algorítmicas que afectan la ruta asistencial del paciente

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. Principios de uso responsable de sistemas de IA en Salud

El uso seguro y responsable de sistemas de Inteligencia Artificial en el ámbito sanitario debe sustentarse en un conjunto de principios operativos claros, que orienten el comportamiento de las personas usuarias finales y sirvan como marco de referencia para la toma de decisiones en situaciones complejas o no previstas. Estos principios no sustituyen a la normativa vigente ni a los protocolos clínicos, pero establecen una base común de actuación que permite reducir riesgos, proteger al paciente y garantizar la confianza en los sistemas sanitarios digitales.

7.1. Responsabilidad compartida

Conforme a los requisitos establecidos en el Reglamento de IA de la UE, todo sistema, especialmente los de alto riesgo, debe diseñarse y operar sujeto a una supervisión humana efectiva. La autonomía de la IA es, por tanto, siempre limitada. Los profesionales responsables deben tener la capacidad real de:

- Comprender las capacidades, limitaciones y el grado de incertidumbre del sistema.
- Interpretar correctamente sus resultados y supervisar su funcionamiento para detectar anomalías.
- Decidir, en una situación particular, no utilizar el sistema o anular, ignorar o invertir un resultado si este se considera incorrecto, inadecuado o puede comprometer la seguridad del paciente.

7.2. Precisión, Robustez y Ciberseguridad

El Reglamento de IA y la normativa de ciberseguridad, como la Directiva NIS2, exigen que los sistemas de IA alcancen un nivel adecuado de precisión, fiabilidad y robustez técnica a lo largo de su ciclo de vida. Este principio implica que el sistema debe ser seguro y funcionar de manera fiable en relación con su finalidad prevista, lo que se desglosa en:

- **Precisión:** El sistema debe funcionar al nivel de rendimiento declarado en su validación, con márgenes de error definidos y comunicados a las personas usuarias.
- **Robustez:** El sistema debe ser resiliente a errores, fallos o inconsistencias en los datos de entrada, así como a intentos de manipulación maliciosa, manteniendo su rendimiento ante situaciones imprevistas.
- **Ciberseguridad:** El sistema debe incorporar las medidas de seguridad de la información necesarias para proteger sus componentes (datos, modelos, infraestructura) contra la alteración, el acceso no autorizado o la interrupción del servicio.

7.3. Principio de Transparencia y Suministro de Información

Tanto el Reglamento de IA como el RGPD imponen estrictas obligaciones de transparencia para permitir que las personas usuarias comprendan e interactúen con los sistemas de IA de forma informada. Este principio garantiza que se suministre información clara y accesible sobre el funcionamiento y las limitaciones del sistema, lo que implica que:

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

- Los sistemas deben estar diseñados para informar a las personas usuarias finales de que están interactuando con una IA.
- Las instrucciones de uso deben proporcionar información concisa y comprensible sobre la identidad del proveedor, la finalidad prevista y las características del sistema.
- Se debe informar a las personas usuarias sobre las limitaciones del sistema, el nivel de precisión esperado y las condiciones bajo las cuales puede generar resultados incorrectos.

7.4. Principio de No Discriminación y Equidad

En consonancia con la Carta de los Derechos Fundamentales de la UE y los requisitos del Reglamento de IA, se deben adoptar todas las medidas necesarias para minimizar el riesgo de discriminación algorítmica. Este principio exige que los sistemas de IA se desarrollen y utilicen de manera que no perpetúen ni amplifiquen sesgos que conduzcan a resultados perjudiciales para personas o grupos protegidos. En el ámbito sanitario, esto implica una vigilancia activa para garantizar la equidad en el diagnóstico y el acceso a los servicios de salud, corrigiendo los sesgos detectados tanto en los datos como en el comportamiento del modelo en producción.

7.5. Principio de Privacidad y Gobernanza de los Datos

El Reglamento General de Protección de Datos (RGPD) y el propio Reglamento de IA establecen un marco estricto para la privacidad y la gobernanza de los datos, especialmente en el tratamiento de datos de salud. Este principio exige el cumplimiento riguroso de las siguientes obligaciones durante todo el ciclo de vida del sistema:

- **Licitud, lealtad y transparencia:** El tratamiento debe tener una base jurídica válida y ser comunicado a los interesados.
- **Limitación de la finalidad:** Los datos solo pueden ser utilizados para los fines específicos, explícitos y legítimos para los que fueron recogidos.
- **Minimización de datos**: Solo se deben tratar los datos estrictamente necesarios para cumplir con dicha finalidad.
- **Integridad y confidencialidad:** Se deben aplicar las medidas de seguridad técnicas y organizativas adecuadas para proteger los datos contra el tratamiento no autorizado, la pérdida o la destrucción.

