Accelerating the digital transformation of the health sector in the Americas

AI-GUARD Tool

Artificial Intelligence Governance & Use Assessment for Responsible Deployment

PAHO
Pan American Health Organization
World Health Organization
Americas Region
IDB
Inter-American Development Bank

Herramienta AI-GUARD

Inteligencia artificial
Evaluación de la gobernanza y el uso para un despliegue responsable

Washington, D.C., 2026

OPS Organización Panamericana de la Salud Organización Mundial de la Salud Región de las Américas BID

Índice

Agradecimientos

Esta publicación se ha elaborado con el apoyo de las siguientes personas y socios:

Organización Panamericana de la Salud
Marcelo D'Agostino, Myrna Marti, Sebastián García-Saiso, Juan Carlos Díaz, Jaime Pedrosa Comino, João Paulo Souza, Marcos Luis Mori, María Alejandra Farias, Francisco Barbosa Junior.

Banco Interamericano de Desarrollo
Jennifer Nelson, Pablo Oreficce.

Hospital Italiano, Buenos Aires, Argentina, Centro Colaborador de la OPS/OMS para Sistemas de Información y Salud Digital
Daniel Luna, Fernando Plazzotta.

Prólogo

La inteligencia artificial está transformando rápidamente los sistemas de salud en todas las Américas. Desde el fortalecimiento de la vigilancia de enfermedades y la detección temprana de brotes hasta la mejora de la prestación de servicios, la optimización de la asignación de recursos, la mejora del apoyo al diagnóstico, el fortalecimiento de la investigación y la racionalización de los procesos administrativos, las tecnologías de IA ofrecen importantes oportunidades para mejorar los resultados de salud pública. Cuando se alinea con prioridades de salud claramente definidas, la IA puede contribuir a reducir los tiempos de espera, mejorar el acceso a servicios especializados, fortalecer el desempeño de la atención primaria de salud, mejorar la planificación de la fuerza laboral en salud y respaldar las decisiones de política basadas en datos. Sin embargo, la integración de la IA en los sistemas de salud no está exenta de complejidad. La implementación de la IA puede plantear retos en materia de gobernanza, equidad, transparencia y rendición de cuentas si no se lleva a cabo con las salvaguardias adecuadas. Pueden surgir riesgos derivados de una supervisión insuficiente, estructuras de rendición de cuentas poco claras, datos de entrenamiento sesgados, la falta de evaluación del desempeño de subgrupos, una transparencia limitada respecto a las limitaciones del sistema o un seguimiento inadecuado una vez implementado. Los sistemas de IA de gran impacto, en particular aquellos que influyen en el diagnóstico, la elegibilidad o la asignación de recursos públicos, requieren mecanismos de gobernanza estructurados para prevenir consecuencias no deseadas y proteger a las poblaciones vulnerables.

Reconociendo tanto el potencial transformador como los retos de gobernanza que plantea la IA en el ámbito de la salud, se ha elaborado la «Evaluación de la gobernanza y el uso de la inteligencia artificial para una implementación responsable» (AI-GUARD), una herramienta de orientación técnica de la OPS y el BID destinada a apoyar a los Estados Miembros y a las instituciones de salud en la toma de decisiones estructuradas, transparentes y responsables en relación con la adopción, la adquisición, el desarrollo y la ampliación de la IA. El instrumento proporciona un marco escalonado y proporcional que adapta la intensidad de la supervisión al nivel de impacto y riesgo asociado a cada iniciativa.

AI-GUARD forma parte de un esfuerzo continuo de la OPS y el BID por fortalecer la capacidad de los Estados Miembros para adoptar de manera responsable la inteligencia artificial en el ámbito de la salud. A través de la cooperación técnica, el diálogo sobre políticas, las iniciativas de desarrollo de capacidades, las evaluaciones de preparación para la IA y el desarrollo de herramientas prácticas de implementación, la OPS, en colaboración con el BID, ha apoyado a los países en el avance de la transformación digital, salvaguardando al

mismo tiempo la equidad, la transparencia y la confianza pública. AI-GUARD complementa estos esfuerzos al ofrecer un instrumento de gobernanza estructurado que traduce los principios estratégicos en prácticas operativas de toma de decisiones.

AI-GUARD es también un componente operativo fundamental del Conjunto de herramientas de evaluación de la preparación para la IA de la OPS y el BID, y contribuye a un enfoque más amplio e integrado que ayuda a los países a evaluar su preparación para la adopción de la inteligencia artificial en los ámbitos de la gobernanza, los datos, la fuerza laboral y la tecnología. En este marco, AI-GUARD proporciona un instrumento práctico de apoyo a la toma de decisiones centrado específicamente en la evaluación de iniciativas individuales de IA, garantizando que las consideraciones sobre la preparación se traduzcan en decisiones de implementación concretas y basadas en el riesgo.

AI-GUARD está diseñado para complementar y fortalecer los marcos de gobernanza establecidos por la Resolución CD60.R9 de la OPS sobre la transformación digital de los sistemas de salud, y para alinearse con los avances normativos nacionales en toda la región. El instrumento promueve una gobernanza que sea proporcionada, basada en la evidencia y alineada con las prioridades nacionales de salud pública. Alienta a las instituciones a evaluar el valor estratégico junto con el grado de preparación, a incorporar una supervisión humana significativa en las decisiones respaldadas por la IA, a evaluar posibles sesgos e impactos diferenciales entre las poblaciones, y a establecer mecanismos de seguimiento que garanticen un desempeño sostenido a lo largo del tiempo.

AI-GUARD no sustituye a los marcos normativos nacionales, los procesos de revisión jurídica ni las normas de contratación pública. Más bien, refuerza los procesos institucionales de toma de decisiones al proporcionar una metodología estructurada para evaluar la preparación y las salvaguardias antes de la implementación. Al hacerlo, AI-GUARD apoya una innovación responsable que contribuye a la resiliencia del sistema de salud, protege la equidad, mejora la confianza pública y garantiza que el avance tecnológico se traduzca en un beneficio medible para la salud pública.

Estructura general de AI-GUARD

Ruta de decisión estructurada de AI-GUARD

- **1.** Definición de la iniciativa
- **2.** Evaluación del valor para la salud pública
- **3.** Clasificación de la exposición al riesgo
- **4.** Evaluación de los pilares de AI-GUARD por niveles
- **5.** Recomendación estructurada y medidas de seguridad

Los cuatro pilares fundamentales de AI-GUARD

Modelo de niveles de AI-GUARD

Nivel 1	Nivel 2	Nivel 3
Uso responsable de la IA	IA a nivel de programa	Gobernanza de la IA de alto impacto

- **Nivel 1.** Uso responsable de la IA para:
  - Uso de IA generativa
  - Automatización administrativa
  - Herramientas de flujo de trabajo interno

Objetivo: Facilitar un uso seguro de la IA en el día a día sin una carga burocrática excesiva.

- **Nivel 2.** IA a nivel de programa para:
  - Herramientas de puntuación de riesgos
  - Aplicaciones de uso público
  - Sistemas de apoyo clínico
  - Cuadros de mando de vigilancia

Objetivo: Proporcionar una evaluación estructurada de la preparación antes de la adopción o la fase piloto.

- **Nivel 3.** Gobernanza de IA de alto impacto para:
  - Detección de brotes a nivel nacional
  - IA diagnóstica
  - Sistemas de asignación de recursos
  - Modelos de elegibilidad

Objetivo: Garantizar una gobernanza responsable antes de una implementación de alto impacto.

Recomendación final:

- Proceder
- Autorización condicional temporal — Pendiente de verificación
- Reconsiderar / Rediseñar

Resumen ejecutivo

Objetivo

La inteligencia artificial (IA) influye cada vez más en la toma de decisiones en los sistemas de salud. Desde la automatización administrativa, la investigación y el apoyo a la toma de decisiones clínicas hasta la vigilancia nacional y la asignación de recursos, las tecnologías de IA ofrecen importantes oportunidades para mejorar la eficiencia, el acceso, la calidad y la preparación en materia de salud pública.

Al mismo tiempo, las iniciativas de IA pueden introducir riesgos relacionados con la gobernanza, la rendición de cuentas, los sesgos, la protección de datos, la sostenibilidad operativa y las consecuencias no deseadas, especialmente cuando se implementan a gran escala o afectan a poblaciones vulnerables.

AI-GUARD se ha desarrollado para ayudar a los responsables de la toma de decisiones en todos los niveles de las instituciones sanitarias a evaluar el grado de preparación y el perfil de riesgo de cualquier iniciativa de IA antes de su adopción, adquisición, desarrollo o ampliación.

AI-GUARD es un instrumento estructurado y por niveles diseñado para:

- Evaluar el valor estratégico para la salud pública de las iniciativas de IA.
- Identificar la exposición al riesgo y el impacto potencial.
- Evaluar la gobernanza institucional y la preparación operativa.
- Detectar y mitigar los sesgos y las preocupaciones relacionadas con la equidad.
- Fortalecer la transparencia, la rendición de cuentas y el despliegue responsable.

El instrumento promueve una toma de decisiones disciplinada y basada en la evidencia, al tiempo que permite la innovación en consonancia con las prioridades de salud pública.

Ámbito de aplicación

AI-GUARD debe aplicarse antes de:

- Adquirir productos o plataformas basados en IA.
- Desarrollar modelos de IA internamente.
- Ampliar iniciativas piloto de IA.

- Integrar la IA en los flujos de trabajo clínicos o de salud pública.
- Introducir herramientas generales de IA (por ejemplo, grandes modelos de lenguaje) en las operaciones institucionales.

El instrumento es aplicable en:

- Ministerios de Salud.
- Organismos de salud pública.
- Hospitales e instituciones clínicas.
- Programas nacionales de salud.
- Unidades de salud digital y sistemas de información.
- Organismos reguladores y de supervisión.

AI-GUARD está diseñado para adaptarse tanto a iniciativas de IA de bajo riesgo como de alto impacto mediante un modelo de evaluación por niveles.

Lo que AI-GUARD no es

AI-GUARD no es:

- Un mecanismo de aprobación reglamentaria.
- Una herramienta de certificación o acreditación.
- Un sustituto de la revisión jurídica, ética o de contratación pública.
- Una barrera para la innovación.

Más bien, es un instrumento de asesoramiento estructurado destinado a reforzar la preparación institucional y la toma de decisiones responsable antes de la implementación de la IA.

- **Transparencia:** Las pruebas, las limitaciones y las actualizaciones del sistema deben documentarse claramente.
- **Sostenibilidad:** La viabilidad operativa a largo plazo y el seguimiento deben planificarse antes de la implementación.

Estructura del instrumento

AI-GUARD funciona a través de un marco adaptativo por niveles:

1. **Análisis ejecutivo inicial:** una breve evaluación para determinar el valor estratégico, la exposición al riesgo y la preparación preliminar.

2. **Clasificación por niveles:** asignación automática a uno de los tres niveles en función del impacto y el riesgo:
    - Nivel 1: Uso responsable de la IA
    - Nivel 2: IA a nivel de programa
    - Nivel 3: gobernanza de la IA de alto impacto
3. **Evaluación de los cuatro pilares fundamentales:** gobernanza, control humano, medidas de protección contra riesgos y sesgos, y preparación para la implementación.
4. **Resumen del panel de control de AI-GUARD:** un informe estructurado que incluye índices y recomendaciones sobre los próximos pasos a seguir.

Esta estructura garantiza la coherencia entre las instituciones, al tiempo que mantiene la flexibilidad para diferentes niveles de complejidad.