7.6. Principio de Responsabilidad y Rendición de Cuentas

El principio de responsabilidad proactiva, fundamental tanto en el RGPD como en el marco de gobernanza del Reglamento de IA, exige la existencia de una atribución clara de responsabilidades en todo el ciclo de vida del sistema. Las organizaciones deben poder demostrar el cumplimiento de sus obligaciones legales y éticas, lo que se materializa a través de un marco de gobernanza que asigna roles, define políticas de uso, establece procedimientos de gestión de riesgos y garantiza la capacidad de responder por las decisiones y los resultados del sistema.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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

7.7. Principio de Trazabilidad y Calidad de los Registros

Para posibilitar la supervisión, la investigación de incidentes y la rendición de cuentas, el Reglamento de IA exige que los sistemas de alto riesgo dispongan de capacidades de registro automático de eventos (logs). Este principio de trazabilidad establece que dichos registros deben ser seguros, inmutables y suficientes para garantizar la reconstrucción de las operaciones del sistema. Su finalidad es permitir:

- Monitorizar el funcionamiento del sistema y verificar el cumplimiento de sus requisitos.
- Investigar incidentes graves y eventos inesperados a posteriori.
- Reconstruir la secuencia de operaciones que llevaron a un resultado específico.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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. Tipos de personas usuarias finales en sistemas de Inteligencia Artificial en salud

El uso seguro y responsable de la IA en salud requiere la participación coordinada de distintos perfiles de personas usuarias finales, cada uno con responsabilidades y capacidades específicas. Esta guía identifica tres tipologías principales, a las que se dirigen las recomendaciones operativas de los capítulos siguientes.

Este bloque establece las condiciones operativas necesarias para garantizar un uso seguro, responsable y conforme a los sistemas de Inteligencia Artificial en el ámbito sanitario, atendiendo a los distintos perfiles de responsabilidad que intervienen en su utilización durante la fase de operación del sistema.

8.1. Personal sanitario

El personal sanitario (médicos, personal de enfermería, farmacéuticos, etc.) constituye el conjunto de personas usuarias clínicas operativas de los sistemas de IA. Su interacción con la tecnología tiene un impacto directo en la atención clínica y en la toma de decisiones que afectan a la salud de los pacientes. Su interacción con la IA tiene un impacto directo en la atención clínica y en la seguridad del paciente.

Sus responsabilidades en el contexto del uso de la IA incluyen:

- Interpretar críticamente los resultados generados por los sistemas de IA (predicciones, alertas, clasificaciones), contrastándolos siempre con su propio juicio clínico, la historia del paciente y otras fuentes de información.
- Integrar las recomendaciones algorítmicas en el contexto global del paciente, considerando factores que el sistema puede no conocer.
- Ejercer la supervisión humana efectiva, decidiendo de forma justificada si acepta, modifica o descarta la salida del sistema.
- Detectar y notificar resultados incoherentes, errores, posibles sesgos o cualquier comportamiento anómalo del sistema que pueda comprometer la seguridad del paciente.
- Informar al paciente de manera clara y comprensible sobre el uso de herramientas de IA en su proceso asistencial, cuando corresponda.

Aunque no es responsable del diseño técnico o la configuración de los sistemas, el personal sanitario es el garante último de que la tecnología se utilice de forma segura y beneficiosa en la práctica clínica.

8.2. Personal de Tecnologías de la Información y Sistemas de Información Sanitaria

El personal de TI es responsable de la infraestructura tecnológica, la seguridad y la integración operativa de los sistemas de IA dentro del ecosistema digital de la organización sanitaria. Su actuación es fundamental para garantizar la confidencialidad, integridad y disponibilidad de los sistemas y de los datos de salud que procesan.

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 el contexto de esta guía, se consideran personas usuarias finales porque interactúan directamente con los sistemas desde un punto de vista técnico y operativo. Sus responsabilidades clave abarcan:

- Gestionar los accesos y permisos a las plataformas de IA, aplicando el principio de mínimo privilegio.
- Asegurar la infraestructura de red, implementando medidas como la segmentación para aislar los sistemas de IA críticos.
- Monitorizar el rendimiento, la disponibilidad y los eventos de seguridad de los sistemas, integrando sus registros (logs) en las herramientas corporativas (p. ej., SIEM).
- Gestionar el ciclo de vida de las actualizaciones de software y modelos en coordinación con los proveedores y los equipos clínicos.
- Actuar como primera línea de respuesta técnica ante incidentes de seguridad o fallos operativos del sistema.

Su rol es esencial para mantener un entorno tecnológico seguro y resiliente que permita el funcionamiento fiable de los sistemas de IA

8.3. Personas usuarias finales, pacientes y ciudadanía

Este grupo constituye las personas usuarias finales no profesionales de los sistemas de IA, interactuando con ellos a través de servicios sanitarios digitales como portales del paciente, aplicaciones móviles, asistentes virtuales (chatbots) de triaje o sistemas de información automatizada. Aunque su interacción no es técnica, sus acciones y su comprensión de la tecnología son cruciales para la seguridad y la eficacia del proceso.

Sus responsabilidades se centran en un uso informado y seguro de las herramientas proporcionadas:

- Utilizar exclusivamente los canales y aplicaciones oficiales proporcionados por el sistema de salud para cualquier interacción que implique datos de salud.
- Proteger sus credenciales de acceso (personas usuarias y contraseñas) para evitar accesos no autorizados a su información personal.
- Interpretar la información generada por la IA (p. ej., una evaluación de síntomas por un chatbot) como una orientación, nunca como un diagnóstico definitivo, y consultar siempre a un profesional sanitario ante cualquier duda sobre su salud.
- Comunicar a través de los canales de atención a la persona usuaria cualquier funcionamiento anómalo, información incorrecta o respuesta inadecuada que detecten en las herramientas de IA.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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. Condiciones de uso responsable y seguro de sistemas de IA por parte del personal sanitario

El personal sanitario actúa como persona usuaria clínica operativa de los sistemas de Inteligencia Artificial y constituye el último y más importante nivel de control humano antes de que los resultados generados por la IA impacten en la atención al paciente. Por este motivo, su actuación es determinante para materializar los beneficios de la tecnología y, a la vez, mitigar sus riesgos. Las siguientes recomendaciones establecen las condiciones de uso seguro y responsable para este perfil.

9.1. Gestión de Cuentas y Credenciales

El acceso a los sistemas de información clínica que integran IA debe ser tratado con la máxima diligencia para proteger la confidencialidad de los datos del paciente.

- Utilización de cuentas personales e intransferibles: Cada profesional debe acceder a los sistemas con una cuenta individual. El uso de cuentas compartidas se desaconseja, ya que dificulta la trazabilidad de las acciones y la correcta atribución de responsabilidades.
- Protección de credenciales: Las contraseñas son estrictamente confidenciales y no deben ser compartidas, anotadas en lugares visibles ni reutilizadas en servicios externos.
- Bloqueo de la sesión de trabajo: Al ausentarse del puesto, es imperativo bloquear la sesión para impedir accesos no autorizados a la información clínica.
- Notificación de anomalías: Se debe informar inmediatamente al servicio de soporte de TI ante la sospecha de que una cuenta ha sido comprometida o si se detectan accesos no autorizados.
- Autenticación multifactor: Como medida adicional de protección, es conveniente implantar autenticación multifactor (MFA) en los sistemas de IA en salud que refuercen las contraseñas tradicionales con medidas adicionales de seguridad como certificados electrónicos, claves FIDO/passkeys u otros mecanismos que añadan un nuevo nivel de seguridad.

9.2. Uso Conforme a la Finalidad Prevista

La seguridad y eficacia de un sistema de IA están estrictamente ligadas a su finalidad validada.

Empleo exclusivo de sistemas autorizados: Se deben emplear únicamente los sistemas de IA que hayan sido formalmente autorizados por la organización e integrados en las plataformas corporativas.

Respeto a las condiciones de uso: Cada herramienta de IA debe utilizarse solo para la población de pacientes, las patologías y los escenarios clínicos para los que ha sido validada. El uso del sistema "fuera de finalidad" (off-label) genera resultados no fiables y puede comprometer la seguridad del paciente.

Conocimiento de las limitaciones del sistema: Antes de utilizar una herramienta de IA, es necesario informarse sobre sus limitaciones conocidas (p. ej., si no está validada para pacientes pediátricos, si requiere una calidad de imagen específica o si su rendimiento es menor en ciertas comorbilidades).

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

9.3. Supervisión Humana y Contención del Sesgo de Automatización

La supervisión humana es la salvaguarda principal contra los errores algorítmicos.

- La IA como herramienta de apoyo: El resultado de la IA (una predicción, una alerta, una clasificación) es un elemento más de información, no una decisión clínica. La responsabilidad final de la decisión recae siempre en el profesional.
- Aplicación de un criterio clínico crítico: No se deben aceptar de forma acrítica las recomendaciones del sistema. Es fundamental contrastarlas siempre con el juicio clínico propio, el contexto del paciente, su historia clínica y otras pruebas diagnósticas.
- Atención especial a los casos atípicos: Los modelos de IA son entrenados con patrones comunes y pueden presentar un rendimiento inferior ante presentaciones clínicas inusuales o pacientes con perfiles complejos. En estos casos, el juicio clínico es aún más indispensable.