Resultados esperados

La aplicación de AI-GUARD ayuda a las instituciones a:

- Mejorar la calidad de las decisiones relacionadas con la IA.
- Identificar las deficiencias en materia de gobernanza y preparación antes de la implementación.
- Reforzar las medidas de protección para los sistemas de alto impacto.
- Reducir la exposición a riesgos operativos, de reputación y relacionados con la equidad.
- Promover una innovación responsable, transparente y sostenible.

Herramienta AI-GUARD

Análisis ejecutivo de entrada

Valor estratégico

Problema de salud pública abordado

¿Cuál es el objetivo principal que persigue la iniciativa?

(Seleccione una opción)

- ☐ Eficiencia administrativa o mejora del flujo de trabajo (1)
- ☐ Reducción de los tiempos de espera o mejora del acceso a los servicios (2)
- ☐ Mejor uso de los datos sanitarios para la planificación, las decisiones políticas o la organización de los servicios (3)
- ☐ Refuerzo de la vigilancia o la detección precoz (3)
- ☐ Asignación estratégica de recursos u optimización del sistema (4)
- ☐ Objetivo exploratorio o indefinido (4)

Justificación del uso de la IA

¿Por qué se considera necesaria la inteligencia artificial para esta iniciativa?

(Seleccione una opción)

- ☐ Las herramientas digitales convencionales son insuficientes para el objetivo previsto (1)
- ☐ La IA puede mejorar la velocidad, la escala, la capacidad de predicción o la capacidad analítica (2)
- ☐ La IA es propuesta principalmente por un socio externo, un donante o un proveedor (3)
- ☐ Se ha considerado la IA sin una justificación comparativa clara (4)

Identificación de la iniciativa

¿Qué tipo de iniciativa de IA se está considerando?

(Seleccione una opción)

- ☐ Uso de una herramienta de IA generativa para apoyo operativo limitado (p. ej., modelo de lenguaje grande, redacción, resumen, automatización administrativa) (1)
- ☐ Adquisición de un producto o plataforma basada en IA de un proveedor externo (2)

- ☐ Ampliación, expansión institucional o implementación más amplia de una iniciativa de IA existente (3)
- ☐ Desarrollo interno de un modelo o sistema de IA (4)

Impacto en la toma de decisiones

¿Qué nivel de impacto en la toma de decisiones tiene la iniciativa?

(Seleccione una opción)

- ☐ No tiene impacto directo en la atención al paciente, las decisiones programáticas o la asignación de recursos (1)
- ☐ Apoya la toma de decisiones, pero no automatiza ni determina las decisiones finales (2)
- ☐ Influye en la priorización, la planificación o la asignación de recursos, o en las intervenciones (3)
- ☐ Influye directamente en el diagnóstico, la elegibilidad, la determinación de prestaciones o las decisiones críticas (4)

Impacto en la población

¿A quién afecta esta iniciativa?

(Seleccione una opción)

- ☐ Solo el personal interno o un número limitado de usuarios operativos internos (1)
- ☐ Una población de programa definida o un grupo específico de beneficiarios (2)
- ☐ Población general o varios grupos de población (3)
- ☐ Poblaciones vulnerables, desatendidas o afectadas de manera desproporcionada (4)

Preparación preliminar

Base empírica

¿Qué nivel de evidencia respalda la eficacia de esta iniciativa?

(Seleccione una opción)

- ☐ Evidencia de validación sólida (independiente, revisada por pares o validada externamente en entornos comparables) (1)
- ☐ Evidencia piloto o validación limitada (2)
- ☐ Afirmaciones del proveedor sin validación independiente (3)
- ☐ No hay pruebas claras que lo respalden (4)

Preparación en materia de gobernanza

¿Están claramente definidos la gobernanza y la rendición de cuentas para esta iniciativa? (Seleccione una opción)

- ☐ Se ha definido claramente la titularidad, la estructura de supervisión formal y la responsabilidad en la toma de decisiones (1)
- ☐ Responsabilidad definida; supervisión informal o en desarrollo (2)
- ☐ Funciones de gobernanza poco claras o asignadas solo parcialmente (3)
- ☐ La gobernanza y la rendición de cuentas aún no están definidas (4)

Guía para la puntuación del Executive Scan

El análisis ejecutivo inicial ofrece tres resultados preliminares:

- Valor estratégico (bajo / moderado / alto)
- Exposición al riesgo (Baja / Moderada / Alta)
- Preparación preliminar (Débil / En desarrollo / Sólida)

Estimación del valor estratégico

El valor estratégico alto suele incluir iniciativas que:

- Abordan prioridades de salud pública claramente definidas.
- Demuestran un impacto cuantificable.
- Mejoran el acceso, la vigilancia o el rendimiento del sistema.

Puede indicarse un valor estratégico limitado cuando los objetivos son exploratorios o no están claramente alineados con las prioridades institucionales.

Estimación de la exposición al riesgo

La exposición al riesgo es elevada cuando:

- La iniciativa influye directamente en las decisiones de diagnóstico o de elegibilidad.
- Afecta a la asignación de recursos públicos.
- Afecta a poblaciones vulnerables o desatendidas.
- Automatiza la toma de decisiones humanas.
- Opera a gran escala o a escala nacional.

Estimación preliminar de la preparación para la gobernanza

La preparación para la gobernanza es mayor cuando:

- Se define claramente la titularidad.
- Existen mecanismos de supervisión.
- Se han establecido procesos de rendición de cuentas y presentación de informes.

Tabla resumen del análisis ejecutivo

Rellene este resumen tras calcular la puntuación media de cada dimensión.

Dimensión	Secciones incluidas	Cálculo	Puntuación media	Interpretación
Valor estratégico	3,1 + 3,2 + 3,3	(3 puntuaciones ÷ 3)	____	Bajo / Moderado / Alto
Exposición al riesgo	3,4 + 3,5	(2 puntuaciones ÷ 2)	____	Bajo / Moderado / Alto
Preparación preliminar	3,6 + 3,7	(2 puntuaciones ÷ 2)	____	Débil / En desarrollo / Sólido

Asignación final de nivel

Clasificación final	Evaluación: consulte la sección 3.10
Nivel asignado	☐ Nivel 1 – Uso responsable de la IA ☐ Nivel 2 – IA a nivel de programa ☐ Nivel 3 – Gobernanza de la IA de alto impacto

Conclusión ejecutiva

Elemento	Evaluación
Tipo de iniciativa	_____ (descripción concisa en texto libre de la iniciativa de IA que se está evaluando)
Recomendación general (completar tras el cálculo del índice; véase la sección 4 para el Nivel 1, la sección 5 para el Nivel 2 y la sección 6 para el Nivel 3)	☐ Proceder ☐ Seguir adelante con condiciones ☐ Reconsiderar / Rediseñar

Determinación del nivel

Nivel 1: Uso responsable de la IA

El nivel 1 se aplica a iniciativas de IA de bajo riesgo que no influyen directamente en las decisiones clínicas, las determinaciones de elegibilidad o la asignación de recursos públicos.

Algunos ejemplos son:

- El uso de grandes modelos de lenguaje para redactar o resumir documentos.
- Herramientas de automatización administrativa.
- Sistemas de optimización de flujos de trabajo internos.
- Herramientas de análisis de datos no clínicos.
- Aplicaciones de traducción o transcripción.

El Nivel 1 garantiza que se apliquen las medidas básicas de gobernanza, rendición de cuentas y protección de datos sin imponer una carga administrativa innecesaria.

Tiempo estimado de realización: 5-7 minutos.

Gobernanza y rendición de cuentas

(Alcance del Nivel 1)

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Se ha definido la titularidad para el uso de herramientas de IA	☐	☐	☐
Orientación interna clara sobre el uso aceptable	☐	☐	☐
Términos del proveedor revisados	☐	☐	☐
Proceso para notificar usos indebidos o incidentes	☐	☐	☐
Transparencia en las actualizaciones o versiones comprendida	☐	☐	☐

Orientación

- La propiedad debe identificar una unidad o un responsable.
- Incluso las herramientas de bajo riesgo requieren una responsabilidad clara.
- El personal debe saber a quién acudir en caso de error o duda.

Uso y control humano

(Ámbito de nivel 1)

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Se requiere verificación humana antes de utilizar el resultado	☐	☐	☐
Los resultados de la IA no se consideran decisiones definitivas	☐	☐	☐
Se informa a los usuarios de las limitaciones y del riesgo de alucinaciones	☐	☐	☐
El uso de la IA no sustituye al criterio profesional	☐	☐	☐
Se ha impartido formación o sensibilización al personal	☐	☐	☐

Orientación

- Los resultados generados por la IA siempre deben ser revisados por una persona.
- El personal debe comprender que las herramientas de IA pueden producir contenido inexacto o inventado.
- La responsabilidad profesional recae en el usuario humano.

Medidas de protección contra riesgos y sesgos

(Ámbito de nivel 1)

Confirme lo siguiente:

Elemento	Sí	Parcial	No
No se introducen datos identificables de pacientes en herramientas de IA públicas	☐	☐	☐
Se dispone de directrices sobre la protección de datos sensibles	☐	☐	☐
Se respetan las políticas de intercambio de datos	☐	☐	☐
Concienciación sobre posibles sesgos en el contenido	☐	☐	☐

Orientación

- Las herramientas de nivel 1 (uso responsable de la IA) no deben utilizarse para procesar datos sanitarios identificables a menos que se disponga de entornos seguros aprobados.
- El personal debe evitar introducir información confidencial, privada o protegida legalmente en sistemas externos.

Preparación para la implementación

(alcance del Nivel 1)

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Herramienta evaluada en cuanto a su relevancia operativa	☐	☐	☐
Caso de uso claramente definido	☐	☐	☐
Existe un mecanismo básico de seguimiento o retroalimentación	☐	☐	☐
Se han tenido en cuenta las implicaciones en materia de sostenibilidad o suscripción	☐	☐	☐

Orientación

- Incluso el uso de IA de bajo riesgo debe tener un propósito operativo definido.
- Las instituciones deben tener en cuenta las implicaciones en materia de licencias, costes y continuidad.

Resumen de puntuación del Nivel 1

(Uso responsable de la IA)

Asigne las puntuaciones de la siguiente manera:

Sí = 2 puntos	Parcial = 1 punto	No = 0 punto

Calcule el total de puntos posibles en todos los elementos del Nivel 1 (y normalice la puntuación como porcentaje del total de puntos posibles).

Interpretación del Nivel 1

Continuar

80-100 % de los puntos totales	PHVI ≥ **50**	Preparación institucional compuesta ≥ **60**

Autorización condicional temporal – Verificación pendiente

60-79 % de los puntos totales

Reconsiderar / Rediseñar

Menos del **60 %** de los puntos totales

Plantilla de recomendación de nivel 1

Resultado de la evaluación:

☐ Continuar

☐ Autorización condicional temporal — Pendiente de verificación

☐ Reforzar las medidas de seguridad antes de su uso

Si se requieren condiciones, especifique las medidas correctivas a continuación:

1. .....

2. .....

Nivel 2: IA a nivel de programa

El Nivel 2 se aplica a las iniciativas de IA que:

- Apoyan o influyen en la toma de decisiones.
- Operan a nivel de programa, institucional o regional.
- Afectan a poblaciones de pacientes identificables.
- Incluyen herramientas de cara al público o componentes de prestación de servicios.
- Requieren una gobernanza y un seguimiento estructurados.

Algunos ejemplos son:

- Sistemas de puntuación de riesgos.
- Herramientas de apoyo a la toma de decisiones clínicas.
- Aplicaciones de salud dirigidas al público.
- Cuadros de mando de vigilancia.
- Sistemas de asistencia para la clasificación de pacientes.
- Análisis predictivo a nivel de programa.