9.4. Detección y Notificación de Errores o Sesgos del Modelo

El personal sanitario es la primera línea de defensa para detectar fallos de rendimiento en el entorno real.

- Vigilancia de patrones de error: La observación de resultados incorrectos o sesgados de forma sistemática para un determinado perfil de paciente (p. ej., por sexo, edad, etnia) puede ser un indicio de sesgo algorítmico o de deriva del modelo (model drift).
- Notificación de todas las incoherencias: Se debe comunicar a través de los canales de notificación de incidentes de la organización cualquier resultado que se considere incoherente, contraintuitivo o clínicamente inverosímil, aunque no haya causado un daño al paciente. Dicha notificación constituye información vital para la monitorización y mejora continua del sistema.

9.5. Gestión de la Confidencialidad y el Riesgo de "Shadow AI"

La protección de los datos del paciente es una obligación legal y ética fundamental.

- PROHIBICIÓN del uso de herramientas de IA públicas con datos de pacientes: Debido a la regulación actual, no se deben introducir, copiar, pegar o subir cualquier tipo de información que pueda identificar a un paciente (textos de historias clínicas, resultados de pruebas, imágenes) en herramientas de IA de acceso público o no autorizadas por la organización (como ChatGPT, Gemini, Copilot u otras).
- Conciencia sobre el riesgo de fuga y corrupción de datos: El uso de "Shadow AI" supone una grave brecha de confidencialidad y una violación del RGPD. Además, estos sistemas pueden generar información falsa ("alucinaciones") que, si se incorpora a la historia clínica, puede corromper el registro y llevar a futuras decisiones erróneas.

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

9.6. Procedimiento de Actuación ante un Fallo Crítico del Sistema

- Priorización de la seguridad del paciente: Si un sistema de IA que apoya una decisión crítica falla o se vuelve inoperativo, se deben activar inmediatamente los procedimientos clínicos estándar y los planes de contingencia definidos. La atención al paciente no puede demorarse por un fallo tecnológico.
- Abstención de intervenciones no autorizadas: Se debe evitar reiniciar equipos o modificar configuraciones sin la debida autorización.
- Comunicación urgente de la incidencia: Es preciso informar inmediatamente al servicio de soporte de TI o al responsable designado, proporcionando toda la información posible sobre el fallo para facilitar una respuesta rápida.

9.7. Información y Comunicación con el Paciente

- Transparencia sobre el uso de la IA: Cuando sea relevante para el proceso asistencial, se debe informar al paciente de forma clara y comprensible de que se están utilizando herramientas de IA como apoyo a la decisión clínica.
- Gestión de las expectativas del paciente: Es importante explicar que la IA es una tecnología de soporte que ayuda al equipo médico a tomar las mejores decisiones, pero no es un sistema infalible ni un sustituto del diagnóstico profesional. Esto contribuye a construir confianza y a evitar malentendidos.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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

10. Condiciones de uso seguro de sistemas de IA por parte del personal de Tecnologías de la Información

El personal de Tecnologías de la Información actúa como responsable técnico y operativo de los El personal de Tecnologías de la Información (TI) es el responsable de la infraestructura tecnológica, la ciberseguridad y la integración operativa de los sistemas de IA. Su actuación es fundamental para garantizar un entorno digital seguro, resiliente y fiable, que proteja la confidencialidad, integridad y disponibilidad tanto de los sistemas de IA como de los datos de salud que procesan.

10.1. Gestión de Accesos y Privilegios

El control de acceso es la primera línea de defensa para proteger los sistemas de IA y sus datos.

- Aplicación del principio de mínimo privilegio: Se deben asignar permisos de acceso a las plataformas de IA, consolas de administración y APIs basándose estrictamente en el rol y las funciones de la persona usuaria. Los accesos deben ser revocados inmediatamente cuando dejen de ser necesarios.
- Uso de cuentas nominales: Cada administrador o persona usuaria técnica debe tener una cuenta individual para garantizar la trazabilidad de las acciones. Está prohibido el uso de cuentas compartidas.
- Implementación de autenticación multifactor (MFA): Como medida adicional de protección, es conveniente implantar autenticación multifactor (MFA) en los sistemas de IA en salud que refuercen las contraseñas tradicionales con medidas adicionales de seguridad como certificados electrónicos, claves FIDO/passkeys u otros mecanismos que añadan un nuevo nivel de seguridad.

10.2. Seguridad de la Infraestructura y las Comunicaciones

La infraestructura que soporta la IA debe ser tratada como un activo crítico.

- Segmentación de la red: Es necesario aislar los servidores y componentes que ejecutan los modelos de IA en un segmento de red específico y protegido. Las reglas del cortafuegos (firewall) deben restringir la comunicación únicamente a los sistemas y servicios autorizados, denegando todo lo demás por defecto.
- Aplicación de configuraciones de seguridad robustas (Hardening): Se deben configurar de forma segura los sistemas operativos, contenedores y servicios que alojan la IA, desactivando todos los puertos, protocolos y servicios no esenciales para reducir la superficie de ataque.
- Cifrado de las comunicaciones: Es preciso asegurar que todo el tráfico de red hacia y desde el sistema de IA, incluidas las llamadas a las APIs, se realice a través de protocolos de comunicación cifrados y seguros.

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

10.3. Gestión de Vulnerabilidades y Amenazas Específicas de IA

Además de las vulnerabilidades tradicionales, se deben considerar las amenazas específicas del ámbito de la IA.

- Realización de análisis de vulnerabilidades periódicos: Es necesario escanear de forma regular el software base, las librerías de IA y las dependencias del sistema para detectar y parchear vulnerabilidades conocidas.
- Protección de la cadena de suministro de IA (MLOps): Se deben asegurar los repositorios de código y los artefactos de los modelos para prevenir su manipulación no autorizada. La integridad de los modelos pre-entrenados y los conjuntos de datos de fuentes externas debe ser verificada antes de su uso.
- Mitigación de los riesgos de ataques de inferencia: Se deben implementar, en la medida de lo posible, controles para proteger los sistemas de IA de alto riesgo contra ataques como los adversarios o de extracción de modelos, por ejemplo, mediante la monitorización de patrones de consulta anómalos.

10.4. Gestión de Actualizaciones y Control de Versiones

El comportamiento de un sistema de IA puede cambiar drásticamente con una actualización.

- Implementación de un control de cambios formal: Cualquier actualización del modelo, del software o de la configuración debe seguir un procedimiento formal de gestión de cambios que incluya una evaluación de impacto.
- Uso de un control de versiones riguroso: Es necesario mantener un registro versionado no solo del código, sino también de los modelos de IA y de los conjuntos de datos para garantizar la reproducibilidad.
- Validación de actualizaciones en un entorno de pre-producción: Se debe probar siempre cualquier cambio significativo en un entorno de pruebas aislado y representativo antes de su despliegue en el entorno clínico real.
- Disponibilidad de un plan de reversión (rollback): Es preciso disponer de un procedimiento preparado y probado para revertir de forma rápida y segura a la versión anterior del sistema en caso de que una actualización cause un incidente.

10.5. Monitorización, Registros y Trazabilidad

La monitorización es clave para cumplir con el principio de trazabilidad.

- Configuración de un registro de auditoría exhaustivo: Se debe asegurar que los sistemas de IA generen logs de todos los eventos relevantes: peticiones de inferencia, resultados, accesos de personas usuarias, errores y cambios de configuración.
- Centralización de registros en un SIEM: Es necesario enviar y almacenar de forma segura los logs en la herramienta corporativa de gestión de eventos e información de seguridad (SIEM) para permitir la correlación de eventos y la investigación de incidentes.

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

- Monitorización del rendimiento y la disponibilidad: Se deben implementar alertas automáticas que notifiquen al equipo de TI sobre caídas del servicio, degradación del rendimiento (p. ej., aumento de la latencia) o picos de error.

10.6. Continuidad de Negocio y Recuperación ante Desastres

Los sistemas de IA críticos deben estar incluidos en los planes de continuidad de la organización.

- Inclusión de los activos de IA en las copias de seguridad: Se debe asegurar que los modelos de IA, las configuraciones críticas y los datos asociados formen parte de la política de copias de seguridad (backup) de la organización.
- Integración de la IA en los planes de recuperación (DRP): Los planes de recuperación ante desastres deben contemplar los pasos necesarios para restaurar los servicios de IA en un tiempo y punto de recuperación objetivos (RTO/RPO) definidos.
- Participación en las pruebas de los planes: Es necesario verificar periódicamente, mediante simulacros, que los procedimientos de restauración de los sistemas de IA funcionan correctamente.

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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

11. Condiciones de uso seguro de sistemas de IA por parte de pacientes y ciudadanía

Los pacientes, personas cuidadoras y la ciudadanía en general interactúan con sistemas de IA a través de servicios sanitarios digitales. Un uso informado de estas herramientas es crucial para la protección de la información personal y la eficacia del proceso asistencial.

11.1. Uso de Canales Oficiales y Protección de Credenciales

La seguridad de la información de salud depende del uso de canales seguros y de la protección de los accesos personales.