El Nivel 2 garantiza una gobernanza estructurada, medidas de protección contra sesgos y mecanismos de supervisión antes de la implementación.

Tiempo estimado de finalización: 15-20 minutos.

Gobernanza y rendición de cuentas

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Propiedad definida y autoridad responsable	☐	☐	☐
Mecanismo de supervisión formal establecido	☐	☐	☐
Obligaciones de los proveedores claramente documentadas	☐	☐	☐
Proceso de notificación de incidentes definido	☐	☐	☐
Procedimientos de notificación de actualizaciones definidos	☐	☐	☐
Documentación clara de las limitaciones del sistema	☐	☐	☐

Orientación

- La gobernanza debe incluir una autoridad responsable designada.
- Los contratos con los proveedores deben abordar la transparencia, las actualizaciones y la presentación de informes de rendimiento.
- Las limitaciones conocidas deben documentarse antes de la implementación.

Uso y control humano

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Se requiere revisión humana antes de la acción	☐	☐	☐
Capacidad de anulación disponible	☐	☐	☐
Acciones de anulación registradas y revisables	☐	☐	☐
Usuarios formados en las capacidades y limitaciones del sistema	☐	☐	☐
Delimitación clara entre la recomendación de la IA y la decisión final	☐	☐	☐

Orientación

- La IA no debe sustituir el criterio profesional.
- Los mecanismos de anulación deben ser técnicamente viables y operativamente claros.
- La formación debe incluir las limitaciones del sistema y los escenarios de error.

Medidas de protección contra riesgos y sesgos

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Fuentes de datos de entrenamiento documentadas	☐	☐	☐
Se ha evaluado la representatividad de la población	☐	☐	☐
Evaluado el rendimiento de los subgrupos	☐	☐	☐
Revisión de las variables sustitutivas para detectar sesgos no intencionados	☐	☐	☐
Consideraciones de imparcialidad o equidad documentadas	☐	☐	☐
Plan para el seguimiento de la equidad tras la implementación	☐	☐	☐

Orientación

- Evalúe si los datos de entrenamiento reflejan la población de implementación prevista.
- Se debe examinar el rendimiento de los subgrupos siempre que sea posible.
- Se deben revisar cuidadosamente los indicadores sustitutivos (por ejemplo, la ubicación geográfica o los indicadores socioeconómicos).
- Se deben tener en cuenta los impactos en materia de equidad antes de la ampliación.

Preparación para la implementación

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Evidencia que respalda la eficacia documentada	☐	☐	☐
Validación externa o independiente disponible	☐	☐	☐
Plan de seguimiento definido	☐	☐	☐
Métricas de rendimiento claramente especificadas	☐	☐	☐
Plan de detección de desviaciones o recalibración definido	☐	☐	☐
Se ha evaluado la sostenibilidad y la capacidad operativa	☐	☐	☐
Estimación de los costes operativos recurrentes durante 24 meses y fuente de financiación identificada	☐	☐	☐
Sostenibilidad de la dotación de personal de supervisión	☐	☐	☐

Orientación

- Las pruebas de validación deben revisarse independientemente de las afirmaciones de marketing del proveedor.
- Los planes de seguimiento deben incluir la frecuencia, la responsabilidad y los mecanismos de notificación.
- Las instituciones deben confirmar la capacidad operativa antes de la implementación.

Resumen de puntuación del Nivel 2

(IA a nivel de programa)

Asigne las puntuaciones de la siguiente manera:

Sí = 2 puntos	Parcial = 1 punto	No = 0 punto

Calcule el total de puntos posibles en todos los elementos del Nivel 2.

A continuación, calcule (consulte la sección 7 para obtener detalles sobre el cálculo y la normalización):

- Índice de Preparación para la Gobernanza (GRI)
- Índice de uso y control humano
- Índice de sesgo y riesgo de equidad (BERI)
- Índice de Evidencia y Transparencia (ETI)

Normalice cada uno de ellos en una escala de 0 a 100 según el marco de puntuación de la sección 7.

Interpretación del Nivel 2

Para las iniciativas de Nivel 3:

Continuar

- Índice de valor para la salud pública ≥ 60
- Preparación institucional compuesta ≥ 65
- Índice de riesgo de sesgo y equidad (BERI) ≥ 60

Autorización condicional temporal — En espera de verificación

- Existe valor estratégico, pero se han identificado deficiencias en la preparación.
- Se requieren medidas de seguridad antes de la implementación.

Reconsiderar / Rediseñar

- Valor estratégico limitado.
- Deficiencias significativas en materia de gobernanza o sesgos.
- Base empírica insuficiente.

Plantilla de recomendaciones de nivel 2

Resultado de la evaluación:

☐ Continuar

☐ Autorización condicional temporal — Pendiente de verificación

☐ Reconsiderar / Rediseñar

Medidas de seguridad necesarias (si procede):

1. .....

2. .....

3. .....

Nivel 3: Gobernanza de la IA de alto impacto

El Nivel 3 se aplica a las iniciativas de IA que:

- Influyen directamente en el diagnóstico, la elegibilidad o la determinación de prestaciones.
- Influyen en la asignación de recursos públicos.
- Operan a nivel nacional o a gran escala.
- Afectan a poblaciones vulnerables o desatendidas.
- Automatizan o sustituyen sustancialmente la toma de decisiones humana.
- Procesar datos de salud identificables a gran escala.

Algunos ejemplos son:

- Sistemas nacionales de detección de brotes.
- Herramientas de diagnóstico asistidas por IA.
- Modelos de asignación de recursos o de establecimiento de prioridades.
- Sistemas de determinación de elegibilidad.
- Modelos predictivos de riesgo a nivel poblacional.

El nivel 3 requiere estructuras de gobernanza integrales, supervisión formal, medidas de protección contra sesgos y un seguimiento continuo antes de su implementación.

Tiempo estimado de finalización: 25-35 minutos.

Gobernanza y rendición de cuentas

Para las iniciativas del Nivel 3, las estructuras de gobernanza deben ser formales y estar documentadas.

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Titularidad institucional definida a alto nivel	☐	☐	☐
Se ha establecido un comité de supervisión o un órgano de revisión formal	☐	☐	☐
Se ha completado una revisión jurídica y normativa clara	☐	☐	☐
Los contratos con los proveedores incluyen cláusulas de transparencia y auditoría	☐	☐	☐
Se ha definido un protocolo de notificación y escalado de incidencias	☐	☐	☐
Procedimientos de control de versiones y notificación de actualizaciones documentados	☐	☐	☐
Evaluación independiente permitida por contrato	☐	☐	☐

● **Requisito obligatorio**

En el Nivel 3, la ausencia de supervisión formal o de una responsabilidad definida limita significativamente la preparación y puede impedir la implementación. La evaluación de AI-GUARD debe ser realizada por un equipo de revisión que no incluya al director del proyecto ni al equipo de desarrollo principal. Esta revisión interna cruzada debe documentarse e incluir una declaración de conflicto de intereses.

La revisión externa puede ser realizada por instituciones académicas, organismos sanitarios homólogos, organismos reguladores nacionales o expertos independientes cualificados. Se puede solicitar cooperación técnica para apoyar esta revisión cuando sea aplicable.

Uso y control humano

Los sistemas de nivel 3 deben preservar una supervisión humana significativa.

Cuadro de definiciones — Sección 6.2

- **Human-in-the-Loop (HITL):** Un ser humano revisa y aprueba la recomendación de la IA antes de que se lleve a cabo cualquier acción.

- **Human-on-the-Loop (HOTL):** El sistema actúa por defecto, pero un ser humano puede intervenir y anularlo dentro de un intervalo de tiempo definido. Mayor riesgo que HITL.
- **Human-in-Command (HIC):** Un ser humano mantiene el control operativo total y puede suspender o desactivar el sistema en cualquier momento. Requerido para la implementación de nivel 3.

(Nota: Esta taxonomía de supervisión se ajusta a los requisitos de «Agencia humana y supervisión» definidos por el Grupo de Expertos de Alto Nivel sobre IA de la Comisión Europea, y respalda el principio rector de la OMS de «Proteger la autonomía humana» en la asistencia sanitaria).

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Se requiere revisión humana antes de la decisión final (HITL implementado)	☐	☐	☐
Capacidad de anulación individual implementada técnicamente	☐	☐	☐
Capacidad de suspensión del sistema o desactivación de emergencia (HIC) implementada	☐	☐	☐
Decisiones de anulación registradas y auditables	☐	☐	☐
Formación de los usuarios documentada formalmente	☐	☐	☐
Escenarios de fallos y errores definidos	☐	☐	☐

● Requisito obligatorio

Los sistemas de nivel 3 deben funcionar bajo supervisión de **«Human-in-the-Loop» (HITL)**, a menos que el cambio a **«Human-on-the-Loop» (HOTL)** o a una ejecución totalmente automatizada esté explícitamente justificado, documentado y aprobado tras una revisión legal y ética.

Independientemente del modo operativo (HITL, HOTL o autónomo), todos los sistemas de Nivel 3 **deben mantener** capacidades de **«Human-in-Command» (HIC)**, garantizando que una autoridad designada pueda desactivar el sistema de forma segura en caso de un incidente de Nivel 3 (véase el Anexo E).

Salvaguardias contra riesgos y sesgos

El Nivel 3 (Gobernanza de la IA de alto impacto) requiere una evaluación estructurada de la equidad y la representatividad.

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Fuentes de datos de entrenamiento completamente documentadas	☐	☐	☐
Representatividad de la población evaluada formalmente	☐	☐	☐
Métricas de rendimiento de los subgrupos evaluadas	☐	☐	☐
Variables proxy analizadas en cuanto al riesgo de sesgo	☐	☐	☐
Métricas de equidad aplicadas y documentadas	☐	☐	☐
Evaluación del impacto en la equidad realizada	☐	☐	☐
Se ha definido un plan para el seguimiento continuo de la equidad	☐	☐	☐

Requisitos obligatorios

Para el Nivel 3:

- Se requiere una evaluación del rendimiento de los subgrupos.
- Se deben documentar los indicadores de equidad.
- La ausencia de pruebas de subgrupos reduce significativamente la preparación.

Preparación para la implementación

Los sistemas de Nivel 3 requieren una validación y supervisión sólidas antes de su implementación.

Confirme lo siguiente:

Elemento	Sí	Parcial	No
Validación externa o independiente completada	☐	☐	☐
Tarjeta del modelo o documentación equivalente disponible	☐	☐	☐
Métricas de rendimiento claramente definidas	☐	☐	☐
Plan de seguimiento y evaluación documentado	☐	☐	☐
Se ha definido un plan de detección de desviaciones o de recalibración	☐	☐	☐
Plan de seguimiento de impactos adversos definido	☐	☐	☐
Se ha evaluado la sostenibilidad operativa	☐	☐	☐
Estimación de los costes operativos recurrentes durante 24 meses y fuente de financiación identificada	☐	☐	☐
Estimación de los costes operativos recurrentes durante 24 meses Sostenibilidad de la dotación de personal de supervisión humana	☐	☐	☐

Requisitos obligatorios

Para el Nivel 3:

- Se recomienda encarecidamente la validación externa.
- Se debe documentar un plan de supervisión continua antes de la implementación.
- Un plan financiero insostenible debe considerarse una deficiencia de preparación que requiere, como mínimo, una «Autorización condicional temporal — Pendiente de verificación».