- Utilización de servicios autorizados: Se debe acceder a la información de salud y a los servicios digitales únicamente a través de los portales web y las aplicaciones móviles oficiales proporcionados por el sistema de salud. Es fundamental desconfiar de enlaces recibidos por correo electrónico o SMS no solicitados.
- Protección de las credenciales de acceso: La cuenta de acceso y la contraseña del portal de salud son personales e intransferibles. Su protección es responsabilidad del titular, quien debe emplear contraseñas robustas y evitar su reutilización en otros servicios de internet para prevenir accesos no autorizados.
- Canal de comunicación de brecha de seguridad: Deberán existir canales de comunicación a través de los cuales se pueda notificar potenciales brechas de seguridad de expongan las credenciales de acceso, a fin de poder reportar de formar ágil y eficaz cualquier potencial incidente que afecte a los sistemas de IA en salud.

11.2. Interpretación de la Información Generada por IA

Es fundamental comprender el rol y las limitaciones de la información proporcionada por un sistema automático para un uso seguro.

- La IA como herramienta de orientación: Cuando una aplicación o un asistente virtual evalúa síntomas u ofrece una recomendación, el resultado debe ser entendido como una orientación basada en algoritmos, y no constituye un diagnóstico médico.
- Prevalencia del criterio profesional: La información proporcionada por un sistema de IA nunca sustituye la consulta, el juicio clínico y el diagnóstico de un profesional de la salud. Ante cualquier duda o preocupación sobre el estado de salud, es imprescindible contactar con el centro sanitario de referencia.

11.3. Derechos y Participación en la Mejora del Sistema

Los ciudadanos tienen un papel activo en la supervisión y mejora de los servicios digitales.

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

- Derecho a la revisión por una persona: Las decisiones con un impacto significativo sobre la salud no deben ser tomadas de forma exclusivamente automática. La normativa vigente reconoce el derecho a solicitar que un profesional cualificado revise cualquier decisión basada en un tratamiento automatizado.
- Notificación de incidencias: En caso de detectar que la información proporcionada por un sistema automático es incorrecta, inadecuada o confusa, o si una aplicación presenta un funcionamiento anómalo, se recomienda comunicarlo a través de los canales de atención a la persona usuaria o al paciente. Dicha comunicación es valiosa para la detección de errores y la mejora continua de los servicios

12. Gestión de incidencias relacionadas con el uso de sistemas de IA en Salud

12.1. Finalidad y alcance de la gestión de incidencias en sistemas de IA

La gestión de incidencias relacionadas con el uso de sistemas de Inteligencia Artificial en el ámbito sanitario constituye un elemento esencial para garantizar la seguridad del paciente, la protección de los datos de salud, la continuidad asistencial y la confianza en los servicios públicos digitales. Dada la naturaleza inferencial y parcialmente autónoma de los sistemas de IA, las incidencias no se limitan a fallos técnicos evidentes, sino que pueden manifestarse a través de resultados incoherentes, comportamientos inesperados, degradaciones progresivas del rendimiento o usos indebidos por parte de las personas usuarias.

Esta sección establece los principios y criterios operativos que deben regir la detección, comunicación, análisis y tratamiento de incidencias asociadas al uso de sistemas de IA en salud, con independencia de su origen técnico, organizativo o humano. Su objetivo no es sustituir los procedimientos generales de gestión de incidentes de la entidad sanitaria, sino complementarlos, incorporando las especificidades propias de la IA y reforzando la coordinación entre perfiles.

La gestión eficaz de incidencias en sistemas de IA no debe concebirse como una actividad reactiva o excepcional, sino como un proceso continuo de aprendizaje y mejora que contribuye a reducir riesgos futuros y a fortalecer la gobernanza de la IA en el sistema sanitario.

12.2. Tipología de incidencias relacionadas con el uso de la IA

Las incidencias asociadas al uso de sistemas de Inteligencia Artificial en salud pueden adoptar múltiples formas y presentar impactos diversos. No todas las incidencias implican necesariamente un fallo técnico del sistema, pero todas requieren una atención adecuada cuando pueden afectar a la seguridad del paciente, a la calidad del servicio o a los derechos de las personas.

En el contexto de esta guía, se consideran incidencias relevantes aquellas situaciones en las que el uso del sistema de IA produce, o puede producir, efectos adversos no previstos, resultados manifiestamente incoherentes, riesgos para la confidencialidad de los datos, interrupciones significativas del servicio o vulneraciones de los principios de uso responsable definidos en las secciones anteriores.

Esta aproximación amplia permite incluir tanto incidentes técnicos clásicos como situaciones derivadas de errores de uso, falta de supervisión humana efectiva, desviaciones respecto a la finalidad autorizada o interacciones problemáticas entre el sistema y su contexto operativo.

12.3. Detección temprana de incidencias y papel de las personas usuarias finales

La detección temprana de incidencias es un factor clave para limitar su impacto y evitar daños mayores. En este ámbito, las personas usuarias finales —personal sanitario, personal de TI 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

desempeñan un papel fundamental como primer nivel de observación de comportamientos anómalos o resultados inesperados.

El personal sanitario puede detectar incidencias a través de incoherencias clínicas, discrepancias entre el resultado de la IA y la valoración profesional, o cambios no explicables en el comportamiento del sistema. El personal de TI puede identificar fallos operativos, degradaciones del rendimiento, anomalías en los flujos de datos o indicios de incidentes de seguridad. Por su parte, los pacientes y la ciudadanía pueden advertir errores en la información proporcionada, dificultades de acceso o respuestas automatizadas inapropiadas.

El uso responsable de la IA exige que estos indicios no se ignoren ni se normalicen, y que las personas usuarias finales dispongan de canales claros y accesibles para comunicar las incidencias detectadas, incluso cuando no exista certeza absoluta sobre su gravedad.

12.4. Comunicación y registro de incidencias

Toda incidencia relacionada con el uso de sistemas de IA en salud deberá ser comunicada y registrada conforme a los procedimientos establecidos por la entidad sanitaria. La comunicación temprana y estructurada es esencial para permitir una evaluación adecuada del impacto, coordinar la respuesta y garantizar la trazabilidad de las actuaciones realizadas.

El registro de incidencias deberá permitir identificar, al menos, el sistema de IA afectado, el contexto en el que se produjo la incidencia, los efectos observados, las personas implicadas y las medidas adoptadas de forma inmediata. Esta información es clave para el análisis posterior, la mejora continua y, en su caso, la rendición de cuentas ante órganos internos o externos.

La gestión responsable de incidencias exige evitar enfoques punitivos o disuasorios que puedan desalentar la comunicación de problemas, y fomentar una cultura organizativa orientada al aprendizaje y a la mejora de la seguridad en el uso de la IA.

12.5. Evaluación del impacto y priorización de la respuesta

Una vez comunicada la incidencia, la entidad sanitaria deberá evaluar su impacto potencial o real sobre la seguridad del paciente, la protección de los datos de salud y la continuidad asistencial. Esta evaluación debe realizarse de manera coordinada, teniendo en cuenta tanto los aspectos técnicos como los clínicos y organizativos.

En aquellos casos en los que exista un riesgo para la seguridad de las personas o para derechos fundamentales, la respuesta deberá priorizar la mitigación inmediata del riesgo, incluso mediante la suspensión temporal del uso del sistema de IA o la activación de procedimientos alternativos. La continuidad del servicio nunca debe anteponerse a la seguridad del paciente cuando exista una duda razonable sobre el comportamiento del sistema.

La evaluación del impacto permitirá, además, determinar si la incidencia requiere la activación de otros procedimientos, como la notificación a autoridades competentes, la revisión del sistema por parte de equipos especializados o la adopción de medidas correctoras a mayor escala.

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

12.6. Medidas correctoras y aprendizaje organizativo

La gestión de incidencias relacionadas con la IA no se agota en la resolución inmediata del problema detectado. Un elemento esencial de este proceso es la identificación de las causas subyacentes y la adopción de medidas correctoras que eviten la repetición de situaciones similares en el futuro.

Estas medidas pueden incluir ajustes en los procedimientos de uso, refuerzo de la formación de las personas usuarias, mejoras en la supervisión humana, modificaciones en la configuración operativa del sistema o, en casos más graves, la revisión de la idoneidad del propio sistema de IA para el contexto en el que se utiliza.

El aprendizaje organizativo derivado de la gestión de incidencias constituye un componente clave de la mejora continua y debe integrarse en los mecanismos de gobierno y gestión de riesgos de la IA definidos por la entidad sanitaria.

12.7. Coordinación con el modelo de gobierno de la IA y la gestión de riesgos de IA

La gestión de incidencias relacionadas con el uso de sistemas de IA en salud debe realizarse de forma coherente con el modelo de gobierno de la IA definido por la entidad sanitaria y con las directrices establecidas otras de guías del proyecto.

En particular, las incidencias relevantes deberán ser comunicadas a los órganos o roles responsables de la supervisión de la IA conforme a lo establecido en la guía IA-01 – Modelo de gobierno de la Inteligencia Artificial, y su análisis deberá alimentarse de la metodología de gestión de riesgos definida en la guía IA-02. Cuando proceda, la clasificación del sistema conforme al Reglamento de IA y las obligaciones asociadas deberán ser revisadas a la luz de las incidencias detectadas.