Resumen de puntuación del Nivel 3

(Gobernanza de la IA de alto impacto)

Asigne las puntuaciones de la siguiente manera:

Sí = 2 puntos	Parcial = 1 punto	No = 0 punto

Calcule las puntuaciones parciales para:

- Índice de Preparación para la Gobernanza (GRI)
- Índice de uso y control humano
- Índice de sesgo y riesgo de equidad (BERI)
- Índice de Evidencia y Transparencia (ETI)

Normalice cada uno de ellos en una escala de 0 a 100 según el marco de puntuación de la sección 7.

Interpretación del Nivel 3

Para las iniciativas de nivel 3:

Continuar

- Índice de valor para la salud pública (PHVI) ≥ 70
- Preparación institucional compuesta ≥ 70
- Índice de riesgo de sesgo y equidad (BERI) ≥ 70

Autorización condicional temporal — En espera de verificación

- Si se pueden identificar deficiencias subsanables.

Reconsiderar / Rediseñar

- Si existen deficiencias estructurales.

Plantilla de recomendaciones de nivel 3

Resultado de la evaluación:

☐ Continuar

☐ Autorización condicional temporal — Pendiente de verificación

☐ Reconsiderar / Rediseñar

Medidas de seguridad obligatorias antes de la implementación:

1. .....

2. .....

3. .....

Organismo de supervisión responsable:

.....

Frecuencia de seguimiento:

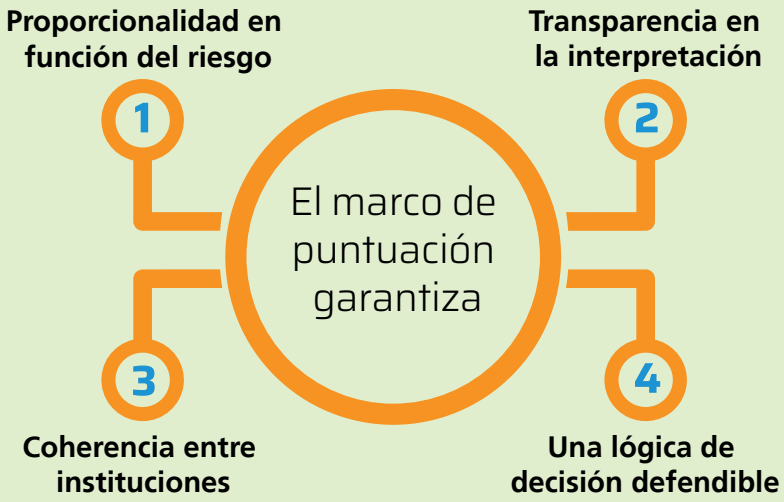
.....

Calendario de reevaluación:

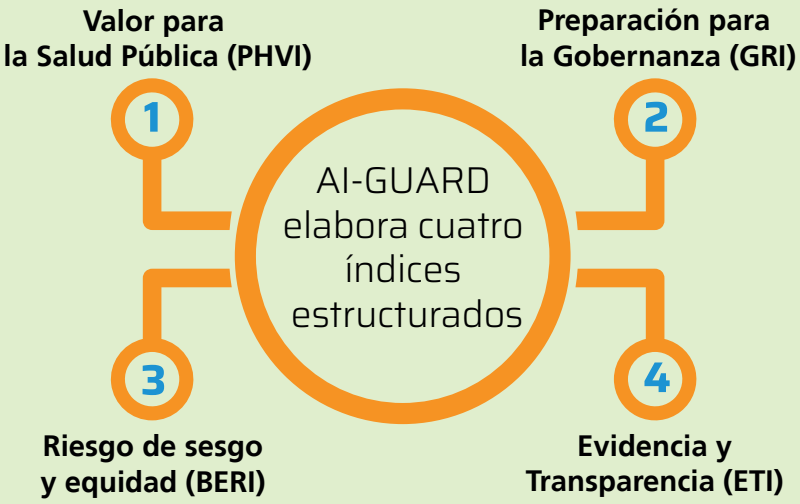
.....

Marco de puntuación y reglas de decisión de AI-GUARD

En esta sección se describe cómo se calculan los resultados de la evaluación de AI-GUARD y cómo se determinan las recomendaciones finales.



Resumen de los índices de AI-GUARD



Además, la clasificación por niveles refleja el grado de exposición al riesgo.

Las recomendaciones finales se basan en la interpretación combinada de:

- Nivel de categoría
- PHVI
- Preparación institucional combinada
- BERI

Índice de Valor de Salud Pública (PHVI)

El Índice de Valor de Salud Pública evalúa la justificación estratégica de la iniciativa de IA.

El PHVI se calcula en función de:

- Claridad del problema
- Resultados medibles
- Alineación con la estrategia nacional o institucional
- Impacto previsto en el acceso, la eficiencia, la vigilancia o el rendimiento del sistema
- Sostenibilidad y capacidad de integración

Cada componente se puntúa en una escala estandarizada y se normaliza a una escala de 0 a 100.

Interpretación del PHVI

Puntuación del PHVI	Interpretación
80–100	Alto valor estratégico
60–79	Valor estratégico moderado
40–59	Valor estratégico limitado
Menos de 40	Justificación débil

Las iniciativas con una justificación débil deben reconsiderarse, especialmente en contextos de Nivel 2 o Nivel 3.

Índice de Preparación para la Gobernanza (GRI)

El GRI refleja la presencia y la solidez de:

- Responsabilidad definida
- Mecanismos de supervisión
- Obligaciones de transparencia de los proveedores
- Estructuras de notificación de incidentes
- Procedimientos de control de versiones y actualización

Las puntuaciones se derivan de elementos de evaluación específicos de cada nivel y se normalizan en una escala de 0 a 100.

Las iniciativas de nivel superior requieren estructuras de gobernanza más formales para alcanzar la plena preparación.

Índice de riesgo de sesgo y equidad (BERI)

El BERI refleja el grado en que se mitigan los riesgos de sesgo y equidad.

Evalúa:

- La representatividad de los datos
- Evaluación del rendimiento de los subgrupos
- Análisis de variables proxy
- Métricas de equidad (obligatorias en el Nivel 3)
- Seguimiento del impacto en la equidad

Las puntuaciones se normalizan en una escala de 0 a 100.

Las puntuaciones más altas indican mayores garantías de mitigación del sesgo.

En las iniciativas del Nivel 3:

- La ausencia de evaluación de subgrupos reduce significativamente el BERI.
- La ausencia de métricas de equidad puede impedir el avance hacia la aprobación total.

Índice de Evidencia y Transparencia (ETI)

El ETI evalúa:

- Evidencia de validación
- Evaluación externa o independiente
- Documentación del modelo (p. ej., ficha del modelo)
- Plan de seguimiento definido
- Mecanismos de detección de desviaciones
- Planificación de la sostenibilidad operativa

Las puntuaciones se normalizan en una escala de 0 a 100.

Las iniciativas de nivel superior requieren pruebas más sólidas para cumplir los umbrales de preparación.

Índice de uso y control humano

Evalúa:

- La presencia de una revisión humana antes de la acción y la claridad del papel humano en las decisiones respaldadas por la IA.
- La disponibilidad y eficacia de los mecanismos de anulación que permiten a los usuarios intervenir o rechazar los resultados de la IA.
- El registro y la auditabilidad de las acciones de anulación y las intervenciones humanas.
- La existencia de capacidades de suspensión del sistema o de desactivación de emergencia, cuando proceda.
- La formación de los usuarios y su conocimiento de las capacidades, limitaciones y escenarios de fallo del sistema.
- Delimitación clara entre las recomendaciones de la IA y las decisiones humanas finales, incluida la asignación de responsabilidades.

Las puntuaciones se normalizan en una escala de 0 a 100.

Las iniciativas de nivel superior requieren mecanismos de control humano más sólidos y formalizados para cumplir los umbrales de preparación. En las iniciativas de Nivel 3, la ausencia de capacidades de anulación humana o de desactivación de emergencia implementadas técnicamente puede impedir el avance hacia la aprobación total.

Preparación institucional compuesta

La preparación institucional compuesta se calcula como la media de:

- Índice de preparación de la gobernanza (GRI)
- Índice de uso y control humano
- Índice de Evidencia y Transparencia (ETI)

Este índice compuesto refleja si la institución está preparada estructuralmente para implementar la iniciativa de IA de manera responsable.

Umbrales por niveles

AI-GUARD aplica umbrales progresivamente más estrictos según el nivel de categoría.

Nivel 1: Uso responsable de la IA	
Proceder cuando:	
PHVI ≥ **50**	Preparación institucional compuesta ≥ **60**

El Nivel 1 hace hincapié en el uso operativo seguro más que en la gobernanza estructural.

Nivel 2: IA a nivel de programa		
Proceder cuando:		
PHVI ≥ **60**	Preparación institucional compuesta ≥ **65**	BERI ≥ **60**

El Nivel 2 (IA a nivel de programa) requiere una gobernanza estructurada y medidas de protección contra sesgos.

Nivel 3: Gobernanza de la IA de alto impacto		
Proceder cuando:		
PHVI ≥ **70**	Preparación institucional compuesta ≥ **70**	BERI ≥ **70**

El nivel 3 requiere una gobernanza sólida, la mitigación de sesgos y la validación de la evidencia antes de la implementación.

Categorías de recomendaciones finales

AI-GUARD genera uno de estos tres resultados:

Seguir adelante

La iniciativa demuestra un valor estratégico adecuado, preparación en materia de gobernanza, medidas de protección contra sesgos y evidencia.

Autorización condicional temporal — Pendiente de verificación

La iniciativa muestra valor estratégico, pero requiere medidas correctivas para subsanar las deficiencias de preparación antes de su implementación y despliegue completo.

Las condiciones pueden incluir:

- Fortalecer las estructuras de gobernanza
- Completar la evaluación de subgrupos
- Realización de una validación externa
- Mitigación de sesgos
- Establecer planes de seguimiento

Hay procedimientos importantes que deben tenerse en cuenta:

- Se debe establecer un plazo de cumplimiento, que no exceda los 90 días a partir de la fecha de evaluación para el Nivel 2, y los 60 días para el Nivel 3.
- Cada requisito debe asignarse a una parte responsable designada.
- La autoridad de supervisión designada debe verificar formalmente el cumplimiento y firmar una Declaración de Verificación de Cumplimiento antes de que el sistema pase a la fase de despliegue completo.

Reconsiderar / Rediseñar

Existen deficiencias significativas en la justificación estratégica, la gobernanza, la mitigación del sesgo o la solidez de la evidencia.

Se requiere un rediseño o una mayor preparación institucional antes de la reevaluación.

Reevaluación y revisión continua

Para las iniciativas de Nivel 2 (IA a nivel de programa) y Nivel 3 (Gobernanza de la IA de alto impacto):

- Se recomienda realizar una reevaluación tras actualizaciones importantes del sistema.
- Se recomienda realizar una reevaluación antes de la ampliación.
- Se debe llevar a cabo una revisión periódica a intervalos adecuados al impacto del sistema.

Reglas de fallo grave: requisitos no compensables

Las siguientes reglas se aplican independientemente de las puntuaciones del índice compuesto. Cuando se activa cualquier criterio de fallo grave, la recomendación general de AI-GUARD se establece automáticamente en «Reconsiderar/Rediseñar», independientemente de los valores de PHVI, GRI, BERI o ETI. Ninguna media de puntuación puede anular una determinación de fallo grave.

#	Criterio de fallo grave	Niveles aplicables	Resultado automático
HF-1	Ausencia de evaluación del rendimiento de los subgrupos	Nivel 3 (obligatorio)	Reconsiderar / Rediseñar
HF-2	Ausencia de validación externa o independiente documentada	Nivel 2 (recomendado)	Reconsiderar / Rediseñar
HF-3	Ausencia de un mecanismo de anulación manual implementado técnicamente y documentado	Niveles 2 y 3	Reconsiderar / Rediseñar
HF-4	No hay una responsabilidad institucional definida a nivel directivo	Niveles 2 y 3	Reconsiderar / Rediseñar
HF-5	Ausencia de un plan de seguimiento continuo previo a la implementación	Nivel 3	Reconsiderar / Rediseñar

Plantilla de resumen de la evaluación de AI-GUARD

El resumen de la evaluación de AI-GUARD ofrece una visión general estructurada de los resultados de la evaluación. Debe completarse tras la evaluación completa y conservarse como parte de la documentación institucional.

Este resumen puede utilizarse en reuniones internas de toma de decisiones, debates sobre adquisiciones, revisiones de supervisión o deliberaciones sobre financiación.

Resumen de la evaluación de AI-GUARD

Título de la iniciativa:

.....

Institución / Programa:

.....

Fecha de evaluación:

.....

Autoridad responsable:

.....

Clasificación por niveles:

☐ Nivel 1 – Uso responsable de la IA

☐ Nivel 2 – IA a nivel de programa

☐ Nivel 3 – Gobernanza de la IA de alto impacto

Índices AI-GUARD

Índice	Puntuación (0–100)	Interpretación
Índice de valor para la salud pública (PHVI)	_____	☐ Alto ☐ Moderado ☐ Limitado ☐ Débil
Índice de preparación para la gobernanza (GRI)	_____	☐ Sólido ☐ Adecuado ☐ En desarrollo ☐ Débil
Índice de riesgo de sesgo y equidad (BERI)	_____	☐ Salvaguardias sólidas ☐ Moderadas ☐ Limitadas ☐ Riesgo elevado
Índice de Evidencia y Transparencia (ETI)	_____	☐ Sólido ☐ Adecuado ☐ Limitado ☐ Insuficiente

Preparación institucional global: _____

Recomendación general

☐ Proceder

☐ Autorización condicional temporal — Pendiente de verificación

☐ Reconsiderar / Rediseñar

Medidas de seguridad necesarias

(si procede)

Enumere las medidas correctivas necesarias antes de la implementación o la ampliación:

1. .....

2. .....

3. .....

4. .....

Supervisión de la gobernanza

- **Nivel 2 (IA a nivel de programa)**
- **Nivel 3 (Gobernanza de la IA de alto impacto)**

Organismo de supervisión responsable:

.....

Frecuencia de seguimiento:

.....

Calendario de reevaluación:

.....

.....

Resumen narrativo

Proporcione una breve descripción (3-5 frases) que resuma los fundamentos de la decisión:

.....

.....

.....

.....

.....

Aprobación definitiva

Nombre y cargo de la autoridad revisora:

.....

Firma: .....

Fecha: .....

Anexos

Anexo A: Cómo utilizar AI-GUARD

AI-GUARD está diseñado para aplicarse de manera estructurada y secuencial. El instrumento puede ser completado por un responsable de la toma de decisiones a título individual, por un equipo técnico o mediante un proceso de revisión institucional facilitado, dependiendo de la escala y la complejidad de la iniciativa de IA que se esté considerando.

La evaluación consta de cuatro etapas principales.

Paso 1: completar el análisis ejecutivo inicial

Todas las iniciativas de IA deben comenzar con el análisis ejecutivo inicial.

Esta breve evaluación:

- Aclara el tipo y el alcance de la iniciativa.
- Identifica el impacto previsto y la población afectada.
- Estima el valor estratégico.
- Evalúa la preparación preliminar en materia de gobernanza.
- Determina el nivel de AI-GUARD adecuado.

El análisis ejecutivo inicial suele tardar unos cinco minutos en completarse.

Una vez completado, la iniciativa se clasificará en uno de los tres niveles:

- **Nivel 1:** Uso responsable de la IA
- **Nivel 2:** IA a nivel de programa
- **Nivel 3:** gobernanza de la IA de alto impacto

El nivel asignado determina la profundidad de la evaluación requerida en los pasos posteriores.

El «Executive Entry Scan» es la fase inicial de toma de decisiones del proceso de evaluación de AI-GUARD. Proporciona una valoración rápida de la relevancia estratégica, el valor esperado para la salud pública, la exposición potencial al riesgo y la preparación preliminar en materia de gobernanza de una iniciativa de inteligencia artificial propuesta, antes de que se lleve a cabo cualquier evaluación técnica detallada. Esta sección debe completarse antes de pasar a la evaluación específica de cada nivel.

Método de puntuación para el análisis ejecutivo inicial

Cada respuesta seleccionada en las secciones **3.1 a 3.6** recibe una puntuación de **1 a 4**, que refleja el nivel de relevancia para la gobernanza, el impacto potencial y la responsabilidad institucional asociados a dicha respuesta.

La escala de puntuación se interpreta de la siguiente manera:

● 1 = Preocupación de gobernanza mínima / condición más simple
● 2 = Moderada-baja
● 3 = Moderada-alta
● 4 = Preocupación de gobernanza más alta / necesidad de gobernanza más fuerte

Una puntuación más alta no indica una mejor calidad. Indica que la iniciativa puede requerir una mayor atención en materia de gobernanza, una evaluación más profunda o salvaguardias adicionales.

Todas las preguntas utilizan la misma **escala del 1 al 4**, incluso cuando una pregunta incluye más de cuatro opciones de respuesta. En tales casos, las puntuaciones se asignan según la importancia en materia de gobernanza de cada respuesta, en lugar del orden en que aparecen las opciones.

Para cada pregunta, solo **debe seleccionarse una opción**: la que mejor represente la condición más alta aplicable en la fase actual de la iniciativa. Si parece aplicable más de una opción, debe elegirse la que tenga **mayor relevancia en materia de gobernanza**.

Tras completar las secciones **3.1 a 3.6**, sume todas las puntuaciones seleccionadas y divida el total entre el número de preguntas respondidas para obtener una **puntuación media**.

La puntuación media se interpreta de la siguiente manera:

● 1,0 a 1,9 = Baja	● 2,0 a 2,9 = Moderada	● 3,0 a 4,0 = Alta

Esta puntuación media respalda la estimación preliminar de:

● Valor estratégico	● Exposición al riesgo	● Preparación preliminar

Estas tres dimensiones se resumen en la sección **3.7** y se utilizan para respaldar **la determinación del nivel en la sección 3.8**.

La puntuación media no debe interpretarse de forma mecánica. Ciertas respuestas, en particular aquellas relacionadas con el diagnóstico, la elegibilidad, la asignación de recursos, las poblaciones vulnerables o el despliegue a gran escala, pueden requerir un nivel más alto incluso cuando la media general sea moderada.

Una puntuación alta **no indica automáticamente que se esté preparado para seguir adelante**. También puede indicar que la iniciativa tiene implicaciones significativas en materia de gobernanza y requiere salvaguardias más sólidas antes de su implementación.

Paso 2: Determinación del nivel

La clasificación por niveles se basa en el nivel de impacto y exposición al riesgo asociados a la iniciativa de IA.

Entre los factores determinantes clave se incluyen:

- Si el sistema influye en las decisiones clínicas o de elegibilidad.
- Si afecta a la asignación de recursos públicos.
- Si se dirige a poblaciones vulnerables o las afecta de manera desproporcionada.
- Si automatiza o sustituye la toma de decisiones humana.
- Si opera a nivel nacional o en el marco de programas a gran escala.
- Si procesa datos de salud identificables.

El nivel de categoría garantiza que la profundidad de la evaluación sea proporcional al impacto potencial.

Las instituciones no deben rebajar manualmente la clasificación de nivel. En caso de incertidumbre, se debe aplicar el nivel más alto como medida de precaución.

En función de las respuestas de la evaluación inicial, asigne la iniciativa a uno de los siguientes niveles. Utilice las definiciones cualitativas que figuran a continuación junto con la interpretación consolidada del valor estratégico, la exposición al riesgo y el grado de preparación preliminar de la sección 3.9. Cuando exista incertidumbre entre dos niveles, se aplicará el nivel más alto.

Definición cualitativa

🟢 **Nivel 1:** Uso responsable de la IA

Bajo impacto en la toma de decisiones, exposición mínima al riesgo e impacto limitado en la población.

🟡 **Nivel 2:** IA a nivel de programa

Impacto moderado, respalda o influye en las decisiones, requiere una revisión de gobernanza estructurada.

🔴 **Nivel 3:** gobernanza de la IA de alto impacto

Influencia directa en el diagnóstico, la elegibilidad o la asignación de recursos; afecta a poblaciones vulnerables; u opera a gran escala.

Si existe incertidumbre entre niveles, se debe aplicar el nivel superior.

Regla de decisión

Utilice la siguiente regla para traducir la interpretación del análisis ejecutivo inicial en una asignación **preliminar** de nivel. Esta regla es **orientativa, no mecánica**, y debe aplicarse junto con los criterios cualitativos de esta sección.

🟠 **Nivel 1 – Uso responsable de la IA:**

Si el Executive Entry Scan arroja tres **«Bajo»** en todas las dimensiones, o **dos «Bajo» más un «Moderado»**, clasifíquelo como **Nivel 1**.

🟠 **Nivel 2 – IA a nivel de programa:**

Si el Executive Entry Scan arroja **dos «Moderado»** (o más), o **tres «Moderado»**, clasifíquelo como **Nivel 2**, a menos que se apliquen los desencadenantes cualitativos del Nivel 3.

🟠 **Nivel 3 – Gobernanza de la IA de alto impacto:**

Si el análisis inicial ejecutivo indica **dos o más** implicaciones **«fuertes/altas»** (por ejemplo, alta exposición al riesgo y alto valor estratégico en contextos nacionales o a gran escala), clasifíquelo como **Nivel 3**. Cualquier factor desencadenante cualitativo del Nivel 3 que se indique a continuación t ambién determina el Nivel 3.

🟠 Paso 3: completar la evaluación específica del nivel

Tras la confirmación del nivel, la iniciativa debe someterse a una evaluación basada en los cuatro pilares fundamentales de AI-GUARD:

- 1 Gobernanza y responsabilidad
- 2 Uso y control humano
- 3 Salvaguardias contra riesgos y sesgos
- 4 Preparación para la implementación

La profundidad y el número de elementos de evaluación aumentan según el nivel de categoría.

- **Nivel 1 - Uso responsable de la IA:** se centra en las medidas de protección básicas para un uso de la IA de bajo riesgo.

- **Nivel 2: IA a nivel de programa:** requiere una revisión estructurada de la gobernanza, la mitigación de sesgos y los planes de supervisión.

- **Nivel 3: gobernanza de la IA de alto impacto:** requiere estructuras de gobernanza exhaustivas, pruebas de validación y mecanismos de supervisión continua.

Los elementos de evaluación se puntúan según el marco de puntuación de AI-GUARD descrito en la sección 5.

Todas las respuestas deben documentarse con pruebas justificativas, cuando estén disponibles.

Paso 4: Revisar el resumen del panel de control de AI-GUARD

Una vez completada la evaluación, los resultados se resumen en el panel de control de AI-GUARD.

El panel incluye cuatro índices:

- 1 Índice de valor para la salud pública (PHVI)
- 2 Índice de riesgo de sesgo y equidad (BERI)
- 3 Índice de preparación para la gobernanza (GRI)
- 4 Índice de Evidencia y Transparencia (ETI)

En función de los resultados combinados, la iniciativa recibirá una de las siguientes recomendaciones:

- Seguir adelante
- Autorización condicional temporal — Pendiente de verificación
- Reconsiderar / Rediseñar

Para las iniciativas que reciban una «Autorización condicional temporal — Pendiente de verificación» o una «Reconsiderar / Rediseñar», AI-GUARD identificará las medidas de seguridad necesarias y las deficiencias de preparación que deben abordarse antes de la implementación.

Responsabilidades institucionales

Los resultados de AI-GUARD deben:

- Revisados por la autoridad responsable designada.
- Documentarse y conservarse como parte de los registros del proyecto.
- Reevaluados si el sistema de IA sufre modificaciones sustanciales, ampliaciones o cambios funcionales.

Para las iniciativas de Nivel 3 (Gobernanza de la IA de alto impacto), se recomienda encarecidamente realizar reevaluaciones periódicas.

Frecuencia de aplicación

AI-GUARD debe aplicarse:

- Antes de las decisiones de adquisición.
- Antes de la implementación piloto.
- Antes del despliegue a escala nacional o del programa.
- Cuando se produzcan actualizaciones significativas del modelo.
- Cuando cambien el contexto operativo o de gobernanza.

El instrumento tiene por objeto apoyar el aprendizaje institucional continuo y la gobernanza responsable de la IA. Los modelos de IA, en particular los utilizados en contextos clínicos o epidemiológicos, están sujetos a una disminución de su rendimiento con el paso del tiempo debido a la deriva de los datos, los cambios en la población y la evolución de la práctica clínica. Por ello, es necesario realizar una reevaluación periódica.

Las evaluaciones de AI-GUARD tienen un periodo de validez definido. Las instituciones deben llevar a cabo una reevaluación completa antes de la fecha de vencimiento aplicable o ante cualquier desencadenante de reevaluación anticipada obligatoria, lo que ocurra primero.

Nivel	Vencimiento estándar	Motivos que dan lugar a una reevaluación anticipada obligatoria
Nivel 1	24 meses	Cambio significativo en el alcance del uso de la herramienta o en las prácticas de manejo de datos
Nivel 2	18 meses	Actualización importante del modelo; cambio en la población objetivo; ampliación a nuevas áreas del programa; informes de acontecimientos adversos
Nivel 3	12 meses	Cualquier actualización del modelo subyacente; ampliación a nuevas áreas geográficas; cambio en el contexto de implementación; detección de impactos adversos; cambios legales o normativos

Anexo B: Simulación - Nivel 1: Uso responsable de la IA

Caso: Uso de un modelo de lenguaje grande para la redacción de políticas

Descripción de la iniciativa

Un departamento del Ministerio de Sanidad propone autorizar el uso de un modelo de lenguaje grande (LLM) empresarial, como ChatGPT o Gemini, para apoyar el trabajo administrativo interno, incluyendo:

- la redacción de resúmenes de políticas
- resumir informes técnicos
- traducir documentos internos
- mejorar la claridad de la redacción

No se introducirán en el sistema datos identificables de pacientes. El uso se limita a funciones administrativas internas y los resultados siguen estando sujetos a revisión humana antes de su uso institucional.

Escaneo de entrada ejecutiva – Puntuación de secciones

Sección	Puntuación seleccionada
3.1 Problema de salud pública abordado	1
3.2 Justificación del uso de la IA	2
3.3 Identificación de la iniciativa	1
3.4 Impacto de la decisión	1
3.5 Impacto en la población	1
3.6 Base empírica	2
3.7 Preparación en materia de gobernanza	2

Tabla resumen del análisis ejecutivo

Dimensión	Secciones incluidas	Cálculo	Puntuación media	Interpretación
Valor estratégico	3,1 + 3,2 + 3,3	(1+2+1) ÷ 3	1,3	Bajo
Exposición al riesgo	3,4 + 3,5	(1+1) ÷ 2	1,0	Baja
Preparación preliminar	3,6 + 3,7	(2+2) ÷ 2	2,0	En desarrollo

Asignación de nivel final

Nivel 1 – Uso responsable de la IA

Conclusión ejecutiva

Elemento	Evaluación
Tipo de iniciativa	Uso de herramientas de IA generativa para apoyo operativo limitado
Recomendación general	Autorización condicional temporal – Pendiente de verificación

Justificación

- La iniciativa mejora la eficiencia operativa interna.
- No influye directamente en las decisiones clínicas, la elegibilidad ni la asignación de recursos.
- La exposición de la población es mínima, ya que el uso sigue siendo interno.
- Los requisitos de gobernanza siguen siendo moderados, pero aún así requieren orientación institucional.

Resumen de la evaluación de nivel 1

Gobernanza y rendición de cuentas

- Responsabilidad definida: Sí (2)

- Orientación sobre el uso interno: Parcial (1)
- Términos del proveedor revisados: Sí (2)
- Definición de notificación de incidentes: Limitada (0)

Uso y control humano

- Se requiere verificación humana: Sí (2)
- El resultado de la IA no se considera una decisión definitiva: Sí (2)
- Formación del usuario proporcionada: Parcial (1)

Medidas de protección contra riesgos y sesgos

- No se introducen datos identificables del paciente: Sí (2)
- Se proporcionan directrices de confidencialidad: Sí (2)
- Concienciación sobre el riesgo de alucinaciones: Parcial (1)

Preparación para la implementación

- Caso de uso claramente definido: Sí (2)
- Se ha tenido en cuenta la sostenibilidad operativa: Sí (2)
- Mecanismo de retroalimentación definido: Parcial (1)

Total: 76,9 ▸ Pendiente de verificación

Medidas de protección recomendadas

- Formalizar las directrices institucionales internas para el uso de la IA generativa.
- Impartir formación estructurada a los usuarios sobre las limitaciones y la verificación de los resultados.
- Establecer un seguimiento periódico de los patrones de uso y los riesgos recurrentes.

Nivel 1: guía de puntuación

Este tutorial ilustra cómo traducir las selecciones **Sí / Parcial / No** en una puntuación porcentual para el Nivel 1, de acuerdo con la Sección **4.5 (Sí = 2; Parcial = 1; No = 0)**. En primer lugar, sume los puntos de todos los elementos del Nivel 1 (Secciones **4.1–4.4**). A continuación, **normalice el total como porcentaje de la puntuación máxima posible** e interprete el resultado utilizando la Sección **4.6**.

Paso 1 — Asignación de puntos (ejemplo de estructura):

- Gobernanza y rendición de cuentas (4 elementos) ▸ [5 / 8]
- Uso y control humano (3 elementos) ▸ [5 / 6]
- Salvaguardias contra riesgos y sesgos (3 elementos) ▸ [5 / 6]
- Preparación para la implementación (3 elementos) ▸ [5 / 6]

Paso 2 — Normalización, PHVI y preparación institucional compuesta:

Normalización:

$$\text{Puntuación de nivel 1 (\%)} \frac{\text{Suma de puntos}}{\text{Máximo puntaje}} \text{x100} = \frac{20}{26} \text{x100} = 76{,}9\%$$

PHVI:

55, Interpretación: Valor estratégico limitado (40–59)

- Claridad del problema: La iniciativa tiene un objetivo claro y delimitado (apoyo administrativo interno), pero no aborda directamente un problema de salud pública definido.
- Resultados medibles: Las mejoras en la eficiencia (ahorro de tiempo, productividad) son medibles, pero están relacionadas indirectamente con los resultados de salud pública.
- Alineación con la estrategia institucional: El uso se alinea con los objetivos generales de transformación digital y eficiencia administrativa, pero no con la prestación de servicios básicos, la vigilancia o las prioridades de salud de la población.
- Impacto esperado: El impacto es interno y operativo, sin efecto directo sobre el acceso, la calidad de la atención, la vigilancia o el rendimiento del sistema a nivel poblacional.
- Sostenibilidad y capacidad de integración: Alta; los LLM empresariales son relativamente fáciles de mantener e integrar para uso administrativo.

- **GRI: 62,5**

Derivado de los elementos de Gobernanza y Rendición de cuentas del Nivel 1: 5/8 puntos ▸ (5/8) × 100.

- **UHCI: 83,3**

Derivado de los elementos de Uso y control humano del Nivel 1: 5/6 puntos ▸ (5/6) × 100.

- **ETI: 83,3**

Derivado de los elementos de «Preparación para el despliegue» de Nivel 1: 5/6 puntos ▸ (5/6) × 100.

- **Preparación institucional compuesta:**

GRI: 62,5 %; UHCI: 83,3 %, ETI: 83,3 %

▸ Compuesto = 76,4 %

Paso 3 — Interpretación (Sección 4.6):

- **Continuar:**
  - 80 – 100 %
  - PHVI ≥ 50
  - Índice compuesto de preparación institucional ≥ 60
- **Verificación pendiente:** 60 – 79 % (SE ENCUENTRA AQUÍ)
- **Rediseño:** Por debajo del 60 %

Anexo C: Simulación – Nivel 2: IA a nivel de programa

Caso: Herramienta de apoyo al triaje en el servicio de urgencias basada en IA

Descripción de la iniciativa

Un hospital regional propone la adopción de una herramienta de triaje asistida por IA para ayudar a predecir el riesgo de deterioro de los pacientes en el servicio de urgencias.

El sistema:

- apoya la priorización de los médicos
- no sustituye las decisiones clínicas finales
- utiliza datos históricos de historias clínicas electrónicas
- influye en los flujos de trabajo de la atención al paciente y en la priorización operativa

La herramienta tiene por objeto mejorar la capacidad de respuesta en la atención de urgencias, manteniendo al mismo tiempo la supervisión clínica humana.

Análisis ejecutivo de entrada – Puntuación de secciones

Sección	Puntuación seleccionada
3.1 Problema de salud pública abordado	2
3.2 Justificación del uso de la IA	2
3.3 Identificación de la iniciativa	2
3.4 Impacto de la decisión	3
3.5 Impacto en la población	2
3.6 Base empírica	2
3.7 Preparación de la gobernanza	2

Tabla resumen del análisis ejecutivo

Dimensión	Secciones incluidas	Cálculo	Puntuación media	Interpretación
Valor estratégico	3,1 + 3,2 + 3,3	(2+2+2) ÷ 3	2,0	Moderado
Exposición al riesgo	3,4 + 3,5	(3+2) ÷ 2	2,5	Moderada
Preparación preliminar	3,6 + 3,7	(2+2) ÷ 2	2,0	En desarrollo

Asignación de nivel final

Nivel 2 – IA a nivel de programa

Conclusión ejecutiva

Elemento	Evaluación
Tipo de iniciativa	Adquisición de un producto basado en IA que apoye la clasificación de pacientes en el servicio de urgencias
Recomendación general	Proceder con condiciones

Justificación

- La iniciativa apoya la priorización clínica, pero no automatiza las decisiones finales.
- Afecta a pacientes identificables y al flujo operativo de la atención.
- Se requieren medidas de gobernanza moderadas, ya que las decisiones influyen en la priorización dentro de un entorno clínico.

Resumen de la evaluación de nivel 2

Gobernanza y responsabilidad

- Responsabilidad definida: Sí
- Mecanismo de supervisión formal: Parcial

- Cláusulas de transparencia de los proveedores: Sí
- Definición de la notificación de incidentes: Parcial

Uso y control humano

- Revisión humana requerida: Sí
- Capacidad de anulación disponible: Sí
- Registro de anulaciones implementado: Parcial
- Formación de usuarios impartida: Parcial

Medidas de protección contra riesgos y sesgos

- Datos de entrenamiento documentados: Sí
- Representatividad de la población evaluada: Parcial
- Rendimiento de los subgrupos evaluado: Parcial
- Análisis de sesgos de proxy realizado: No
- Plan de seguimiento de la equidad definido: Parcial

Preparación para la implementación

- Validación piloto disponible: Sí
- Validación externa: No
- Plan de seguimiento documentado: Parcial
- Plan de detección de desviaciones definido: No

Medidas de seguridad recomendadas

- Realizar una evaluación del rendimiento de subgrupos antes de una implementación más amplia.
- Realizar un análisis de sesgo de las variables proxy.
- Establecer procedimientos estructurados de seguimiento y detección de desviaciones.
- Formalizar una revisión periódica de supervisión.

Recomendación de reevaluación

Se recomienda realizar una reevaluación antes de una expansión institucional o regional más amplia.

Nivel 2: guía para el cálculo del índice

Este tutorial ilustra cómo traducir las selecciones **«Sí», «Parcial»** y **«No»** en **índices normalizados** para el Nivel 2, de conformidad con la **Sección 5.5** (Sí = 2; Parcial = 1; No = 0) y el **Marco de puntuación de AI-GUARD (Sección 7)**. En primer lugar, asigne puntos a los elementos específicos del nivel (Secciones **5.1–5.4**). A continuación, **normalice** cada índice de pilar a **0-100**, calcule la **preparación institucional compuesta** (sección **7.7**) e interprete el resultado utilizando **los umbrales del Nivel 2** (secciones **5.6 / 7.8**).

Paso 1 — Asignación de puntos (ejemplo de esquema)

- Gobernanza y rendición de cuentas (elementos seleccionados utilizados en este ejemplo: 4) ▸ [6 / 8]
  - Sí (Responsabilidad definida, Transparencia del proveedor) = 2 + 2
  - Parcial (Supervisión formal, Notificación de incidentes) = 1 + 1
- Uso y control humano (elementos seleccionados utilizados en este ejemplo: 4) ▸ [6 / 8]
  - Sí (Revisión humana, Capacidad de anulación) = 2 + 2
  - Parcial (Registro de anulaciones, formación de los usuarios) = 1 + 1
- Medidas de protección contra riesgos y sesgos (elementos seleccionados utilizados en este ejemplo: 5) ▸ [5 / 10]
  - Sí (Datos de entrenamiento) = 2
  - Parcial (representatividad, rendimiento de subgrupos, supervisión de la equidad) = 1 + 1 + 1
  - No (Análisis de sesgos de proxy) = 0
- Preparación para la implementación (elementos seleccionados utilizados en este ejemplo: 4) ▸ [3 / 8]
  - Sí (Validación piloto) = 2
  - Parcial (plan de supervisión) = 1
  - No (Validación externa, detección de desviaciones) = 0 + 0

Paso 2 — Normalización, PHVI e índice compuesto de preparación institucional

Índice de valor para la salud pública (PHVI):

Derivado de los cinco componentes del PHVI (claridad del problema, resultados medibles, alineación estratégica, impacto esperado, sostenibilidad/integración). Para este caso de uso

(apoyo al triaje en urgencias para mejorar la capacidad de respuesta con supervisión humana), la evaluación a nivel de componentes arroja: PHVI = 65 ▸ Valor estratégico moderado.

Normalización a una escala de 0 a 100 (por pilar):

- **GRI** = (6/8) × 100 = **75**
- **UHCI** = (6/8) × 100 = **75**
- **BERI** = (5/10) × 100 = **50**
- **ETI** = (3/8) × 100 = **37,5**

Preparación institucional compuesta (por §7.7):

$$\text{"Composite"} = \frac{\text{"GRI"} + \text{"UHCI"} + \text{"ETI"}}{3} = \frac{75+75+37.5}{3} = 62{,}5\%$$

Paso 3 — Interpretación (Secciones 5.6 / 7.8)

Nivel 2: continuar cuando se cumplan todos los requisitos:

- PHVI ≥ 60 ▸ Cumplido (65)
- Preparación institucional compuesta ≥ 65 ▸ No cumplida (62,5)
- BERI ≥ 60 ▸ No cumplida (50)

Recomendación general: Autorización condicional temporal — En espera de verificación (se incluye aquí)

Aspectos que deben abordarse antes de la implementación completa (véase B.8):

- Sesgo y equidad: completar la evaluación del rendimiento de los subgrupos, el análisis del sesgo de los indicadores sustitutivos y formalizar la supervisión de la equidad.
- Evidencia y transparencia: obtener validación externa/independiente; definir umbrales de rendimiento (incluida la calibración); formalizar el seguimiento y la detección de desviaciones.
- Gobernanza y control humano: mecanismo de supervisión formal; registro de anulaciones / pista de auditoría; formación estructurada de los usuarios y límites de funciones (recomendación de IA frente a decisión final).

Anexo D: Simulación – Nivel 3: Gobernanza de la IA de alto impacto

Caso: Sistema nacional de detección temprana de brotes basado en IA

Descripción de la iniciativa

Un Ministerio de Sanidad propone el despliegue de un sistema nacional de vigilancia respaldado por IA para detectar señales tempranas de brotes de enfermedades infecciosas utilizando:

- historiales médicos electrónicos
- sistemas de notificación de laboratorio
- datos de vigilancia sindrómica

El sistema tiene por objeto apoyar:

- la asignación de recursos de emergencia
- la priorización de las estrategias de respuesta de salud pública
- la identificación temprana de riesgos a nivel poblacional

La iniciativa procesa datos sanitarios identificables a escala nacional y puede influir en las decisiones que afectan a las poblaciones vulnerables.

Análisis ejecutivo de entrada – Puntuación de secciones

Sección	Puntuación seleccionada
3.1 Problema de salud pública abordado	3
3.2 Justificación del uso de la IA	1
3.3 Identificación de la iniciativa	4
3.4 Impacto de la decisión	4

Sección	Puntuación seleccionada
3.5 Impacto en la población	4
3.6 Base empírica	3
3.7 Preparación en materia de gobernanza	4

Tabla resumen del análisis ejecutivo

Dimensión	Secciones incluidas	Cálculo	Puntuación media	Interpretación
Valor estratégico	3,1 + 3,2 + 3,3	(3+1+4) ÷ 3	2,7	Moderado-Alto
Exposición al riesgo	3,4 + 3,5	(4+4) ÷ 2	4,0	Alta
Preparación preliminar	3,6 + 3,7	(3+4) ÷ 2	3,5	Bajo

Asignación final de nivel

Nivel 3 – Gobernanza de la IA de alto impacto

Conclusión ejecutiva

Elemento	Evaluación
Tipo de iniciativa	Desarrollo interno de un sistema nacional de vigilancia basado en IA
Recomendación general	Reconsiderar / Rediseñar

Justificación

- La iniciativa opera a escala nacional.
- Influye en la priorización de las medidas de salud pública de emergencia y en la asignación de recursos.

- Afecta a poblaciones con distintos grados de vulnerabilidad.
- La gobernanza y la evidencia siguen siendo insuficientes para su implementación inmediata.

Resumen de la evaluación de nivel 3

Gobernanza y rendición de cuentas

- Se ha definido la responsabilidad de los altos cargos: Sí
- Comité de supervisión formal: Aún no se ha establecido
- Cláusulas de auditoría de proveedores: Parciales
- Protocolo de notificación de incidentes: Limitado

Uso y control humano

- Supervisión humana definida: Parcial
- Registro de anulaciones implementado: No
- Programa de formación formal implementado: No

Medidas de protección contra riesgos y sesgos

- Datos de entrenamiento documentados: Sí
- Rendimiento de los subgrupos evaluado: No
- Métricas de equidad aplicadas: No
- Evaluación del impacto en la equidad realizada: No

Preparación para la implementación

- Validación externa completada: No
- Ficha del modelo disponible: Solo borrador
- Plan de seguimiento definido: Parcial
- Plan de detección de desviaciones definido: No

Medidas de seguridad obligatorias previas a la implementación piloto

- Establecer un comité de supervisión multidisciplinar formal.
- Realizar una evaluación del rendimiento y la equidad de los subgrupos.
- Realizar estudios de validación independientes.

- Definir procedimientos de monitorización continua y detección de desviaciones.
- Consolidar la documentación de gobernanza y las líneas de responsabilidad.

Requisito de reevaluación

Se requiere una reevaluación completa antes de la implementación piloto.

Nivel 3: guía para el cálculo de índices

Este tutorial detalla cómo calcular los **índices del Nivel 3** a partir de la lista de verificación **C.7** utilizando **Sí = 2; Parcial = 1; No = 0**, y luego **normalizarlos a 0–100**. En primer lugar, asigne puntos a los elementos específicos del nivel (Secciones **6.1–6.4**). A continuación, **normalice** cada índice de pilar a una escala de **0 a 100**, calcule la **preparación institucional compuesta** e interprete los resultados en función de los **umbrales del Nivel 3** y **las reglas de fallo grave**.

Paso 1 — Asignación de puntos (ejemplo de formato)

- **Gobernanza y rendición de cuentas ▸ [4 / 8]**
  - **Sí:** Se ha definido la responsabilidad de los altos cargos = **2**
  - **No:** aún no se ha establecido un comité de supervisión formal = **0**
  - **Parcial:** Cláusulas de auditoría de proveedores = **1**
  - **Parcial:** Protocolo de notificación de incidentes (limitado) = **1**
- **Uso y control humano ▸ [1 / 6]**
  - **Parcial:** Supervisión humana definida = **1**
  - **No:** Registro de anulaciones implementado = **0**
  - **No:** Programa de formación formal implementado = **0**
- **Salvaguardias contra riesgos y sesgos ▸ [2 / 8]**
  - **Sí:** Datos de entrenamiento documentados = **2**
  - **No:** Rendimiento de subgrupos evaluado = **0**
  - **No:** Métricas de equidad aplicadas = 0
  - **No:** Se ha realizado una evaluación del impacto en la equidad = **0**
- **Preparación para la implementación ▸ [2 / 8]**
  - **No:** Validación externa/independiente completada = **0**

- **Parcial:** Tarjeta de modelo disponible (borrador) = **1**
- **Parcial:** Plan de seguimiento definido = **1**
- **No:** Plan de detección de desviaciones definido = **0**

Paso 2 — Normalización, PHVI y preparación institucional compuesta

Índice de Valor para la Salud Pública (PHVI) (según el apartado 7.2):

Derivado de los cinco componentes del PHVI (claridad del problema, resultados medibles, alineación estratégica, impacto esperado, sostenibilidad/integración). En el caso de un **sistema nacional de detección temprana de brotes**, la valoración a nivel de componentes suele ser alta en cuanto al valor para la salud pública:

- **Claridad del problema** (detección temprana de brotes a nivel poblacional) = **85**
- **Resultados medibles** (detección oportuna, respuesta más temprana, asignación de recursos) = **70**
- **Alineación con la estrategia** (prioridad central de la vigilancia nacional) = **85**
- **Impacto esperado** (asignación de recursos, priorización de la respuesta de emergencia) = **80**
- **Sostenibilidad e integración** (integración multisistémica compleja; elevados requisitos de operación y mantenimiento) = **60**

$$\text{"PHVI"} = \frac{85+70+85+80+60}{5} = 76\ (\text{"High Strategic Value"})$$

Normalización a una escala de 0 a 100 (por pilar):

- **GRI** = (4/8) × 100 = **50,0**
- **UHCI** = (1/6) × 100 = **16,7**
- **BERI** = (2/8) × 100 = **25,0**
- **ETI** = (2/8) × 100 = **25,0**

Preparación institucional compuesta (según el apartado 7.7):

$$\text{"Composite"} = \frac{\text{"GRI"}+\text{"UHCI"}+\text{"ETI"}}{3} = \frac{50.0+16.7+25.0}{3} = 30{,}6\%$$

Paso 3 — Interpretación

Nivel 3 — Umbrales de proceder (deben cumplirse todos):

- PHVI ≥ 70 ▸ Cumplido (76)
- Preparación institucional compuesta ≥ 70 ▸ No cumplida (30,6)
- BERI ≥ 70 ▸ No cumplido (25,0)

Reglas de fallo grave (según 7.10):

- **HF-1:** Ausencia de **evaluación del rendimiento de los subgrupos ▸ Activada** (C.7 = No).
- **HF-2:** Ausencia de **validación externa/independiente documentada ▸ Activada** (C.7 = No).
- **HF-3:** Ausencia de **anulación manual implementada técnicamente ▸ Probable** (no documentada; si se confirma que **No**, activa HF-3).

Recomendación general: Reconsiderar / Rediseñar

Aspectos que deben abordarse antes de cualquier proyecto piloto (véase C.8):

- **Gobernanza:** establecer un comité de supervisión multidisciplinar formal, definir la revisión legal y normativa, documentar el control de versiones y la notificación de actualizaciones, y permitir la evaluación independiente mediante contrato.
- **Control humano:** implementar HITL (o justificar HOTL), garantizar la desactivación de emergencia del HIC y el registro de anulación; impartir formación formal a los usuarios y definir escenarios de fallo.
- **Sesgo y equidad:** realizar análisis de rendimiento de subgrupos, aplicar métricas de equidad, evaluar el impacto en la equidad y planificar un seguimiento continuo de la equidad.
- **Evidencia y transparencia:** completar la validación externa/independiente, finalizar la ficha del modelo, formalizar el seguimiento y la detección de desviaciones, definir métricas de rendimiento (incluidos umbrales y alertas) y documentar la sostenibilidad operativa y los recursos.

Anexo E: Requisitos de evidencia y ficha del modelo de los proveedores de la Iniciativa para la Salud y la Tecnología

Este anexo proporciona una plantilla estructurada para solicitar la documentación y las pruebas esenciales a los proveedores o desarrolladores de sistemas de IA destinados a su implementación en entornos sanitarios.

Para las iniciativas de Nivel 2 (IA a nivel de programa) y Nivel 3 (Gobernanza de la IA de alto impacto), este anexo debe transmitirse formalmente a los proveedores como parte de los procesos de diligencia debida o de contratación.

Requisitos mínimos de documentación de los proveedores

Se debe solicitar la siguiente documentación antes de la adquisición o el despliegue:

1. Información general del sistema

- Nombre y versión del sistema
- Organización desarrolladora
- Punto de contacto para la responsabilidad técnica
- Uso previsto y población destinataria
- Supuestos sobre el contexto de implementación

2. Descripción del modelo

- Tipo de modelo de IA (p. ej., aprendizaje automático, aprendizaje profundo, PLN)
- Descripción de las fuentes de datos de entrada
- Descripción del formato de salida
- Función de apoyo a la toma de decisiones frente a función de toma de decisiones automatizada
- Limitaciones conocidas del sistema

3. Transparencia de los datos de entrenamiento

Los proveedores deben proporcionar:

- Descripción de las fuentes de los datos de entrenamiento
- Origen geográfico de los datos de entrenamiento
- Periodo de tiempo cubierto
- Características de la población (edad, sexo, región, datos demográficos relevantes)
- Criterios de inclusión y exclusión de datos

Para las iniciativas de nivel 3, se recomienda encarecidamente la divulgación de la representación de los subgrupos.

4. Métricas de rendimiento

Los proveedores deben proporcionar:

- Métricas de rendimiento principales
- Sensibilidad y especificidad (cuando proceda)
- Tasas de falsos positivos y falsos negativos
- Rendimiento de la calibración (en caso de puntuación de riesgo)
- Rendimiento en subgrupos (cuando proceda)
- Características del conjunto de datos de validación

Para las iniciativas de nivel 3, se espera una evaluación del rendimiento en subgrupos.

5. Medidas de equidad y mitigación del sesgo

Los proveedores deben revelar:

- Si se han realizado pruebas de sesgo
- La metodología utilizada para la evaluación del sesgo
- Las métricas de equidad aplicadas
- Las disparidades identificadas (si las hubiera)
- Las medidas de mitigación implementadas
- Planes de supervisión continua de la equidad

En el caso de las iniciativas de Nivel 3 (Gobernanza de la IA de alto impacto), la ausencia de documentación sobre equidad debe considerarse una importante carencia en la preparación.

6. Validación y evaluación

- Métodos de validación internos

- Estudios de validación externos o independientes
- Publicaciones revisadas por pares (si están disponibles)
- Resultados de pruebas piloto en el mundo real
- Autorizaciones reglamentarias (si procede)

Los sistemas de nivel 3 deben demostrar la validación externa siempre que sea posible.

7. Gobernanza y transparencia de las actualizaciones

Los proveedores deben especificar:

- Frecuencia de las actualizaciones del modelo
- Procedimientos de notificación de cambios
- Documentación sobre el control de versiones
- Disponibilidad del registro de auditoría
- Procedimientos de colaboración para la notificación de incidencias

8. Supervisión y gestión de desviaciones

- Estrategia de supervisión tras la implementación
- Indicadores de supervisión del rendimiento
- Métodos de detección de desviaciones
- Procedimientos de recalibración
- Responsabilidades en la gestión del deterioro del rendimiento

Plantilla de ficha de modelo (campos mínimos)

Las instituciones pueden solicitar a los proveedores que rellenen la siguiente ficha de modelo estructurada.

● Ficha del modelo: información mínima requerida

1. Descripción general del modelo

- Nombre y versión del modelo
- Desarrollador
- Fecha de lanzamiento
- Uso previsto
- Usuarios previstos

2. Entradas y salidas del modelo

- Descripción de las variables de entrada
- Formato de los resultados e interpretación
- Definiciones de umbrales (si procede)

3. Resumen de los datos de entrenamiento

- Fuentes de datos
- Cobertura de la población
- Periodo de tiempo de los datos
- Pasos de preprocesamiento de datos

4. Métricas de rendimiento

- Rendimiento general
- Rendimiento de subgrupos (si procede)
- Métricas de calibración (si procede)

5. Evaluación de la equidad

- Métricas de equidad aplicadas
- Disparidades identificadas
- Estrategias de mitigación

6. Limitaciones y advertencias

- Limitaciones conocidas
- Situaciones en las que no se debe utilizar el modelo
- Grupos de población en los que el rendimiento puede variar

7. Plan de seguimiento

- Frecuencia de seguimiento tras la implementación
- Mecanismos de detección de desviaciones
- Punto de contacto para notificar incidencias

Orientación sobre la interpretación

La ausencia de documentación en cualquiera de los ámbitos anteriores debe reflejarse en:

- Índice de Preparación para la Gobernanza (GRI)
- Índice de Riesgo de Sesgo y Equidad (BERI)
- Índice de Evidencia y Transparencia (ETI)

En el caso de las iniciativas de Nivel 3, la ausencia de una evaluación del rendimiento de los subgrupos o de pruebas de validación puede impedir el paso al estado «Continuar».

Impacto estratégico del anexo D

Este anexo:

- Protege a las instituciones durante la contratación
- Aumenta la responsabilidad de los proveedores
- Fomenta la transparencia
- Se ajusta a las mejores prácticas globales en materia de gobernanza de la IA
- Refuerza la credibilidad de AI-GUARD

Modelos desarrollados internamente y adaptados al código abierto

D.4.1 Ámbito de aplicación

Esta sección se aplica cuando una institución o un equipo:

- Desarrolla un modelo de IA internamente utilizando datos sanitarios institucionales o nacionales.
- Ajusta un modelo base de código abierto utilizando conjuntos de datos sanitarios locales, regionales o nacionales.
- Implemente un modelo de código abierto en el contexto de un sistema sanitario sin modificaciones sustanciales, pero asuma la responsabilidad operativa de sus resultados.

D.4.2 Principio de responsabilidad equivalente

El equipo de desarrollo interno o la institución responsable asume todas las obligaciones de documentación, pruebas de sesgo y supervisión que las secciones D.1 y D.2 asignan a los proveedores comerciales externos. La ausencia de un proveedor externo no reduce estas obligaciones.

D.4.3 Requisitos adicionales específicos para la adaptación de código abierto

Además de todos los requisitos de D.1 y D.2, se debe documentar lo siguiente:

- Identidad y versión del modelo base utilizado (por ejemplo, Llama 3.1 8B, Mistral 7B v0.3)
- Limitaciones, sesgos y riesgos documentados conocidos del modelo base, tal y como los han comunicado sus desarrolladores originales.
- Descripción del conjunto de datos de ajuste fino: origen, tamaño, cobertura demográfica, periodo de tiempo y pasos de preprocesamiento.
- Sesgos introducidos o heredados a través del ajuste fino, incluido el rendimiento de los subgrupos en el conjunto de datos local.
- Autoridad institucional designada responsable del mantenimiento continuo, la supervisión y la respuesta a incidentes.

Para las aplicaciones de Nivel 3, los equipos de desarrollo internos deben completar la plantilla completa de la Ficha del Modelo (Sección D.2) antes de cualquier implementación, incluidas las fases piloto.

Anexo F - Protocolo de respuesta a incidentes y desactivación de emergencia

Objetivo

Este anexo define los requisitos mínimos para la identificación, clasificación, escalado y desactivación de emergencia de incidentes en los sistemas de IA desplegados bajo los niveles 2 y 3 de AI-GUARD. Todos los sistemas de nivel 3 deben completar este protocolo antes del despliegue. Se recomienda encarecidamente a los sistemas de nivel 2 que lo adopten.

Clasificación de incidentes

Los siguientes tipos de eventos constituyen incidentes que deben notificarse:

- **Nivel 1 - Menor:** Errores aislados en los resultados sin impacto en el paciente o la población; corregidos mediante la intervención del usuario.
- **Nivel 2 - Moderado:** Errores sistemáticos que afectan a un grupo de población definido, interrupción del flujo de trabajo o indicios de un deterioro emergente del rendimiento (desviación de datos).
- **Nivel 3 - Crítico:** indicios de daño significativo al paciente, falsos negativos/positivos masivos que afectan a la respuesta de salud pública, resultados discriminatorios entre subgrupos de población o pérdida de la capacidad de supervisión humana.

Vía de escalamiento

Antes de la implementación, la institución responsable debe definir y documentar:

- **La autoridad responsable designada** para cada nivel de incidente.
- **Tiempo máximo de respuesta por nivel** (sugerido: N.° 1 = 5 días laborables; N.° 2 = 48 horas; N.° 3 = inmediato).
- **Vía de comunicación** con las partes interesadas afectadas y, cuando proceda, con las autoridades reguladoras nacionales.

Criterios de desactivación de emergencia (protocolo de reversión)

Las siguientes condiciones deben dar lugar a la suspensión inmediata del sistema en espera de la investigación y la corrección:

- Detección de un incidente de nivel 3 tal y como se define en E.2.
- Pérdida de la capacidad de intervención manual.
- Deterioro del rendimiento por debajo de los umbrales predefinidos (que se establecerán durante la evaluación de la preparación para el despliegue).
- Incidentes de nivel 2 no resueltos que persistan más allá del plazo de respuesta establecido.

La desactivación debe incluir un plan documentado de retorno al proceso manual para garantizar la continuidad de la atención o la función de salud pública durante el período de suspensión.

Requisitos de revisión posterior al incidente

Tras cualquier incidente de nivel 2 o nivel 3:

- Se debe completar y documentar el análisis de la causa raíz.
- Se requiere una reevaluación de AI-GUARD antes de la reasignación.
- La autoridad supervisora debe aprobar formalmente la reimplementación.

OPS
Organización Panamericana de la Salud
Organización Mundial de la Salud
Región de las Américas
BID
www.paho.org