Esta coordinación garantiza que la gestión de incidencias no se limite a una respuesta aislada, sino que contribuya de manera efectiva a la mejora global del uso de la IA en el sistema sanitario.

13. Referencias

- Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (AI Act). https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32024R1689

- Reglamento (UE) 2016/679 (RGPD), de 27 de abril de 2016, relativo a la protección de datos personales. https://eur-lex.europa.eu/eli/reg/2016/679/oj

- Directiva (UE) 2022/2555 (NIS2), relativa al nivel elevado común de ciberseguridad en la Unión. https://eur-lex.europa.eu/eli/dir/2022/2555/oj

- Reglamento (UE) 2022/868 (Data Governance Act) sobre la gobernanza europea de datos. https://eur-lex.europa.eu/eli/reg/2022/868/oj

- Reglamento (UE) 2022/2065 (Digital Services Act – DSA) sobre servicios digitales. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

- Ley Orgánica 3/2018 (LOPDGDD), de Protección de Datos Personales y garantía de los derechos digitales. https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673

- Esquema Nacional de Seguridad (ENS) – Real Decreto 311/2022, de 3 de mayo. https://www.boe.es/eli/es/rd/2022/05/03/311

- Carta de Derechos Digitales (Gobierno de España, 2021). https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/Carta_Derechos_Digitales.pdf

- NIST AI Risk Management Framework 1.0 (2023). https://www.nist.gov/itl/ai-risk-management-framework

- ISO/IEC 42001:2023 – AI Management System. https://www.iso.org/standard/81230.html

- ISO/IEC 23894:2023 – AI Risk Management. https://www.iso.org/standard/77304.html

- ISO/IEC 38507:2022 – Governance implications of AI. https://www.iso.org/standard/80382.html

- ISO/IEC 27001:2022 – Information security management systems.

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

https://www.iso.org/standard/82875.html

- ISO 31000:2018 – Risk management guidelines.
  https://www.iso.org/standard/65694.html

- UNE-EN 301549:2022 – Accesibilidad de productos y servicios TIC.
  https://www.aenor.com/norma-une-en-301549-2022-n0062490

Financiado por la Unión Europea NextGenerationEU
GOBIERNO DE ESPAÑA MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PUBLICA 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

14. Otras referencias de interés

El presente apartado recoge un resumen de los principales temas abordados en la guía, así como la relación de productos y servicios mencionados en ella. Su finalidad es ofrecer una visión global de los ámbitos tratados y facilitar la identificación de posibles referencias o elementos relevantes para futuras consultas o ampliaciones documentales.

14.1. Lista de materias tratadas en la guía

- Uso responsable y seguro de sistemas de Inteligencia Artificial en el ámbito sanitario público
- Objetivo, alcance y exclusiones de la guía en fase de uso operativo
- Marco normativo aplicable: Reglamento (UE) 2024/1689, RGPD, LOPDGDD, NIS2, ENS, MDR e IVDR
- Conceptos operativos y definiciones clave: supervisión humana efectiva, trazabilidad, finalidad prevista, deriva y sesgos
- Tipologías de sistemas de IA en salud: apoyo clínico, gestión sanitaria, interacción con ciudadanía e IA generativa con impacto asistencial
- Riesgos técnicos: calidad de datos, degradación del rendimiento en producción, vulnerabilidades e integraciones
- Riesgos operativos: sesgo de automatización, fatiga por alertas, uso fuera de finalidad y Shadow AI
- Riesgos estratégicos, éticos y sociales: equidad, transparencia, auditabilidad, confianza institucional y rendición de cuentas
- Riesgos de confidencialidad y protección de datos: minimización, limitación de finalidad, transferencias no autorizadas y reidentificación
- Riesgos de disponibilidad y continuidad asistencial: dependencia tecnológica, degradación silenciosa y planes de contingencia
- Riesgos para la seguridad del paciente: falsos negativos y positivos, recomendaciones no aplicables y errores de priorización
- Principios operativos de uso responsable: seguridad del paciente, supervisión humana, transparencia, integridad, disponibilidad y no discriminación
- Perfiles de personas usuarias finales: personal sanitario, personal de TI, pacientes y ciudadanía
- Condiciones de uso responsable y seguro por perfil de persona usuaria final
- Gestión de incidencias: detección temprana, comunicación, registro, evaluación de impacto, medidas correctoras y aprendizaje organizativo
- Coordinación con gobernanza institucional y gestión de riesgos en la organización sanitaria

14.2. Lista de productos mencionados en la guía

No se mencionan productos específicos en la guía.

14.3. Lista de servicios mencionados en la guía

No se mencionan productos específicos en la guía.

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

Uso responsable y seguro de sistemas de IA en Salud para personas usuarias finales

Febrero de 2026

Versión 1.1

Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia