



HUDERIA

Metodología y modelo

**Metodología para evaluar los riesgos e
impactos de los sistemas de inteligencia
artificial desde la perspectiva de los
derechos humanos, la democracia y el
Estado de Derecho**

**Modelo de recursos para el análisis de riesgos basado en el
contexto (COBRA)**

HUDERIA

Metodología y modelo

El Comité de Ministros
aprobó el modelo HUDERIA
el 25 de febrero de 2026
y la metodología HUDERIA
el 26 de febrero de 2025

Consejo de Europa

Edición en
francés: *HUDERIA –
Méthodologie et modèle* ISBN
978-92-871-9690-3
ISBN 978-92-871-9691-0 (PDF)

Las opiniones expresadas en este documento son responsabilidad de los autores y no necesariamente reflejan la política oficial del Consejo Europeo.

Se autoriza la reproducción de extractos (hasta 500 palabras), excepto para fines comerciales, siempre que se preserve la integridad del texto, que el extracto no se utilice fuera de contexto, que no proporcione información incompleta o que, de otra manera, induzca a error al lector sobre la naturaleza, el alcance o el contenido del texto. El texto original debe citarse siempre de la siguiente manera: «© Consejo de Europa, año de publicación». Todas las demás solicitudes relativas a la reproducción/traducción total o parcial del documento deberán dirigirse a la División de Publicaciones e Identidad Visual (PVID), Consejo de Europa (F-67075 Estrasburgo-Cedex) o publishing@coe.int.

Toda otra correspondencia relativa a este documento deberá dirigirse a la Secretaría del Comité Director para las Tecnologías Digitales Nuevas y Emergentes (CDNET)/ Departamento de Desarrollo Digital y Deporte del Consejo de Europa,
Consejo de Europa, F-67075 Estrasburgo-Cedex, Correo electrónico: secretariat.cdnet@coe.int

Fotografía: ©Shutterstock Portada y diseño gráfico: División de Publicaciones e Identidad Visual (DPIV)
ISBN 978-92-871-9692-7
ISBN 978-92-871-9693-4 (PDF)
© Consejo de Europa, abril de 2026
Impreso en el Consejo de Europa

La metodología HUDERIA fue creada por el Comité sobre Inteligencia Artificial (CAI) del Consejo de Europa para respaldar una evaluación coherente y práctica de los riesgos e impactos de los sistemas de IA desde la perspectiva de los derechos humanos, la democracia y el Estado de Derecho. Partiendo de conceptos ya asentados sobre los riesgos para los derechos humanos, traduce estas ideas en directrices adaptadas a la realidad del ciclo de vida de la IA, que ayudan a los actores públicos y privados a identificar, evaluar, prevenir y mitigar riesgos en los más diversos usos de la inteligencia artificial. HUDERIA combina directrices de alto nivel y tecnológicamente neutras (la metodología HUDERIA) con materiales de apoyo y adaptables (el modelo HUDERIA: recursos para el análisis de riesgos basado en el contexto (COBRA)), y ofrece herramientas y recomendaciones escalables que pueden ajustarse a distintos contextos, necesidades y niveles de capacidad. Como instrumento no vinculante jurídicamente, HUDERIA ofrece a las Partes en el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho una referencia flexible que puede utilizarse o adaptarse —en todo o en parte— junto con otros marcos o herramientas, para propiciar enfoques de evaluación de riesgos e impactos acordes con el Convenio Marco.

Índice

PARTE I - METODOLOGÍA PARA EVALUAR LOS RIESGOS E IMPACTOS DE LOS SISTEMAS DE IA DESDE LA PERSPECTIVA DE LOS DERECHOS HUMANOS, LA DEMOCRATIA Y EL ESTADO DE DERECHO 6

INTRODUCCIÓN 8

¿Qué es HUDERIA? 8

Relación con el Convenio Marco 8

Objetivos principales de HUDERIA 8

¿CUÁL ES EL ENFOQUE DE HUDERIA? 11

Enfoque sociotécnico 11

Directrices generales y específicas 11

Adaptabilidad y flexibilidad 11

Enfoque gradual y diferenciado 12

ESTRUCTURA DE HUDERIA 14

1. ANÁLISIS DE RIESGOS BASADO EN EL CONTEXTO (COBRA) 16

Introducción 16

Enfoque preliminar 16

Explicaciones 16

Análisis de los factores de riesgo 17

Explicaciones 18

Cartografía de los posibles impactos en los derechos humanos, la democracia y el Estado de Derecho; 18

Explicaciones 19

Determinación del nivel de riesgo 19

Selección 20

Enfoque adaptable de la selección 20

"Preguntas cero" 21

2. PROCESO DE PARTICIPACIÓN DE LAS PARTES INTERESADAS (SEP) 23

Introducción 23

Explicaciones 23

Análisis de las partes interesadas 23

Reflexión sobre la posición 23

Establecimiento de los objetivos de participación 24

Determinación del método de participación 24

Aplicación 25

3. EVALUACIÓN DE RIESGOS 26

Introducción 26

Explicaciones relativas a las preguntas e indicaciones de la evaluación de riesgos e impactos 26

Escala 27

Alcance 27

Reversibilidad 28

Probabilidad 28

Resultado de la evaluación de riesgos e impactos 28

4. PLAN DE MITIGACIÓN 30

Introducción 30

Explicaciones 30

Obligaciones jurídicas	30
Jerarquía de mitigación	30
Acceso a recursos	31
Resultado de este elemento del proceso HUDERIA	32
REVISIÓN ITERATIVA	34
Introducción	34
Factores de producción, aplicación y despliegue	34
Factores del entorno real	34
Aplicación de la revisión iterativa	35
PARTE II - MODELO DE RECURSOS PARA EL ANÁLISIS DE RIESGOS BASADO EN EL CONTEXTO (COBRA)¹³⁶	
PREFACIO	38
¿Qué es HUDERIA?	38
Relación con el Convenio Marco	38
Directrices generales y específicas	38
RECURSOS COBRA	41
Introducción	41
Panorámica de los recursos COBRA A, B y C	41
Recursos COBRA A (Lista de factores de riesgo que surgen en el contexto de aplicación del sistema de IA)	43
Recursos COBRA B (Lista de factores de riesgo que surgen en el contexto de diseño y desarrollo del sistema de IA)	47
Recursos COBRA C (Lista de factores de riesgo que surgen en el contexto de despliegue del sistema de IA)	52
Recursos COBRA E (Ámbitos ilustrativos de posible preocupación desde el punto de vista de los derechos humanos, la democracia y el Estado de Derecho)	53
Recursos COBRA F (Lista ilustrativa de sectores/esferas y ámbitos de posible preocupación)	68

¹ El Comité sobre Inteligencia Artificial (CAI) examinó la posible inclusión de un recurso COBRA D; no obstante, sin perjuicio de cualquier decisión futura sobre su inclusión, el modelo HUDERIA: recursos para el análisis de riesgos basado en el contexto, adoptado el 5 de noviembre de 2025, comprende los recursos COBRA A, B, C, E y F.

Parte I

METODOLOGÍA PARA EVALUAR LOS RIESGOS E IMPACTOS DE LOS SISTEMAS DE IA DESDE LA PERSPECTIVA DE LOS DERECHOS HUMANOS, LA DEMOCRACIA Y EL ESTADO DE DERECHO

Introducción

¿Qué es HUDERIA?

La evaluación de los riesgos e impactos de los sistemas de inteligencia artificial (IA) desde la perspectiva de los derechos humanos, la democracia y el Estado de Derecho («HUDERIA») es una guía que ofrece un enfoque estructurado para evaluar los riesgos e impactos de los sistemas de IA, específicamente diseñada para la protección y la promoción de los derechos humanos, la democracia y el Estado de Derecho. Está llamada a desempeñar un papel único y esencial en la intersección entre las normas internacionales en materia de derechos humanos y los marcos técnicos existentes sobre la gestión de riesgos en el contexto de la IA.

HUDERIA puede ser utilizada tanto por actores públicos como privados para ayudar a identificar y abordar los riesgos e impactos sobre los derechos humanos, la democracia y el Estado de Derecho a lo largo de todo el ciclo de vida de los sistemas de IA. Tiene su origen en los trabajos del Comité Ad Hoc sobre Inteligencia Artificial (CAHAI) (2019-2021) y, más concretamente, de su Grupo de Elaboración de Políticas, que encargó al Instituto Alan Turing —el instituto nacional de ciencia de datos e IA del Reino Unido— la elaboración de una propuesta original que hiciera operativas las líneas generales de un modelo de evaluación de impacto sobre los derechos humanos, la democracia y el Estado de Derecho. La metodología HUDERIA fue adoptada por el Comité sobre Inteligencia Artificial (CAI) del Consejo de Europa el 28 de noviembre de 2024.

Relación con el Convenio Marco

HUDERIA es un conjunto autónomo de orientaciones no jurídicamente vinculante que, como tal, carece de efectos jurídicos. No tiene carácter obligatorio y no pretende servir de instrumento interpretativo del Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, en lo sucesivo denominado «el Convenio Marco». Para contribuir a la gestión de los riesgos e impactos de la IA se pueden emplear numerosos marcos, políticas, directrices, normas o herramientas, existentes o futuros, incluido HUDERIA.

Las Partes en el Convenio Marco disponen de flexibilidad para utilizar o adaptar la guía, en su totalidad o en parte, con el fin de desarrollar nuevos enfoques para la evaluación de riesgos, o para utilizar o adaptar enfoques ya existentes de conformidad con su legislación aplicable, siempre que cumplan plenamente sus obligaciones en virtud del Convenio Marco, incluidos los requisitos mínimos de gestión de riesgos e impactos establecidos en su capítulo V.

Objetivos principales de HUDERIA

HUDERIA persigue los siguientes fines:

- contribuir a determinar en qué medida pueden resultar necesarias actividades de gestión de riesgos relacionadas con los derechos humanos, la democracia y el Estado de Derecho, y ofrecer una metodología de identificación, evaluación, prevención y mitigación de riesgos e impactos, que sea aplicable a una amplia variedad de tecnologías y contextos de aplicación de la IA, y que

responda a la evaluación futura de las tecnologías y aplicaciones de la IA;

- promover la compatibilidad y la interoperabilidad con las directrices, normas y marcos, existentes y futuros, elaborados por las organizaciones u órganos técnicos, profesionales y de otra índole pertinentes (como la Organización Internacional de Normalización (ISO), la Comisión Electrotécnica Internacional (IEC), la Unión Internacional de Telecomunicaciones (UIT), el Comité Europeo de Normalización (CEN), el Comité Europeo de Normalización en Electrónica y Electrotécnica (CENELEC), el Institute of Electrical and Electronics Engineers (IEEE), la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y el National Institute of Standards and Technology (NIST)), incluidos el Marco de Gestión de Riesgos de la IA del NIST y la gestión de riesgos y la evaluación de impacto sobre los derechos fundamentales previstas en el Reglamento de Inteligencia Artificial de la Unión Europea.

¿Cuál es el enfoque de HUDERIA?

HUDERIA combina los conocimientos actuales sobre los procesos y mecanismos de gobernanza técnica y sociotécnica que pueden facilitar actividades responsables dentro del ciclo de vida de los sistemas de IA con los procedimientos de diligencia necesarios para proteger y promover los derechos humanos, la democracia y el Estado de Derecho.

HUDERIA parte de variables, conceptos y un lenguaje bien conocidos para la evaluación de riesgos para los derechos humanos (escala, alcance, probabilidad y reversibilidad de los posibles impactos adversos en los derechos humanos). Su objetivo es facilitar su examen aportando orientaciones adicionales que tengan en cuenta la complejidad sociotécnica del ciclo de vida de la IA.

Enfoque sociotécnico

HUDERIA adopta un enfoque sociotécnico, que considera que todos los aspectos del ciclo de vida del sistema de IA se ven afectados por la relación interconectada entre la tecnología, las decisiones humanas y las estructuras sociales. Con este enfoque, la gestión de los riesgos e impactos de los sistemas de IA tiene en cuenta tanto sus aspectos técnicos como los contextos jurídicos, sociales, políticos, económicos, culturales y tecnológicos en los que operan. Un enfoque de este tipo promueve el desarrollo de una IA segura, fiable y de confianza que, además de tener un buen rendimiento, favorezca el respeto de los derechos humanos, la democracia y el Estado de Derecho.

Directrices generales y específicas

HUDERIA aporta estructura al combinar directrices generales y específicas, y flexibilidad al dejar margen para la adaptación en su aplicación práctica.

En el plano general, la metodología HUDERIA describe conceptos, procesos y elementos de alto nivel que orientan las actividades de evaluación de riesgos e impactos de los sistemas de IA que puedan repercutir en los derechos humanos, la democracia y el Estado de Derecho.

En el plano específico, el modelo HUDERIA proporciona materiales y recursos de apoyo (como herramientas flexibles pertinentes para los distintos elementos del proceso HUDERIA y recomendaciones escalables) que pueden facilitar la aplicación de la metodología HUDERIA. Estos recursos se mencionan a lo largo del texto y constituirán un conjunto de conocimientos que puede facilitar el examen de los riesgos e impactos relacionados con los derechos humanos, la democracia y el Estado de Derecho, también en el marco de otros enfoques de gestión de riesgos.

Adaptabilidad y flexibilidad

Tanto la metodología HUDERIA como el modelo HUDERIA dejan margen para la adaptación a distintos contextos, necesidades y capacidades, al establecer metas, principios y objetivos, dejando al mismo tiempo un margen de apreciación para decidir cómo alcanzarlos y ofrecer un abanico de opciones de política y de gobernanza que pueden adaptarse a contextos diversos.

Enfoque gradual y diferenciado

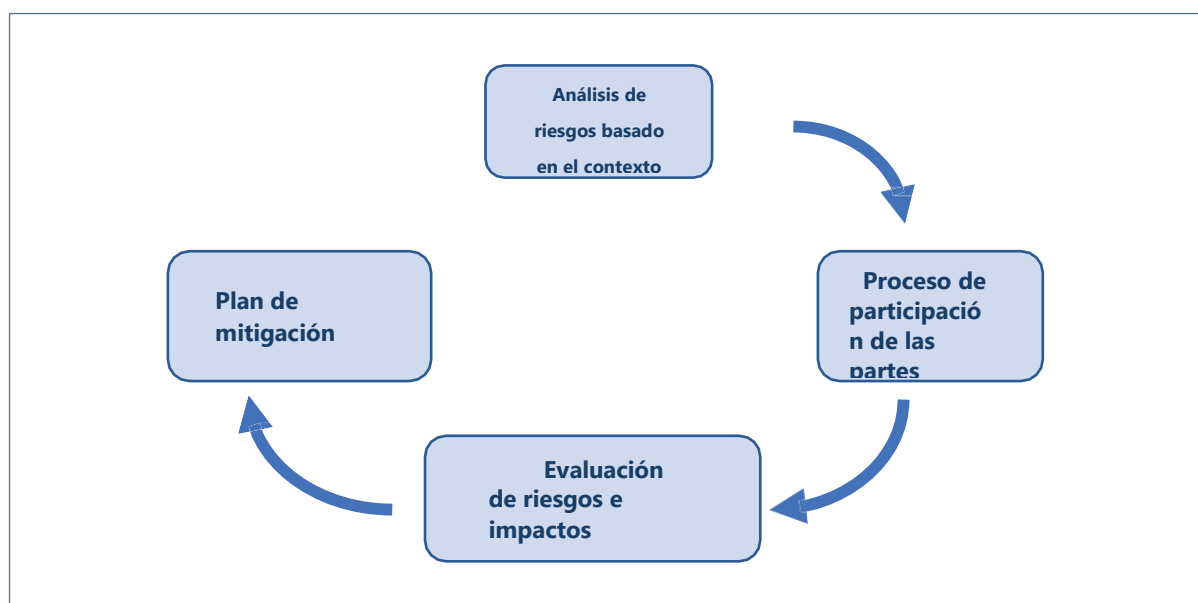
HUDERIA tiene por objeto establecer un enfoque gradual y diferenciado de las medidas de identificación, evaluación, prevención y mitigación de riesgos e impactos, que tenga en cuenta la gravedad y la probabilidad de que se produzcan impactos adversos sobre los derechos humanos, la democracia y el Estado de Derecho, así como los factores contextuales pertinentes.

Estructura de HUDERIA

La metodología HUDERIA consta de cuatro elementos:

1. El **análisis de riesgos basado en el contexto** (COBRA) ofrece un enfoque estructurado para recopilar y cartografiar la información necesaria para identificar y comprender los riesgos que un sistema de IA podría suponer para los derechos humanos, la democracia y el Estado de Derecho, atendiendo a su contexto sociotécnico. Contribuye además a determinar en un primer momento si el sistema de IA es la solución adecuada al problema planteado.
2. El **proceso de participación de las partes interesadas** (SEP) propone un enfoque para posibilitar y hacer operativa la participación, cuando proceda, de las partes interesadas pertinentes, con el fin de recabar información sobre las personas potencialmente afectadas y de contextualizar y corroborar los posibles daños y las medidas de mitigación;
3. La **evaluación de riesgos e impactos** ofrece posibles pasos para evaluar los riesgos e impactos relacionados con los derechos humanos, la democracia y el Estado de Derecho.
4. El **plan de mitigación** ofrece posibles pasos para definir las medidas de mitigación y reparación, incluidos el acceso a recursos y la revisión iterativa.

Aunque lo lógico es llevar a cabo en primer lugar el elemento COBRA, en función de las necesidades y los enfoques puede optarse por alterar la secuencia de los elementos y/o aplicar o utilizar de otro modo únicamente determinadas partes de la metodología, atendiendo a los enfoques existentes en materia de gobernanza de la IA y a los contextos, necesidades y capacidades concretos.



1. Análisis de riesgos basado en el contexto (COBRA)

Introducción

COBRA ayuda a identificar los distintos factores de riesgo —características o propiedades de un sistema de IA y de su contexto que inciden en la probabilidad de que se produzcan impactos adversos en los derechos humanos, la democracia y el Estado de Derecho—. Estos factores no deben considerarse necesariamente como causas de impactos adversos, sino más bien como condiciones que guardan relación con una mayor probabilidad de daño y que, por tanto, es preciso anticipar y tener en cuenta en las labores de gestión de riesgos y mitigación de impactos. Los factores de riesgo se clasifican en tres grandes ámbitos: el contexto de aplicación del sistema, su contexto de diseño y desarrollo, y su contexto de despliegue.

El examen de los factores de riesgo tiene como objetivo facilitar la cartografía de los posibles impactos adversos en los derechos humanos, la democracia y el Estado de Derecho. Los resultados de este análisis cartográfico de factores de riesgo e impacto están destinados a orientar el alcance del enfoque de los elementos posteriores de HUDERIA, entre otras cosas, estableciendo la proporcionalidad de las actividades HUDERIA subsiguientes.

Los resultados de este análisis cartográfico de factores de riesgo e impacto pueden ayudar, además, a precisar cuáles son los contextos sociotécnicos concretos, a lo largo del ciclo de vida del sistema, que requieren de una atención específica en materia de gobernanza.

El elemento COBRA consta de cuatro pasos:

1. alcance preliminar;
2. análisis de los factores de riesgo;
3. cartografía de los posibles impactos en los derechos humanos, la democracia y el Estado de Derecho;
4. selección.

Enfoque preliminar

Objetivos

La finalidad principal de esta etapa es llevar a cabo la investigación preliminar de base necesaria para orientar las actividades posteriores de identificación de los factores de riesgo y de la cartografía de impactos.

Explicaciones

El proceso COBRA comienza con la labor de alcance preliminar, en la que:

- se describen a grandes rasgos la finalidad del sistema, sus componentes principales, los contextos en los que está previsto utilizarlo, el ámbito o ámbitos en qué operará, el grado de intervención humana, así como la naturaleza y el volumen de los datos que tratará y con los que será entrenado, y se deja constancia de las comprobaciones que se hayan podido efectuar

previamente para evaluar el sesgo en el conjunto de datos o en el modelo.

- ▶ se identifican a las personas o los grupos que podrían verse afectados por el sistema, o que puedan influir en él, centrándose en las características contextuales pertinentes de las personas y los grupos identificados, incluidas las características protegidas y los factores de vulnerabilidad; y se efectúa una delimitación preliminar de los posibles impactos adversos en los derechos humanos, la democracia y el Estado de Derecho mediante el examen de los ámbitos de preocupación ilustrativos²

-
- ▶ y se ofrece una cartografía inicial de las roles y responsabilidades a lo largo del ciclo de vida del sistema de IA.³

Esta actividad de alcance preliminar podría basarse en documentos de la organización (por ejemplo, la justificación económica del proyecto, la prueba de concepto o el acta constitutiva del proyecto), en la colaboración y en la investigación documental (si fuera necesario). Esta actividad de alcance preliminar, así como los elementos posteriores del proceso HUDERIA, deberían desarrollarse, cuando proceda, en el seno de un equipo multidisciplinar integrado por expertos con especializaciones complementarias diversas⁴ y con perfiles tanto técnicos como no técnicos.

Análisis de los factores de riesgo

Objetivos

La finalidad principal de esta etapa es recabar la información pertinente sobre los factores de riesgo relacionados con el contexto de aplicación previsto del sistema, su contexto de diseño y desarrollo y su contexto de despliegue. Estos factores de riesgo facilitarán la cartografía de los posibles impactos adversos en los derechos humanos, la democracia y el Estado de Derecho, así como la subsiguiente evaluación de las variables de riesgo fundamentales: gravedad (escala, alcance y reversibilidad)⁵ y probabilidad.

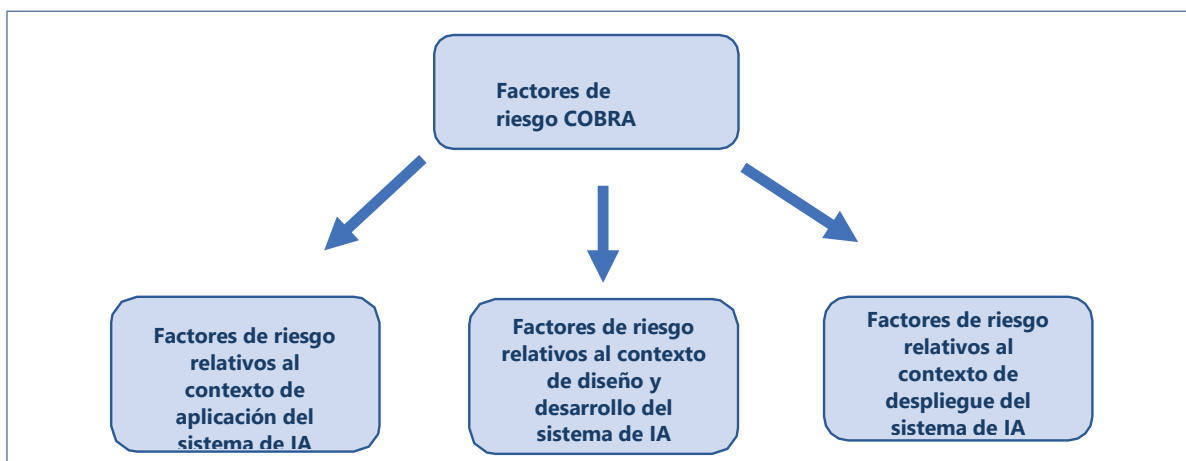
² Los recursos COBRA E (ámbitos ilustrativos de posible preocupación desde el punto de vista de los derechos humanos, la democracia y el Estado de Derecho) constituyen una herramienta que podría utilizarse para realizar u orientar esta evaluación.

³ La sección «Roles y responsabilidades» del modelo HUDERIA [que elaborará y adoptará el CDNET en 2026] ofrecerá orientaciones en relación con este aspecto de la metodología.

⁴ Los conocimientos especializados pertinentes pueden abarcar, según proceda, cuestiones de derechos humanos, privacidad y protección de datos personales, ciencia de datos, gestión de conjuntos de datos, seguridad, riesgos de la IA y pruebas, evaluación, verificación y validación de la IA.

⁵ Siguiendo el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos y los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos, el a los efectos de HUDERIA se entiende que el término «gravedad» se compone de una combinación de las variables de escala, alcance y reversibilidad.

Explicaciones



Los sistemas de IA se diseñan, desarrollan y utilizan en una gran variedad de contextos y de muchas formas diferentes, por lo que resulta importante realizar una evaluación holística de diversos factores relacionados con el contexto de aplicación del sistema, su contexto de diseño y de desarrollo, y su contexto de despliegue.

El **contexto de aplicación del sistema** de IA⁶ comprende información sobre el sector y al ámbito de aplicación del sistema, los entornos legales y reglamentarios en los que se desarrolla y utiliza, la finalidad prevista del sistema y otros datos pertinentes de su contexto de aplicación, como los antecedentes conocidos de sesgo y discriminación.

El **contexto de diseño y desarrollo del sistema** de IA⁷ comprende las características técnicas pertinentes del sistema. Puede incluir las limitaciones conocidas del sistema, consideraciones relativas a la recogida, el enriquecimiento, el almacenamiento, el uso y la retirada de datos, así como consideraciones relativas al propio algoritmo o modelo. Revisten particular relevancia las características técnicas relacionadas con la privacidad y la protección de datos, el sesgo y la discriminación, así como la explicabilidad y la interpretabilidad.

El **contexto de despliegue del sistema** de IA⁸ comprende los factores que determinan cómo pueden manifestarse y gestionarse en la práctica los riesgos potenciales, como las medidas que se adoptarán para proteger la privacidad y los datos personales, mitigar el sesgo nocivo, garantizar una formación adecuada, prevenir usos no previstos y asegurar la rendición de cuentas y el cumplimiento de la legalidad.

Cartografía de los posibles impactos en los derechos humanos, la democracia y el Estado de Derecho;

Objetivos

En esta etapa de la cartografía se identifican a las personas o grupos que podrían verse afectados y se lleva a cabo una evaluación inicial de las principales variables de riesgo: gravedad (magnitud, alcance, reversibilidad) y probabilidad. La cartografía contribuye a orientar los elementos posteriores de la metodología y el alcance de la intervención de la gobernanza y de las medidas de mitigación que pueden resultar adecuadas (véase la sección «Selección», más adelante). El análisis de las variables de

⁶ Los recursos COBRA A (lista de factores de riesgo que surgen en el contexto de aplicación del sistema) constituyen una herramienta que podría utilizarse para realizar u orientar esta evaluación.

⁷ Los recursos COBRA B (lista de factores de riesgo que surgen en el contexto de diseño y desarrollo del sistema) constituyen una herramienta que podría utilizarse para realizar u orientar esta evaluación.

⁸ Los recursos COBRA C (lista de factores de riesgo que surgen en el contexto de despliegue del sistema) constituyen una herramienta que podría utilizarse para realizar u orientar esta evaluación.

riesgo fundamentales resulta esencial para obtener una visión de conjunto clara y estructurada de dónde es más probable que se produzcan las amenazas y de su impacto potencial.

Explicaciones

Los recursos COBRA E y F podrían utilizarse para determinar los ámbitos de aplicación potencialmente sensibles y los ámbitos de preocupación potencialmente pertinentes en relación con los derechos humanos, la democracia y el Estado de Derecho.⁹

Sirviéndose de la información recabada en los pasos anteriores:

- a. se determina si el sistema operará en el marco de una o varias actividades (decisiones o acciones) que puedan producir impactos en las personas afectadas en los sectores o ámbitos pertinentes;
- b. se identifican y se enumeran los ámbitos de preocupación pertinentes¹⁰ y, teniendo esto presente, se responde a la pregunta de si el sistema puede tener impactos potenciales o reales en derechos humanos concretos, la democracia y el Estado de Derecho;
- c. para cada impacto potencial y real identificado, se describe la naturaleza de dicho impacto,¹¹ teniendo en cuenta los impactos diferenciados sobre las personas y los grupos afectados en función de las características contextuales pertinentes, incluidas las características protegidas y los factores de vulnerabilidad.

El análisis de estos puntos aportará información para la evaluación inicial de las variables de riesgo fundamentales:

- la gravedad (escala, alcance y reversibilidad) y la probabilidad, que ayudan a determinar el riesgo y a elegir el enfoque adecuado para los elementos posteriores de la metodología, lo que contribuirá a garantizar que las intervenciones de gobernanza y las medidas de mitigación se ajusten a las necesidades a lo largo de todo el ciclo de vida del sistema de IA.

Los resultados de este análisis también pueden ayudar a detectar oportunidades de utilizar el sistema de IA con el fin de respaldar actuaciones positivas que promuevan los derechos humanos, incluidas la promoción y la garantía de la no discriminación.

Determinación del nivel de riesgo

Para indexar el nivel de riesgo de cada uno de los posibles impactos adversos en los derechos humanos, la democracia y el Estado de Derecho que se hayan identificado como resultado del ejercicio de cartografía pueden emplearse las variables siguientes:

1. La **escala**¹² de los posibles impactos adversos (la gravedad del daño potencial);
2. El **alcance** de los posibles impactos adversos (incluidos el número de personas afectadas, las características protegidas o la vulnerabilidad de las personas o los grupos y el horizonte temporal de los impactos).
3. La **reversibilidad**¹³ de los posibles impactos adversos es la información sobre la posible

⁹ Las referencias que se hacen en los recursos COBRA a los derechos humanos consagrados en diversos instrumentos internacionales de derechos humanos se incluyen a título ilustrativo. Dichas referencias solo son aplicables a los Estados que son partes en esos instrumentos. Se espera que cada Estado aplique su propia legislación aplicable y sus obligaciones internacionales.

¹⁰ Los recursos COBRA E y F podrían utilizarse para determinar los posibles ámbitos de preocupación relacionados con los derechos humanos, la democracia y el Estado de Derecho, así como para cartografiar los sectores y ámbitos pertinentes.

¹¹ Para mayor claridad de la evaluación deberían tenerse en cuenta tanto los impactos adversos o restrictivos como los beneficiosos, potenciadores o positivos de otro modo producidos por los sistemas de IA, ya que pueden plantearse diversas cuestiones de sesgo y discriminación respecto de sistemas que producen ambos tipos de impacto.

¹² El término «escala» (*scale*) se denomina en ocasiones «gravedad» (*gravity*) en los contextos de evaluación de riesgos.

¹³ El término «reversibilidad» (*reversibility*) se denomina en ocasiones «remediabilidad» (*remediability*) en los contextos de

reparabilidad o restitución de las personas afectadas a su situación anterior al impacto o a una situación equivalente.

4. La **probabilidad**¹⁴ de los posibles impactos adversos.

Los equipos competentes deberían repasar cada uno de los impactos potenciales que se hayan identificado y examinar, para cada ámbito de preocupación relacionado con los derechos humanos, la democracia y el Estado de Derecho, y para cada grupo afectado, la escala, el alcance, la reversibilidad y la probabilidad del efecto adverso potencial o real. La legislación o las políticas nacionales pueden ofrecer definiciones más detalladas que sirvan para orientar esta determinación del riesgo y para determinar enfoques adecuados y proporcionados de las actividades HUDERIA posteriores (como la participación de las partes interesadas).

Cabe plantearse establecer un método de combinación de estas variables que permita calibrar el riesgo y determinar enfoques adecuados y proporcionados de las actividades HUDERIA posteriores (como la participación de las partes interesadas), así como del alcance y la profundidad de las intervenciones de gobernanza y de las medidas de gestión de riesgos y de mitigación. Ello puede suponer la formulación de métodos cuantitativos o semicuantitativos de cálculo del riesgo, de matrices de riesgo o procedimientos de carácter más cualitativos o basados en reglas. Todo mecanismo de calibración del riesgo resultante de la combinación de estas variables a efectos de la evaluación HUDERIA podrá tener en cuenta:

- ▶ en lo que respecta a los derechos humanos, los efectos de bajo alcance y elevada gravedad, así como los efectos de elevado alcance y escasa gravedad, sobre cada persona afectada;
- ▶ en lo que respecta a la democracia y al Estado de Derecho, los efectos de elevado alcance y de larga duración sobre las personas, las instituciones y la sociedad en general.

Selección

Objetivos

La finalidad principal de esta etapa es partir de la información recabada en las actividades COBRA anteriores y, por tanto:

- ▶ facilitar la identificación y la clasificación de los sistemas que presentan un riesgo significativo, de modo que la metodología HUDERIA no resulte onerosa para los sistemas de IA con riesgo mínimo o bajo;
- ▶ efectuar una determinación inicial de si el sistema de IA debería desarrollarse o desplegarse, en función de si los beneficios de desarrollarlo o desplegarlo superan sus riesgos, atendiendo en particular a sus posibles impactos en los derechos humanos, la democracia y el Estado de Derecho, así como de si el uso del sistema de IA es incompatible con el respeto de los derechos humanos, la democracia y el Estado de Derecho.

Enfoque adaptable de la selección

Las actividades previas de esta etapa ofrecen una imagen preliminar del perfil de riesgo del sistema de IA.

La información recopilada puede bastar, por ejemplo, para determinar que es improbable que el sistema tenga impacto alguno en los derechos humanos, la democracia o el Estado de Derecho, lo que hace que innecesarios los elementos posteriores de HUDERIA. Se podría llegar a una conclusión similar si los impactos identificados fueran insignificantes o improbables. Si los impactos identificados llevan a decidir que un sistema de IA no se desarrollará ni se desplegará por considerarse incompatible con el respeto de los derechos humanos, la democracia y el Estado de Derecho, los elementos posteriores

evaluación de riesgos.

¹⁴ El término «probabilidad» (*probability*) se denomina en ocasiones «verosimilitud» (*likelihood*) en los contextos de evaluación de riesgos.

de HUDERIA también resultan innecesarios. Por último, cuando se detecten impactos potenciales graves, puede estar justificado recurrir a un abanico de estrategias y respuestas de gestión de riesgos (incluido el proceso de participación de las partes interesadas descrito en la sección siguiente). Para hacer frente a esta complejidad, HUDERIA no prescribe directrices detalladas para ajustar las tareas de gestión de riesgos, sino que se limita a exponer los elementos que se proponen y que pueden aplicarse según proceda en función del riesgo de los impactos adversos en los derechos humanos, la democracia y el Estado de Derecho. Pueden aplicarse diferentes enfoques para determinar las medidas de gestión de riesgos a partir de los impactos potenciales y reales identificados —o de una combinación de ambos— en función del marco o entorno normativo nacional concreto, del sector, del sistema y del contexto (por ejemplo, enfoques basados en umbrales, escenarios, proporcionalidad, dinámicos o específicos del contexto).

La determinación final de si debe emplearse un método cualitativo, cuantitativo, mixto o de cualquier otro tipo queda a la discreción de las autoridades o, en su caso, de los equipos de proyecto de IA responsables del sistema.

“Preguntas cero”

Para ayudar a determinar si los beneficios de construir o desplegar el sistema de IA —incluidos los beneficios sociales adicionales que puedan derivarse de un uso del sistema que vaya más allá de su finalidad principal— superan sus riesgos, a la vista de los factores de riesgo y de los impactos potenciales identificados, considérese:

- ▶ si el uso del sistema es adecuado a la vista de la naturaleza del problema que el sistema de IA pretende resolver, es decir, en qué medida las tecnologías y los procesos ya existentes para resolver el problema considerado están en mejores condiciones de hacerlo, atendiendo al perfil de riesgo y a los posibles impactos adversos del sistema previsto, con especial atención, cuando proceda, a todo riesgo marginal que se añada al introducir la IA en el contexto actual;
- ▶ en qué medida el sistema previsto podrá satisfacer las necesidades y expectativas del responsable del despliegue;
- ▶ en qué medida los impactos del sistema previsto serán equitativos entre los distintos grupos afectados;
- ▶ en qué medida la calidad y la representatividad de los datos disponibles en la actualidad o potencialmente disponibles son suficientes para que el sistema previsto sea eficaz y seguro y evite de forma razonable un posible sesgo nocivo;
- ▶ en qué medida se dispone de recursos suficientes (humanos y materiales) capaces de satisfacer los requisitos técnicos y de llevar a cabo las actuaciones técnicas y de gobernanza necesarias para mitigar adecuadamente los riesgos identificados;
- ▶ los posibles contextos de uso del sistema y los riesgos de uso indebido o abusivo, incluso mediante su despliegue más allá de la finalidad prevista.

2. Proceso de participación de las partes interesadas (SEP)

Introducción

Con el fin de mejorar la calidad de la información destinada al siguiente elemento de HUDERIA —la evaluación de riesgos e impactos— cabe plantearse la decisión de llevar a cabo este paso incorporando los puntos de vista de las personas potencialmente afectadas, incluidas las que se encuentran en situación de vulnerabilidad.

La participación de las partes interesadas, tal como se contempla en la metodología HUDERIA, puede adoptar diversas formas. El nivel de participación de las personas afectadas debería determinarse en función de los factores de riesgo y de los impactos potenciales y reales identificados durante la etapa COBRA. Hacer participar a las partes interesadas a lo largo del ciclo de vida del sistema de IA puede aportar además otros muchos beneficios, como fomentar la transparencia, generar confianza y mejorar la usabilidad y el rendimiento del sistema de IA.

Explicaciones

El SEP comprende cinco pasos fundamentales:¹⁵ el análisis de las partes interesadas, la reflexión sobre la posición, el establecimiento de objetivos de participación, la determinación del método de participación, y la ejecución.

Análisis de las partes interesadas

En el análisis de las partes interesadas se determinan los grupos de partes interesadas que pueden verse afectados por las actividades desarrolladas dentro del ciclo de vida del sistema, o puedan influir en ellas. Dicho análisis¹⁶ evalúa los intereses, derechos, vulnerabilidades potenciales y existentes y ventajas relativos de las partes interesadas identificadas, así como la relevancia de los grupos de las partes interesadas identificados. En este paso, se debe considerar la posibilidad de incorporar de manera efectiva los puntos de vista de quienes:

1. están expuestos de forma desproporcionada a los riesgos derivados del uso del sistema;
2. son especialmente vulnerables a los daños potenciales;
3. tienen una capacidad especialmente limitada para influir en el modo en que se diseña y se utiliza el sistema (por ejemplo, los grupos marginados, desfavorecidos o infrarrepresentados en la actualidad o históricamente, o las personas en situación de vulnerabilidad o con necesidades específicas).

Reflexión sobre la posición

El paso siguiente consiste en reflexionar sobre la posición adoptada en relación con las partes

¹⁵ El proceso descrito en esta sección tiene carácter ilustrativo, ya que la determinación final sobre el proceso de participación de las partes interesadas queda a la discreción de las autoridades o, en su caso, de los equipos de proyecto de IA responsables del sistema.

¹⁶ Los recursos SEP (lista de preguntas para evaluar la relevancia relativa de las partes interesadas) [que elaborará y adoptará el CDNET en 2026] ofrecen preguntas detalladas y herramientas para orientar la identificación de las partes interesadas pertinentes.

interesadas afectadas, con el fin de reconocer las limitaciones de la perspectiva de los usuarios de HUDERIA y de detectar los puntos de vista ausentes que reforzarían la evaluación de los impactos potenciales y reales del sistema. En función de los factores de riesgo pertinentes y los impactos potenciales identificados durante el elemento COBRA, esto podría incluir una evaluación de las características demográficas autodeclaradas de los usuarios de HUDERIA, de su educación y formación, de su origen socioeconómico, así como del contexto institucional y de equipo.

Las principales cuestiones sobre las que deberían reflexionar los usuarios de HUDERIA al abordar esta etapa de la metodología son las siguientes:

- ▶ ¿En qué medida mis características personales, mis identificaciones grupales, mi situación socioeconómica, mi trayectoria educativa, formativa y laboral, la composición de mi equipo y mi marco institucional constituyen fuentes de poder y de ventaja, o bien fuentes de marginación y de desventaja?
- ▶ ¿Cómo influye esta posición en mi capacidad y en la de mi equipo para identificar y comprender a las partes interesadas afectadas y los posibles impactos del sistema de IA?

En función de los factores de riesgo identificados durante el proceso COBRA, los usuarios de HUDERIA deberían plantearse asimismo recurrir a partes interesadas externas o a consultores con conocimientos especializados concretos, como conocimientos en derecho de los derechos humanos, en relación con los impactos potenciales y reales del sistema en los derechos humanos.

Establecimiento de los objetivos de participación

La fijación de objetivos claros para la participación de las partes interesadas tiene por finalidad crear un entendimiento nítido de cómo y por qué se llevan a cabo las actividades de participación. Estos objetivos facilitan la intervención inclusiva, informada y efectiva de las personas potencialmente afectadas.¹⁷

Determinación del método de participación

La determinación del método o los métodos apropiados de participación de las partes interesadas exige evaluar y atender las necesidades de las personas potencialmente afectadas, tomando en consideración, según proceda, los resultados del proceso COBRA y otros factores pertinentes, como las limitaciones de recursos, las dificultades para llegar a grupos aislados o socialmente excluidos, las limitaciones de capacidad – como los obstáculos derivados de las brechas digitales o de las lagunas de información –, los plazos, etc.¹⁸

Los criterios siguientes pueden servir de orientación para el elemento SEP.

1. **Participación:** la intervención significativa de las personas afectadas o potencialmente afectadas se integra en los elementos pertinentes del proceso.
2. **Igualdad y prohibición de la discriminación:** los procesos de participación y consulta son inclusivos, tienen en cuenta la perspectiva de género y atienden a las necesidades de las personas y los grupos con características protegidas o que pueden hallarse en riesgo de vulnerabilidad o de marginación.
3. **Empoderamiento:** se tienen en cuenta la adecuación a la edad y a las necesidades de accesibilidad, y se desarrollan las capacidades de las personas y los grupos con características protegidas o que puedan hallarse en riesgo de vulnerabilidad o de marginación, con el fin de garantizar su intervención efectiva.
4. **Transparencia:** se prevé el intercambio de información significativa e inteligible entre las partes

¹⁷ Los recursos SEP [que elaborará y adoptará el CDNET en 2026] ofrecen preguntas detalladas a título indicativo y la descripción de las opciones de participación de las partes interesadas.

¹⁸ Los recursos SEP (Ejemplos de métodos de participación pertinentes con las preguntas correspondientes) [que elaborará y adoptará el CDNET en 2026] ofrecen posibles ejemplos de métodos de participación pertinentes y una lista de preguntas pertinentes que pueden ayudar a determinar los grupos de partes interesadas adecuados.

interesadas a intervalos pertinentes y regulares, se pone a disposición información sobre la comprensión de las posibles implicaciones y de los impactos en los derechos humanos, cuando proceda, y se comunican públicamente las conclusiones de HUDERIA y los planes de gestión de impactos (planes de acción).

5. **Rendición de cuentas:** la responsabilidad de la aplicación, el seguimiento y el control de las medidas de mitigación se atribuye a entidades, personas o funciones concretas dentro de la organización.

Aplicación

Una vez completadas las cuatro actividades precedentes, puede iniciarse el proceso de participación apropiado. Esto debería ser coherente con los resultados del análisis de las partes interesadas, la reflexión sobre la posición y los objetivos y métodos de participación establecidos, y debería documentarse adecuadamente.

3. Evaluación de riesgos e impactos

Introducción

La evaluación de riesgos e impactos tiene por finalidad ofrecer valoraciones detalladas de los impactos potenciales y reales que las actividades desarrolladas dentro del ciclo de vida de un sistema de IA podrían tener en los derechos humanos, la democracia y el Estado de Derecho.

De acuerdo con la selección efectuada en la fase COBRA, llevar a cabo la evaluación de riesgos e impactos reviste especial importancia en el caso de los sistemas de IA que puedan plantear riesgos significativos para los derechos humanos, la democracia y el Estado de Derecho. Tras la selección del análisis COBRA, este paso puede resultar necesario únicamente para determinados sistemas de IA, en particular aquellos que se consideren que presentan riesgos significativos para los derechos humanos, la democracia y el Estado de Derecho.

La evaluación de riesgos e impactos tiene por objeto:

- ▶ reexaminar, contextualizar y corroborar los daños potenciales y reales identificados en la fase COBRA;
- ▶ identificar y analizar otros daños potenciales y reales mediante una reflexión ampliada que permita precisar lagunas en la exhaustividad y la integridad de los ejemplos de daño enumerados previamente;
- ▶ evaluar las variables de riesgo de escala, alcance, reversibilidad y probabilidad de los posibles impactos adversos, de modo que sus riesgos puedan valorarse mejor para priorizarlos, gestionarlos y mitigarlos posteriormente.

La evaluación de riesgos e impactos parte de la identificación inicial de los factores de riesgo, basados en el contexto, para los derechos humanos, la democracia y el Estado de Derecho, así como de la cartografía de los posibles impactos en los derechos humanos, la democracia y el Estado de Derecho, llevadas a cabo en la fase COBRA, y de los posibles conocimientos aportados por el SEP, con el fin de abordar los impactos potenciales y reales del sistema de IA.

Ello se hace de manera efectiva mediante un proceso en dos pasos que permite elaborar un plan de mitigación y establecer el acceso a recursos durante el paso siguiente de la metodología HUDERIA.

Explicaciones relativas a las preguntas e indicaciones de la evaluación de riesgos e impactos

Introducción

La evaluación de riesgos e impactos en el contexto de HUDERIA se organiza en dos etapas.

En la primera etapa, la atención se centra en identificar los impactos potenciales y, más concretamente, "cómo" podrían producirse los impactos potenciales y reales identificados en las fases COBRA y SEP, lo que permite adoptar un enfoque más abierto y exploratorio que posibilita un análisis más profundo de los contextos concretos y del alcance, la escala y la reversibilidad de los impactos, en particular en lo que respecta a las personas en situación de vulnerabilidad o a los grupos vulnerables.

En la segunda fase, se procede a evaluar las variables de riesgo de escala, alcance, reversibilidad y probabilidad de los impactos potenciales o reales identificados. Un examen minucioso y sensible al contexto de estas variables ayuda a priorizar las actuaciones de mitigación al diferenciar la gravedad de los impactos del sistema de IA.

Escala

La escala de un impacto adverso potencial y real se refiere a la gravedad de la consecuencia esperada del daño potencial.

El examen de la gravedad de todo daño potencial debería incluir una reflexión sobre las diferentes formas y los distintos grados en que las personas o los grupos (en particular quienes poseen características que podrían hacerlos más vulnerables al impacto adverso) podrían sufrir ese daño. Las consideraciones sobre la escala deberían tener en cuenta las preguntas adicionales siguientes.

1. Para cada impacto adverso potencial y real identificado, ¿hay personas o grupos que presenten características que puedan hacerlos más vulnerables a ese impacto? En caso afirmativo, ¿cuáles son esas características y podrían quienes las poseen sufrir el daño de forma más aguda o grave que los demás?
2. Para cada impacto adverso potencial y real identificado, ¿qué personas o grupos podrían sufrir el impacto más grave derivado del daño considerado?

Las respuestas a estas preguntas cumplirán posteriormente una función importante en la etapa de planificación de la mitigación, cuando se examinen la reparación y la priorización de los daños potenciales.

Alcance

El alcance de un impacto adverso potencial y real se refiere a la estimación tanto del número de personas afectadas como del horizonte temporal de los impactos.

Las estimaciones del alcance de los impactos adversos potenciales y reales identificados se analizan una por una, prestando especial atención a los niveles de exposición al daño de determinados grupos de personas afectadas, así como a los impactos acumulativos o agregados del sistema sobre las personas y los grupos de personas potencialmente afectados, presentes y futuros.

Las consideraciones sobre el alcance pueden incluir el examen de estas preguntas.

1. Para el impacto adverso potencial y real identificado, ¿hay grupos que posean características que podrían hacerlos vulnerables a niveles más elevados de exposición^{19,20} a ese impacto? En caso afirmativo, ¿a cuánta exposición podrían enfrentarse esos grupos?
2. Para el impacto adverso potencial y real identificado, examínese la escala temporal global de los impactos del sistema de IA en el derecho o el ámbito de preocupación considerado (en el caso de la democracia o el Estado de Derecho). ¿Existen impactos acumulativos o agregados del sistema sobre las personas afectadas y las futuras personas afectadas que pudieran ampliar los impactos del sistema más allá del alcance del impacto ya identificado?
3. Algunas preguntas "de conjunto" sobre las que reflexionar al evaluar los impactos acumulativos o agregados pueden ser las siguientes.
 - ¿Podrían el suministro y el uso del sistema contribuir a impactos adversos más amplios en los derechos humanos, la democracia o el Estado de Derecho cuando su despliegue se coordine con el de otros sistemas que cumplen funciones o finalidades similares (o se produzca en paralelo a él)?

¹⁹ Los recursos SEP A y B [que elaborará y adoptará el CDNET en 2026] ofrecen preguntas detalladas (lista de preguntas para evaluar la relevancia relativa de las partes interesadas) y la descripción de los formatos de participación de las partes interesadas (lista de factores que determinan los objetivos y los niveles de participación de las partes interesadas) que pueden ayudar a los equipos de proyecto a determinar los grupos de partes interesadas especialmente pertinentes y los objetivos de la participación.

²⁰ El término «nivel de exposición» se entiende aquí como la proporción de un grupo que se ve afectada negativamente por un sistema de IA: cuando la fracción afectada del grupo es pequeña, sus miembros presentan niveles de exposición bajos, y cuando la fracción afectada es muy grande, sus miembros presentan niveles de exposición altos. A modo de ejemplo, los miembros de un grupo caracterizado por una situación socioeconómica baja pueden presentar un nivel de exposición alto a los posibles impactos adversos de un modelo de IA que se utilice para asignar prestaciones públicas.

- ¿Podrían el suministro y el uso del sistema replicar, reforzar o amplificar daños heredados o características inherentes socio-históricamente arraigados, de manera que pudieran generar efectos en cadena para las personas y los grupos afectados?
- ¿Cabría entender que el suministro y el uso del sistema contribuyen a impactos adversos agregados más amplios (sobre el medio ambiente o la salud pública, por ejemplo) cuando su despliegue se considere en combinación con el de otros sistemas que puedan tener impactos similares?

Reversibilidad

Como se ha explicado anteriormente, la reversibilidad se refiere a la información sobre el grado de reparabilidad o de restitución posible para las personas potencialmente afectadas como resultado de las medidas dirigidas a superar el impacto adverso considerado y a restituir a las personas afectadas a una situación similar o equivalente a la que tenían antes del impacto. Al igual que ocurre con las consideraciones relativas a la escala de un impacto potencial, llegar a comprender hasta qué punto es reversible un daño dependerá tanto del conocimiento del contexto concreto del daño como del de las personas afectadas que lo sufren. Determinar el grado de reversibilidad de un impacto adverso potencial exigen valorar el esfuerzo necesario para superar y (en su caso) revertir el daño.

Los miembros de distintos grupos pueden necesitar niveles de esfuerzo diferentes para superar los impactos adversos, en función de su edad, su posición en la sociedad y las circunstancias del daño (pues los grupos vulnerables y marginados suelen tener menos resiliencia que otros grupos dominantes, privilegiados o mayoritarios).

Probabilidad

Evaluar la probabilidad de un riesgo supone estimar la verosimilitud de que se produzca un determinado impacto adverso, sobre la base, según proceda, del juicio cualitativo, el análisis cuantitativo y la comprensión del contexto.

Determinar el nivel de probabilidad de un riesgo implica un análisis amplio de las condiciones contextuales y operativas y viene determinado por lo general por el nivel (tipo, cantidad y calidad) de información de que el riesgo pueda llegar a materializarse. Así se garantiza que las evaluaciones de riesgos se sustenten tanto en datos como en el criterio de los expertos, lo que facilita la priorización y la mitigación de los riesgos potenciales.

Resultado de la evaluación de riesgos e impactos

Una vez completadas las preguntas e indicaciones sobre la identificación y evaluación de los impactos adversos potenciales y reales, puede iniciarse la priorización y la planificación de la prevención y la mitigación de los impactos. El proceso de planificación de la mitigación de impactos y de establecimiento del acceso a recursos se aborda en la fase siguiente.

4. Plan de mitigación

Introducción

Una vez identificados y evaluados los impactos adversos potenciales y reales, debe elaborarse un plan de mitigación y debe reflexionarse, según proceda, sobre la puesta a disposición de recursos para las personas potencialmente afectadas.

Esta parte del proceso HUDERIA especifica las actuaciones y los procesos destinados a hacer frente a los impactos adversos potenciales y reales mediante:

- ▶ la formulación de medidas de mitigación;
- ▶ la elaboración de un plan de mitigación basado en la gravedad y la probabilidad de los daños identificados;
- ▶ cuando proceda, el establecimiento del acceso a recursos para personas potencialmente afectadas y para otras partes pertinentes.

Explicaciones

Delimitación y priorización

Una planificación diligente de la prevención y de la mitigación de riesgos e impactos comienza con una etapa de delimitación y de priorización. Con las aportaciones de las personas afectadas, si procede y en la medida en que proceda, debe revisarse cada impacto adverso potencial y real identificado y cartografiarse las interrelaciones e interdependencias entre ellos, así como los factores de riesgo sociales circundantes identificados durante la fase COBRA (por ejemplo, las vulnerabilidades y la precariedad específicas del contexto).

Cuando sea necesario priorizar las actuaciones de prevención y de mitigación (por ejemplo, cuando las demoras en la gestión de un daño potencial o la vulnerabilidad específica de una persona o un grupo afectados puedan reducir su reversibilidad), la toma de decisiones debería guiarse por consideraciones relativas a la probabilidad y a la gravedad con respecto de los impactos considerados.

Obligaciones jurídicas

Un aspecto importante en la elaboración de un plan de mitigación es que, en esta fase del proceso HUDERIA, deberían tenerse en cuenta las obligaciones jurídicas relativas al respeto de los derechos humanos, la democracia y el Estado de Derecho, tal como se establecen en el derecho internacional y nacional aplicable, al valorar si los impactos adversos potenciales pueden mitigarse y los impactos adversos reales pueden abordarse y, en caso afirmativo, de qué modo.

La disponibilidad y la eficacia de los recursos jurídicos, incluidas la restitución o la indemnización como recursos jurídicos, vienen determinadas por el derecho internacional y nacional aplicable.

Jerarquía de mitigación

A la hora de decidir el abanico de actuaciones disponibles para prevenir o mitigar los impactos adversos potenciales, puede utilizarse un enfoque estructurado denominado «jerarquía de mitigación» (evitar, mitigar, restituir, indemnizar).

En las primeras etapas del ciclo de vida de un sistema de IA, los impactos considerados aún no se habrán producido, por lo que resultarán más pertinentes las opciones de mitigación «evitar» y «mitigar». En cambio, en iteraciones posteriores de seguimiento, revisión y reevaluación (es decir, durante la etapa de despliegue), los impactos adversos pueden haberse producido ya, lo que hace pertinentes las opciones de mitigación «restituir» e «indemnizar» junto a «evitar» y «mitigar».

Las descripciones de las opciones incluidas en la jerarquía de mitigación son las siguientes.

EVITAR	MITIGAR	RESTITUIR	INDEMNIZAR
Introducir cambios, desde el principio, en los procesos de diseño, desarrollo o despliegue que sustentan la producción y el uso del sistema de IA, o en el propio sistema de IA, para evitar el impacto adverso. Conviene señalar que evitar no equivale a ignorar los posibles impactos negativos.	Aplicar medidas en los procesos de diseño, desarrollo o despliegue que sustentan la producción y el uso del sistema de IA, o introducir cambios en el propio sistema de IA, para minimizar el impacto adverso.	Introducir cambios para restituir o rehabilitar a las personas afectadas hasta una situación similar, o al menos equivalente, a la que tenían antes del impacto adverso.	Indemnización en especie o por otros medios, cuando sea factible y cuando los demás enfoques de mitigación no sean posibles ni eficaces.
Nivel 1	Nivel 2	Nivel 3	Nivel 4
Opción preferente	...		Opción menos preferente

El empleo del término «jerarquía de mitigación» sugiere que debe darse prioridad, en primer lugar, a evitar por completo los impactos adversos potenciales y reales y, después, a reducirlos y repararlos. Cabe señalar, además, que, en las últimas etapas del ciclo de vida de un sistema de IA, cuando las opciones de restitución e indemnización cobran mayor relevancia, puede resultar pertinente más de una de estas opciones de mitigación (por ejemplo, cuando es necesario rehabilitar a una persona afectada al tiempo que se adoptan medidas inmediatas para minimizar daños ulteriores).

En todos los casos, las decisiones sobre qué actuación o actuaciones de prevención o de mitigación deben de emprenderse han de guiarse por consideraciones que prioricen la protección de los derechos humanos, la democracia y el Estado de Derecho, y las opciones dirigidas a evitar y mitigar los impactos adversos deberían de preferirse a las opciones de indemnizar o de compensar a las personas potencialmente afectadas por los daños sufridos.

Teniendo en cuenta toda la información obtenida en esta fase del proceso HUDERIA, existe la posibilidad de reexaminar las preguntas cero. Esta información también podría resultar útil para orientar los debates destinados a determinar si las actividades del ciclo de vida del sistema de IA en cuestión (en fase de desarrollo o ya en uso, en caso de revisión iterativa) se ajustan a los derechos humanos, la democracia y el Estado de Derecho.

Acceso a recursos

Las medidas destinadas a hacer frente a los impactos adversos no se limitan a los recursos jurídicos. Tales impactos pueden abordarse recurriendo a otras medidas de mitigación, como las previstas en políticas, directrices u otros instrumentos.

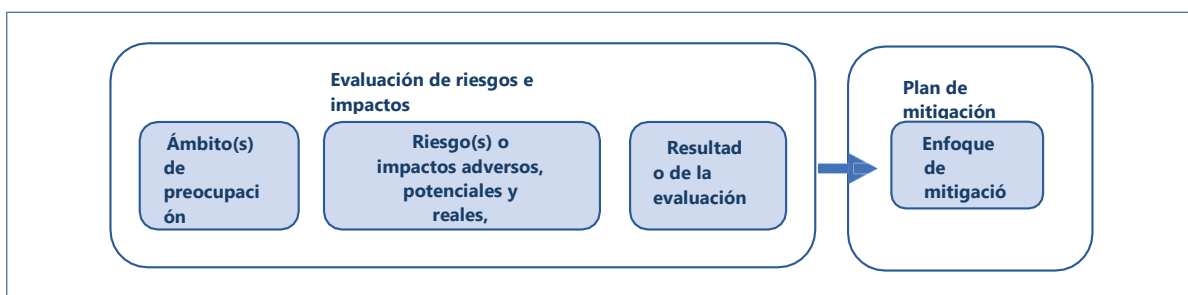
Al establecer dichas medidas podrían tomarse en consideración los puntos siguientes:

- a. Si existen ya medidas y mecanismos de rendición de cuentas en relación con los derechos humanos, la democracia y el Estado de Derecho. Es fundamental que estos marcos ya existentes se apliquen al contexto de los sistemas de IA.
- b. La complejidad técnica, la opacidad y el carácter intensivo en datos de algunos sistemas de IA puede limitar su transparencia. Ello puede generar un desequilibrio considerable en el acceso a la información, en su comprensión o en su control entre las distintas partes que intervienen en el ciclo de vida del sistema de IA. Las medidas destinadas a documentar el sistema de IA y sus impactos y a informar de ellos a las personas afectadas pueden facilitar la puesta a disposición y la accesibilidad de recursos efectivos frente a los impactos adversos que sufran.
- c. La información facilitada mediante esas medidas debería ser adecuada al contexto, clara y significativa, de modo que las personas puedan utilizarla de manera efectiva para ejercer sus derechos en los procedimientos relacionados con las decisiones que les afectan.
- d. En su caso y según proceda, puede ser necesario ofrecer a las personas afectadas garantías procesales y salvaguardas efectivas adicionales, de conformidad con el derecho internacional y nacional aplicable.

Resultado de este elemento del proceso HUDERIA

Este elemento debería dar lugar a una descripción clara de las medidas y actuaciones destinadas a hacer frente a los riesgos e impactos identificados, junto con una aclaración de las funciones y responsabilidades de los distintos agentes que intervienen en la mitigación, la gestión y el seguimiento. En su caso, este elemento debería producir, además, una exposición accesible de los mecanismos y de las medidas de reparación a disposición de las personas afectadas.

Además (véase la sección "Revisión iterativa" más adelante), se establece un plan para el seguimiento de las labores de mitigación y para la reevaluación y revaloración iterativas de dichas labores a lo largo de las fases posteriores del ciclo de vida del sistema de IA.



Revisión iterativa

Introducción

Llevar a cabo el proceso HUDERIA al inicio del ciclo de vida de un sistema de IA no es más que un primer paso, aunque esencial, de un proceso más largo e iterativo de seguimiento y de reevaluación responsables. El proceso de revisión iterativa garantiza que la evaluación de riesgos e impactos siga siendo eficaz a lo largo de todo el ciclo de vida del sistema de IA. Se trata de un proceso continuo que ofrece oportunidades periódicas de detectar nuevos impactos y de actualizar el plan de mitigación.

Con el tiempo, es probable que los impactos del sistema de IA evolucionen como consecuencia de las decisiones adoptadas durante su desarrollo y aplicación, de las aplicaciones contextuales o de cambios externos en el entorno real. Estos cambios, que pueden afectar al ciclo de vida de los datos, el desarrollo y el diseño del sistema de IA, los procesos de contratación, las variaciones en las técnicas de IA, la integración o la puesta en funcionamiento del sistema, las vulnerabilidades de seguridad o los acontecimientos o hechos significativos que den lugar a consecuencias dañinas o no previstas, pueden influir en el rendimiento del sistema de IA o en su impacto sobre las personas y los grupos afectados.

Tales cambios exigen un procedimiento de revisión que garantice que los derechos humanos, la democracia y el Estado de Derecho se respeten de forma continuada a lo largo del ciclo de vida del sistema de IA. Debería prestarse especial atención al modo en que esos cambios afectan al rendimiento del sistema y a su impacto sobre las personas y las comunidades.

Factores de producción, aplicación y despliegue

Las decisiones adoptadas en cualquier momento del ciclo de vida del sistema, así como los acontecimientos que se produzcan durante su despliegue, pueden exigir una revisión de las decisiones y evaluaciones previas, en particular de las derivadas del proceso HUDERIA, lo que hace necesarias la reevaluación, la reconsideración y la modificación.

Estos cambios, en concreto los relativos al ciclo de vida de los datos, el desarrollo y el diseño del sistema de IA, las variaciones en las técnicas de IA, la integración o la puesta en funcionamiento del sistema, las vulnerabilidades de seguridad o los acontecimientos o hechos significativos que den lugar a consecuencias dañinas o no previstas, pueden influir en el rendimiento del sistema de IA o en su impacto sobre las personas y los grupos afectados. Los procesos que se desarrollan a lo largo del ciclo de vida de un sistema de IA son iterativos y a menudo no lineales, por lo que con frecuencia exigen revisiones y actualizaciones, según proceda.

Factores del entorno real

Los cambios que se producen en los entornos sociales, reglamentarios, políticos o jurídicos en los que el sistema se encuentra en fase de desarrollo o de uso pueden afectar al buen funcionamiento del sistema de IA y a la forma en que este afecta a los derechos de las personas o de los grupos afectados.

Del mismo modo, pueden producirse cambios reglamentarios y de política, así como cambios en los métodos de registro de datos, entre la población afectada, de tal forma que se vea afectada la capacidad de los datos utilizados para entrenar el modelo de reflejar con exactitud los fenómenos, las poblaciones o los factores conexos.

En esa misma línea, pueden producirse cambios culturales o de comportamiento en las poblaciones afectadas que alteren la distribución subyacente de los datos y mermen el rendimiento de un modelo entrenado con datos recogidos antes de tales cambios. Todas estas alteraciones de las condiciones de contexto pueden incidir de manera significativa en el rendimiento de un sistema de IA y en el modo en que éste afecta a las personas, los grupos, las comunidades y la sociedad en general.

Aplicación de la revisión iterativa

Si bien la metodología HUDERIA ofrece cierta flexibilidad en lo que respecta a las modalidades, umbrales, factores desencadenantes y mecanismos de seguimiento y de gobernanza concretos del proceso de revisión iterativa, pueden tenerse en cuenta los principios siguientes:

- a. Una revisión continua del proceso HUDERIA desempeña un papel esencial para mantener su eficacia y fiabilidad.
- b. Se establece un plan para el seguimiento de los impactos y la reevaluación y revalorización del proceso HUDERIA durante cada fase del ciclo de vida del proyecto, hasta la retirada o el desmantelamiento del sistema.
- c. Los procesos empleados para la revisión iterativa deberían ser tan sensibles como sea posible al modo en que el sistema de IA interactúa con sus entornos operativos y con las personas afectadas (posibles ámbitos de aplicación del sistema de IA, aparición de nuevas formas de uso indebido del sistema, etc.).
- d. En contextos que evolucionan o cambian rápidamente, podría ser necesario intervenir con mayor frecuencia mediante reevaluaciones o revalorizaciones.

Parte II

MODELO DE RECUSOS PARA EL ANÁLISIS DE RIESGOS BASADO EN EL CONTEXTO (COBRA)

Prefacio

¿Qué es HUDERIA?

La evaluación de los riesgos e impactos de los sistemas de inteligencia artificial (IA) desde la perspectiva de los derechos humanos, la democracia y el Estado de Derecho («HUDERIA») es una guía que ofrece un enfoque estructurado para evaluar los riesgos e impactos de los sistemas de IA, específicamente diseñada para la protección y la promoción de los derechos humanos, la democracia y el Estado de Derecho. Está llamada a desempeñar un papel único y esencial en la intersección entre las normas internacionales en materia de derechos humanos y los marcos técnicos existentes sobre la gestión de riesgos en el contexto de la IA.

HUDERIA puede ser utilizada tanto por los actores públicos como privados para ayudar a identificar y abordar los riesgos e impactos sobre los derechos humanos, la democracia y el Estado de Derecho a lo largo de todo el ciclo de vida de los sistemas de IA.

Relación con el Convenio Marco

HUDERIA es un conjunto de orientaciones autónomo y no jurídicamente vinculante que carece de efectos jurídicos. No tiene carácter obligatorio y no pretende servir de instrumento interpretativo del Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, en lo sucesivo denominado «el Convenio Marco». Además, aunque HUDERIA tiene una función facilitadora, no constituye un medio para la aplicación del Convenio Marco. Para contribuir a la gestión de los riesgos e impactos de la IA pueden emplearse numerosos marcos, políticas, directrices, normas o herramientas, existentes o futuros, incluida HUDERIA.

Las Partes en el Convenio Marco disponen de flexibilidad para utilizar o adaptar la guía, en su totalidad o en parte, con el fin de desarrollar nuevos enfoques para la evaluación de riesgos, o para utilizar o adaptar enfoques ya existentes de conformidad con su legislación aplicable, siempre que cumplan plenamente sus obligaciones en virtud del Convenio Marco, incluidos los requisitos mínimos de gestión de riesgos e impactos establecidos en su capítulo V.

Directrices generales y específicas

HUDERIA comprende tanto la [metodología HUDERIA](#) como el modelo HUDERIA.

En el plano general, la metodología HUDERIA describe conceptos, procesos y elementos de alto nivel que orientan las actividades de evaluación de riesgos e impactos de los sistemas de IA que puedan repercutir en los derechos humanos, la democracia y el Estado de Derecho.

En el plano específico, el modelo HUDERIA proporciona materiales y recursos de apoyo (como herramientas flexibles pertinentes para los distintos elementos del proceso HUDERIA, recursos de carácter ilustrativo y recomendaciones escalables) que pueden facilitar la aplicación de la metodología HUDERIA. El presente documento forma parte del modelo HUDERIA y contiene los recursos COBRA

(análisis de riesgos basado en el contexto), que contribuyen a hacer operativo el proceso COBRA mediante preguntas estructuradas, explicaciones y ejemplos concebidos para respaldar la identificación y evaluación sistemática de los riesgos.

Estos recursos constituyen un conjunto de conocimientos que puede facilitar el examen y el tratamiento de los riesgos e impactos relacionados con los derechos humanos, la democracia y el Estado de Derecho, también en el marco de otros enfoques de gestión de riesgos. Para mayor seguridad, los recursos y directrices del modelo HUDERIA no se ofrecen como ejemplos de buenas prácticas ni establecen normas mínimas; ahora bien, pueden servir de inspiración para que tanto los agentes públicos como los privados desarrollen enfoques de gestión de riesgos adaptados a cada caso, y que los Estados podrán adoptar

Recursos COBRA

Introducción

El presente documento pretende ofrecer directrices ilustrativas para el análisis de riesgos basado en el contexto, primer elemento de la metodología HUDERIA, y consta de distintos elementos, entre ellos el material que puede utilizarse para orientar la reflexión en el contexto de la evaluación de los riesgos e impactos potenciales de los sistemas de IA. De acuerdo con el enfoque flexible de la metodología, estos elementos pueden ajustarse según sea necesario. Este material también podrá utilizarse además en el futuro para crear herramientas interactivas (como plataformas en línea o programas informáticos) que faciliten la realización de las evaluaciones de riesgos e impactos.

Los elementos de esta sección pueden utilizarse para ayudar a identificar y evaluar los factores de riesgo que inciden en el alcance, la escala, la probabilidad y la reversibilidad de los impactos adversos en los derechos humanos, la democracia y el Estado de Derecho que surgen en el contexto de aplicación del sistema de IA (recursos COBRA A), en sus contextos de diseño y desarrollo (recursos COBRA B) y en su contexto de despliegue (recursos COBRA C). Puede ser necesario revisarlos y examinarlos de forma iterativa a lo largo del ciclo de vida de un sistema de IA.

Con el fin de facilitar la cartografía de los posibles impactos en los derechos humanos, la democracia y el Estado de Derecho, los recursos COBRA E enumeran los ámbitos de posible preocupación desde el punto de vista de los derechos humanos, la democracia y el Estado de derecho, ofrecen una descripción ilustrativa de dichos ámbitos y aportan ejemplos de posibles riesgos relacionados con la IA para cada uno de ellos. La lista podría utilizarse junto con los recursos COBRA F, que enumeran y describen brevemente los sectores y ámbitos potencialmente pertinentes y los ponen en relación con los ámbitos de posible preocupación de los recursos COBRA E.

Panorámica de los recursos COBRA A, B y C

Esta sección ofrece ejemplos ilustrativos de posibles factores de riesgo que inciden en la probabilidad de que se produzcan impactos adversos en los derechos humanos, la democracia y el Estado de Derecho, y que surgen en los ámbitos siguientes.

- ▶ El contexto de aplicación de un sistema de IA (recursos COBRA A), como los entornos legales y reglamentarios en los que se desarrolla y utiliza el sistema, su finalidad prevista, las categorías de usuarios previstos y de personas afectadas y otros posibles datos relativos al contexto de aplicación del sistema, como los antecedentes conocidos de sesgo y de discriminación.
- ▶ El contexto de diseño y de desarrollo del sistema de IA (recursos COBRA B), en particular sus características técnicas, como las limitaciones conocidas del sistema, las consideraciones relativas a la recogida, el enriquecimiento, el almacenamiento, el uso y la retirada de los datos, y las

consideraciones relativas al propio algoritmo o modelo, así como las características técnicas relacionadas con la privacidad y la protección de datos personales, el sesgo y la discriminación, y la explicabilidad y la interpretabilidad;

- El contexto de despliegue del sistema de IA (recursos COBRA C), como las características de sus usuarios previstos y de los grupos afectados, y las medidas adoptadas para garantizar la seguridad general, prevenir daños y preservar la dignidad humana, proteger la privacidad y los datos personales, mitigar el sesgo nocivo, garantizar una formación adecuada, prevenir usos no previstos y asegurar la rendición de cuentas y el cumplimiento de la legalidad.

Los recursos tienen carácter ilustrativo, no exhaustivo y abierto, lo que significa que será necesario adaptarlos con el tiempo para tener en cuenta la evolución, por ejemplo, en los ámbitos de la tecnología y los regímenes jurídicos y normativos aplicables.

Recursos COBRA A (Lista de factores de riesgo que surgen en el contexto de aplicación del sistema de IA)

Ejemplos de posibles factores de riesgo	
1. Sector o ámbito en el que se construye el sistema de IA	A.1.1 A la vista de su finalidad prevista, ¿desempeñará el sistema de IA un papel significativo en un contexto de gran impacto —como el transporte, la atención social, la sanidad o la educación— en el que sus resultados podrían influir en decisiones o actuaciones importantes? ¿Cuál es el contexto en el que opera el sistema? <i>En tales contextos, cualquier fallo de funcionamiento o degradación del rendimiento a lo largo del ciclo de vida del sistema puede dar lugar a decisiones perjudiciales, interrupciones del servicio o incidentes de seguridad. Considérese si el sistema podría afectar a ámbitos en los que pudieran estar en juego los derechos humanos, la democracia o el Estado de Derecho. Los recursos COBRA E exponen ejemplos de daños potenciales e intereses protegidos que pueden resultar pertinentes, y los recursos COBRA F contienen la lista de sectores de posible relevancia desde el punto de vista de los derechos humanos, la democracia y el Estado de Derecho.</i> <i>Ejemplo: un sistema de IA híbrido utilizado para controlar la cinemática de un robot quirúrgico que asiste a los médicos que practican una intervención médica de urgencia.</i>
	A.1.2 ¿Desempeñará el sistema de IA una función de gran impacto, independientemente del sector en el que opere? ¿Cuál es la función del sistema? <i>Incluso fuera de los sectores de gran impacto, este tipo de funciones pueden dar lugar a decisiones perjudiciales, interrupciones del servicio u otros resultados adversos que puedan afectar a los derechos humanos de las personas potencialmente afectadas o repercutir en la democracia y el Estado de Derecho —véanse los recursos COBRA E—, aunque el sistema no se despliegue en un contexto tradicionalmente de gran impacto.</i> Ejemplo: el sistema que se está desarrollando es una herramienta de selección de personal asistida por IA para una función de la administración pública. La herramienta no se sitúa en un sector crítico para la seguridad; sin embargo, el sistema podría tener un impacto negativo en grupos de personas en situación de vulnerabilidad.
2. Legislación vigente y entorno regulatorio	A.2.1 ¿Está el sector o ámbito en el que se desplegará el sistema de IA sujeto a una amplia supervisión jurídica o regulatoria? En caso afirmativo, ¿cuáles son las obligaciones o normas específicas (por ejemplo, en materia de cumplimiento, certificación, supervisión o ética) que pueden aplicarse al sistema de IA o a sus resultados? <i>Una supervisión jurídica o regulatoria puede constituir, por sí misma, un indicio de contexto de gran impacto. Cuando existe tal supervisión, es importante identificar las obligaciones y normas pertinentes que puedan interactuar con las consideraciones de HUDERIA o servirles de orientación. La evaluación HUDERIA se centra específicamente en los posibles impactos en los derechos humanos, la democracia y el Estado de Derecho. Otras obligaciones sectoriales (por ejemplo, en materia de estabilidad financiera, seguridad sanitaria o ciberseguridad) siguen sujetas a sus respectivos marcos reglamentarios. Cuando resulte pertinente, las evaluaciones o certificaciones existentes (obligatorias o voluntarias) pueden citarse como documentación de apoyo para reforzar la base probatoria del análisis HUDERIA, pero no lo sustituyen ni se integran en él.</i> Ejemplo: un banco que utiliza una herramienta de evaluación de riesgos para predecir el impago de los prestatarios opera en el sector financiero, donde históricamente existe una amplia regulación en materia de abuso de mercado, gestión de riesgos, derecho de sociedades y derecho de la competencia.

3. Alcance del despliegue	A.3.1 En un escenario en el que el sistema de IA está plenamente desplegado de acuerdo con su diseño, ¿en qué escala afectará el sistema de IA, directa o indirectamente, a las personas y los grupos (poblaciones locales, poblaciones nacionales, poblaciones mundiales)? <i>Esta pregunta se centra en el alcance espacial y jurisdiccional del impacto potencial del sistema de IA. La escalabilidad puede alterar sustancialmente el perfil de riesgo del sistema, con posibles efectos sobre las poblaciones locales y sobre las comunidades y los grupos que las integran, así como sobre los procesos sociales y políticos. Estos impactos pueden ser significativos desde el punto de vista de los derechos humanos, la democracia y el Estado de derecho (véanse los recursos COBRA E, en particular los ámbitos (16) Democracia y (17) Estado de derecho). Cuando los sistemas se introducen de forma gradual o se someten a ensayo en fases piloto o de prueba con un grupo limitado, ese despliegue escalonado también puede constituir una medida de mitigación, al permitir identificar, evaluar y gestionar los impactos potenciales antes de una implantación más amplia.</i> <i>Ejemplo (local): un sistema de vigilancia policial predictiva utilizado por un departamento de policía municipal y que opera exclusivamente dentro de los límites del municipio.</i> Ejemplo (nacional): un sistema de clasificación utilizado por un organismo público para determinar si las personas cumplen los requisitos para acceder a prestaciones sociales y que afecta a personas y grupos de la población nacional. Ejemplo (global): un sistema de recomendación utilizado por una plataforma internacional de redes sociales para personalizar la difusión de noticias y que afecta a personas y grupos de la población mundial.
---------------------------	---

Ejemplos de posibles factores de riesgo	
3. Alcance del despliegue	A.3.2 En un escenario en el que el sistema de IA escale de forma óptima, ¿a cuántas personas potencialmente afectadas afectará el sistema de IA de forma directa o indirecta? <i>Esta pregunta se centra en la magnitud numérica y demográfica del impacto potencial. El uso del sistema de IA a gran escala o de forma masiva puede generar efectos significativos en los procesos e instituciones sociales y políticos, con implicaciones particulares para los derechos humanos, la democracia y el Estado de Derecho (véanse los recursos de COBRA E, en particular los ámbitos (16) Democracia y (17) Estado de derecho).</i> Ejemplo (pequeña): una organización no gubernamental local con 60 empleados utiliza un chatbot de IA para apoyar a las víctimas de violencia doméstica, proporcionándoles información sobre sus derechos legales y los centros de acogida disponibles. Ejemplo (moderada): una agencia regional de empleo utiliza un clasificador basado en el aprendizaje automático para filtrar y priorizar las candidaturas a contratos de corta duración en el sector público, con unas 10.000 candidaturas al año. Ejemplo (grande): una aseguradora sanitaria nacional despliega un sistema predictivo para señalar a los pacientes «en riesgo» de cara a programas de intervención temprana, lo que afecta a aproximadamente 1.2 millones de asegurados. Ejemplo (muy grande): una plataforma estatal de identidad digital integra la verificación de identidad basada en IA (por ejemplo, el reconocimiento facial y el escaneado automático de documentos) para controlar el acceso a los regímenes de protección social, lo que afecta a más de 50 millones de personas.
	A.3.3 Atendiendo a los posibles impactos directos e indirectos del sistema de IA sobre las personas, las comunidades y el medio ambiente, ¿cuál es la escala temporal más amplia dentro de la cual el sistema de IA podría afectar a personas y grupos? <i>El uso del sistema de IA puede dar lugar a efectos a más largo plazo que se prolonguen más allá del período de su aplicación y que repercutan en los derechos humanos de las personas potencialmente afectadas, así como en los procesos y las instituciones sociales y políticos, en particular en relación con la democracia y el Estado de Derecho (véanse los recursos COBRA E, en particular los ámbitos (16) Democracia y (17) Estado de Derecho).</i> Ejemplo (corto plazo): una oficina municipal de vivienda utiliza un sistema de IA conversacional para responder a las consultas de los inquilinos sobre impagos de alquiler, notificaciones de desahucio y procedimientos de reclamación. Ejemplo (medio plazo): una autoridad educativa regional utiliza un sistema de elaboración de perfiles basado en IA para predecir la probabilidad de fracaso académico del alumnado, lo que da lugar a intervenciones específicas o a una reorientación hacia itinerarios formativos profesionales. Ejemplo (largo plazo): una agencia nacional de migración despliega un sistema de clasificación por IA para evaluar como de riesgo bajo, medio o alto las solicitudes de inmigración, visado de larga duración o nacionalidad.
4. Antecedentes de sesgo	A.4.1 ¿Presentan el sector o los ámbitos en los que operará el sistema de IA, y de los que proceden los datos utilizados para entrenarlo, patrones de discriminación, desigualdad, sesgo o inexactitud que lleven sistemáticamente a un trato injusto de grupos de personas en situación de vulnerabilidad (o de grupos de personas con características protegidas)? En caso afirmativo, ¿qué probabilidad hay de que esos patrones se repliquen o se amplifiquen en el funcionamiento del sistema o en sus resultados y en sus impactos a corto, medio y largo plazo? Considérese centrar la atención en las consideraciones de igualdad y no discriminación relativas a los posibles impactos del sistema de IA sobre las personas afectadas. <i>Si existen tales patrones, el sistema puede reproducirlos o amplificarlos en el desarrollo y el despliegue, lo que daría lugar a resultados diferenciados según los grupos y a una discriminación indirecta. A gran escala y en horizontes de corto, medio y largo plazo, ello puede consolidar las disparidades y menoscabar la igualdad y la no discriminación.</i>

	Ejemplo: el uso de un sistema de cribado de candidaturas en el sector científico y tecnológico basado en datos históricos de contratación replica los patrones de discriminación en la contratación y genera resultados desfavorables para grupos de personas en situación de vulnerabilidad.
A.4.2	Teniendo en cuenta el sector o ámbito en el que opera el sistema de IA, ¿incluirán las personas potencialmente afectadas a quienes puedan verse afectados de forma significativa o desproporcionada por el diseño y el uso del sistema, en particular grupos de personas en situación de vulnerabilidad (o grupos de personas con características protegidas)? <i>La vulnerabilidad puede ser situacional, derivarse de condiciones económicas, sociales, institucionales, tecnológicas o de otro tipo propias del contexto. Considérese centrar la atención en las consideraciones de igualdad y no discriminación relativas a los posibles impactos del sistema de IA sobre las personas potencialmente afectadas. Préstese especial atención a los factores de riesgo B.5.1 a B.5.4 de los recursos COBRA B.</i> Ejemplo: una herramienta de evaluación de riesgos utilizada por los trabajadores sociales para determinar el acceso a las prestaciones sociales opera en el ámbito de la administración social, donde las personas en situación de vulnerabilidad económica o social se ven afectadas de forma significativa.

Ejemplos de posibles factores de riesgo	
4. Antecedentes de sesgo	A.4.3 ¿Es probable que los niños accedan al sistema de IA o se vean afectados por él? <i>Se debe prestar especial atención a los niños, ya que la evolución de sus facultades, sus relaciones de dependencia y su limitada capacidad para prestar un consentimiento informado o para reclamar reparación generan perfiles de riesgo específicos que las evaluaciones generales de igualdad o de vulnerabilidad pueden no captar adecuadamente. Los impactos sobre los niños pueden ser desproporcionadas y duraderos debido a sus vulnerabilidades relacionadas con el desarrollo, a las asimetrías de poder y a la mayor sensibilidad de sus datos. Estos riesgos suscitan preocupaciones particulares en relación con el bienestar mental, la igualdad y la no discriminación, así como con los derechos del niño (véanse los recursos COBRA E, en particular el ámbito (14) Infancia). Pueden surgir asimismo riesgos adicionales de una participación insuficiente de los niños o de las personas responsables de su bienestar (como padres, tutores, organizaciones de la sociedad civil o defensores de la protección de la infancia) a lo largo del ciclo de vida del sistema de IA (véase el elemento SEP de la metodología HUDERIA).</i>
5. Contexto medioambiental	A.5.1 ¿Cuáles son los posibles impactos directos e indirectos del sistema de IA sobre el medio ambiente (en relación, por ejemplo, con el consumo de energía, las emisiones de carbono, el uso de recursos o los residuos electrónicos)? ¿En qué medida están su desarrollo, su despliegue y su uso sujetos a normas, reglamentaciones o compromisos de sostenibilidad medioambientales ya existentes, y son esos marcos adecuados para hacer frente a los posibles impactos del sistema? <i>El sistema de IA puede generar impactos directos e indirectos sobre el medio ambiente, con posibles consecuencias para la eficiencia energética, las emisiones de carbono, el uso de recursos y la gestión de residuos. Estos impactos medioambientales pueden, a su vez, afectar al disfrute de los derechos humanos, en particular para las personas en situación de vulnerabilidad o los grupos con características protegidas (véanse los recursos COBRA E).</i> Ejemplo: un modelo de IA utilizado para generar cuadros de mando analíticos diarios sigue reprocesando cada noche un conjunto de datos invariable de más de 100 TB, con un consumo

	considerable de capacidad de cálculo y de energía, sin producir ningún resultado nuevo ni significativo. Esta decisión de diseño da lugar a emisiones y a un consumo de recursos innecesarios, sin beneficio alguno para el público ni para los usuarios.
--	--

Recursos COBRA B (Lista de factores de riesgo que surgen en el contexto de diseño y desarrollo del sistema de IA)

Ejemplos de posibles factores de riesgo	
1. Definición de diseñar y definición del problema y del resultado	B.1.1 ¿Qué otros enfoques, aparte de construir el sistema de IA, podrían atender de forma viable la necesidad o necesidades previstas? En comparación con el sistema de IA, ¿qué beneficios y riesgos presentarían estos otros enfoques, en particular en cuanto a los impactos en los derechos humanos, la democracia y el Estado de Derecho? <i>La decisión de construir y desplegar un sistema de IA debería evaluarse críticamente frente a enfoques alternativos que pudieran atender mejor las necesidades previstas con menos impactos negativos. Entre los riesgos figuran los daños marginales o nuevos asociados a la introducción de la IA en un contexto ya existente, especialmente en lo que respecta a los derechos humanos, la democracia y el Estado de Derecho. La evaluación de las alternativas exige considerar: a) los posibles efectos adversos en los derechos humanos, la democracia y el Estado de derecho (incluidos los riesgos marginales de la introducción de la IA); b) si las tecnologías y los procesos ya existentes, incluidas las técnicas de mitigación y de gobernanza, podrían atender la necesidad; c) la suficiencia de los recursos y datos disponibles; d) la complejidad de los contextos de uso; e) los posibles beneficios de emplear una solución basada en IA, incluidos los costes de oportunidad de no seguir adelante; f) la naturaleza del problema que se pretende resolver, en particular si se trata de resolver un problema político o social. No evaluar estas alternativas puede llevar a adoptar una solución que aumente innecesariamente el riesgo o la complejidad.</i>
2. Madurez tecnológica	B.2.1 ¿Se basa el diseño del sistema de IA en técnicas de validación bien conocidas y ampliamente reconocidas para una finalidad prevista similar en el mismo sector? En caso negativo, ¿qué riesgos puede entrañar una madurez tecnológica limitada para la calidad del rendimiento del sistema y qué efectos podrían derivarse de ello para los derechos humanos de las personas potencialmente afectadas o para la democracia y el Estado de Derecho? <i>Si el diseño y el desarrollo del sistema de IA no se basan en técnicas bien conocidas y previamente validadas para finalidades similares en el mismo sector —incluidas las normas del sector y las buenas prácticas aplicables— y teniendo en cuenta el uso previsto del sistema y el umbral de error aceptable, existe el riesgo de que su inmadurez tecnológica menoscabe la calidad, la fiabilidad y la robustez de su rendimiento. Tales deficiencias pueden aumentar la probabilidad de errores, reducir la confianza que merece el sistema e incrementar el riesgo de impactos adversos sobre las personas potencialmente afectadas. Sin embargo, la innovación suele implicar enfoques novedosos o experimentales que se apartan de la práctica establecida, lo que puede aumentar la incertidumbre y el riesgo de fallos de rendimiento o de impactos imprevistos. En tales casos, la ausencia de precedentes sectoriales debería reconocerse como un factor de riesgo agravado que exige medidas de control compensatorias, como una supervisión humana reforzada, la revisión por expertos independientes o despliegues piloto cuidadosamente supervisados. Estos mecanismos son esenciales para gestionar la incertidumbre, verificar la fiabilidad del sistema y garantizar que el avance tecnológico se produzca de manera segura y responsable.</i>

Ejemplos de posibles factores de riesgo	
3. Sistema preexistente	B.3.1 Si el sistema de IA va a sustituir a un sistema preexistente, ¿se han identificado y mitigado los defectos y riesgos del sistema al que sustituye? <i>Cuando un sistema de IA sustituye a un sistema humano, técnico o híbrido que cumple una función similar, o cuando su despliegue lleva de otro modo a que los operadores humanos confíen en exceso en los resultados de la IA, pueden surgir riesgos si las limitaciones, los defectos o los daños documentados del sistema preexistente no se comprenden y abordan adecuadamente. No tenerlos en cuenta puede dar lugar a la réplica o la amplificación de los riesgos existentes, en lugar de a una mejora. La calidad y el impacto del rendimiento del sistema de IA dependerán, en parte, de en qué medida aprenda de las deficiencias del sistema sustituido, las corrija o las compense.</i>
	B.3.2 ¿Sustituye el sistema de IA a un sistema humano, técnico o híbrido que constituye una infraestructura crítica o que cumple una función de gran impacto? En caso afirmativo, ¿qué riesgos (como interrupciones o perturbaciones) puede entrañar el proceso de actualización o sustitución del sistema y cómo podrían repercutir en los derechos humanos de las personas potencialmente afectadas? <i>Si el sistema de IA sustituye a un sistema humano, técnico o híbrido que constituye una infraestructura crítica o que cumple una función de gran impacto, el proceso de actualización o sustitución del sistema puede plantear riesgos significativos. Puede tratarse de riesgos derivados de interrupciones, de perturbaciones de servicios esenciales o de fallos que podrían afectar directamente a los derechos humanos de las personas potencialmente afectadas, a la democracia o al Estado de Derecho. El carácter crítico del sistema sustituido amplifica la escala potencial de tales impactos.</i>
4. Contexto de ciberseguridad	B.4.1 ¿Qué motivaciones y oportunidades ofrece el sistema de IA a los agentes malintencionados para vulnerarlo, corromperlo o utilizarlo indebidamente? ¿Qué riesgos plantean los intentos de comprometer su seguridad y su robustez (por ejemplo, mediante ataques adversarios, envenenamiento de datos, inversión del modelo o violaciones de la seguridad de los datos)? ¿Qué riesgos podría entrañar tal explotación malintencionada para los derechos humanos de las personas potencialmente afectadas, la democracia o el Estado de Derecho? <i>El sistema de IA puede generar motivaciones y oportunidades para que agentes malintencionados lo vulneren, lo corrompan o lo manipulen de otro modo. Tales actuaciones podrían obedecer a un afán de lucro, a objetivos políticos o a otros beneficios percibidos, y pueden dar lugar a que el sistema se utilice indebidamente para facilitar abusos de los derechos humanos, socavar la confianza o causar daños a las personas potencialmente afectadas.</i>
5. Calidad de los datos y protección de datos personales	B.5.1 ¿Son los datos utilizados en el diseño y el desarrollo del sistema de IA suficientemente representativos de las personas y los grupos afectados, suficientemente exactos, completos, fiables, pertinentes, adecuados y actualizados, y de cantidad y calidad suficientes para el caso de uso, el ámbito, la función y la finalidad previstos? En caso negativo, ¿qué riesgos pueden derivarse para las personas afectadas de las deficiencias detectadas? ¿Qué técnicas se emplearon para realizar esta evaluación? <i>Si los datos utilizados en el diseño y el desarrollo del sistema de IA no son suficientemente representativos, suficientemente exactos, completos, fiables, pertinentes, adecuados y actualizados, ni de cantidad y calidad suficientes para el caso de uso, el ámbito, la función y la finalidad previstos, pueden surgir riesgos significativos. Puede tratarse de riesgos derivados de: a) las consecuencias de utilizar datos inexactos, incoherentes o incompletos; b) la falta de un registro, una trazabilidad y una auditabilidad adecuados de la procedencia y el linaje de los datos a lo largo del ciclo de vida del sistema; c) los errores de medición o los sesgos introducidos durante la recogida de datos (con o sin intervención humana); d) los datos ausentes o inutilizables en los conjuntos de datos recogidos o adquiridos; e) la falta de transparencia y de accesibilidad de los procesos de etiquetado o anotación a efectos de auditoría, supervisión y revisión; f) los sesgos introducidos por etiquetadores y anotadores humanos, incluso a través de variables indirectas, sin salvaguardas adecuadas; g) los sesgos introducidos por el etiquetado o la anotación automatizados, si no están sujetos a una supervisión humana suficiente y a procesos transparentes. Tales deficiencias pueden menoscabar la fiabilidad, la confianza que merece el sistema de IA y su equidad, y pueden afectar negativamente a los derechos humanos de las personas potencialmente afectadas.</i>

6. Contexto de desarrollo y de implantación del modelo	B.5.2	¿Implican el diseño y el desarrollo del sistema de IA el uso de datos personales? En caso afirmativo, ¿qué riesgos pueden surgir para la privacidad y la protección de los datos personales? <i>Si se utilizan datos personales en el diseño y el desarrollo del sistema de IA, existe el riesgo de que no se respeten los requisitos de protección de datos personales. Según proceda, pueden existir riesgos asociados, entre otras cosas, a: a) la insuficiencia de la información facilitada a las personas afectadas y a las partes interesadas sobre el consentimiento u otra base legítima del tratamiento; b) la invocación de un consentimiento tácito o de bases jurídicas poco claras sin consultar a las personas afectadas y a las partes interesadas sobre la aceptabilidad del uso de los datos; c) la insuficiencia de las salvaguardas de que disponen las personas respecto del uso de sus datos una vez compartidos o utilizados para el entrenamiento, incluida, cuando proceda, la posibilidad de retirarlos o de limitar su uso ulterior; d) las situaciones de posible divergencia entre las normas aplicables de protección de datos personales, como la limitación de la finalidad, y las prácticas de IA de reutilización de datos a gran escala o iterativa; e) la posibilidad de desanonimización o reidentificación mediante la vinculación de datos con conjuntos de datos existentes, públicamente accesibles o fácilmente obtenibles, si no se gestiona adecuadamente. Tales deficiencias pueden menoscabar, entre otras cosas, la privacidad y la protección de los datos personales.</i>
	B.5.3	¿Utilizará el sistema de IA datos dinámicos recopilados y procesados en tiempo real (o casi real) para garantizar un aprendizaje continuo, una adaptación o una optimización del rendimiento? En caso afirmativo, ¿qué riesgos pueden surgir en relación con la seguridad, la fiabilidad, la robustez, la calidad de los datos, la integridad de los datos y, cuando proceda, la no discriminación y la mitigación del sesgo? <i>Si el sistema de IA utiliza datos dinámicos recopilados y tratados en tiempo real (o casi real) con fines de aprendizaje continuo, la adaptación o la optimización del rendimiento, pueden surgir nuevos riesgos. Puede tratarse de riesgos asociados a amenazas para la seguridad, la protección, la fiabilidad, la solidez, la calidad de los datos y la integridad de los datos. La adaptación en tiempo real puede además introducir o amplificar riesgos de sesgo y discriminación si no se gestiona adecuadamente. Tales riesgos pueden socavar la confianza y afectar negativamente a los derechos humanos de las personas afectadas.</i>
	B.5.4	¿En qué medida el ámbito y el tipo de datos recogidos o adquiridos plantean riesgos de desplazamientos o derivas distributivas rápidas o inesperadas que pudieran repercutir negativamente en la exactitud y el rendimiento del sistema de IA? <i>El ámbito en el que se recogen o adquieren los datos, así como el tipo de datos utilizados, pueden plantear riesgos de desplazamientos o derivas distributivas rápidas o inesperadas. Tales cambios pueden menoscabar la exactitud, la fiabilidad y el rendimiento del sistema de IA y dar lugar a errores o a resultados dañinos para las personas potencialmente afectadas si no se anticipan y se abordan. La ausencia de evaluación, reevaluación, validación y seguimiento dinámicos incrementa estos riesgos.</i>
	B.6.1	¿Qué posibles riesgos de sesgo y de discriminación indirecta pueden surgir en el marco del diseño, el desarrollo y la implantación del modelo? <i>El diseño, el desarrollo y la implantación del modelo pueden introducir riesgos de sesgo y discriminación indirecta. Estos riesgos pueden surgir: a) durante la ingeniería de características (manual o automatizada), mediante la agrupación, la desagregación o la exclusión de variables de entrada relacionadas con características protegidas o sensibles —o con sus variables indirectas—, lo que genera un sesgo emergente; b) a través de inferencias generadas por los mecanismos de aprendizaje del modelo que resulten irrazonables o injustas, que tengan un impacto dispar o que estén influidas por variables indirectas ocultas de características discriminatorias, y que configuren así los resultados de manera discriminatoria. Si no se abordan, tales riesgos pueden dar lugar a un trato desigual y a discriminación sistémica, y menoscabar los derechos humanos de las personas potencialmente afectadas.</i>

B.6.2	¿Se construye el sistema de IA sobre técnicas intrínsecamente difíciles de explicar, predecir o verificar por completo, como los sistemas no deterministas, los modelos probabilísticos o los modelos evolutivos o dinámicos? En caso afirmativo, ¿cuáles son los posibles riesgos de que repercuta negativamente en los derechos humanos de las personas potencialmente afectadas, especialmente cuando el sistema de IA interactúa con personas? <i>Si el sistema de IA se construye sobre técnicas intrínsecamente difíciles de predecir, verificar o explicar por completo —como los sistemas no deterministas (resultados distintos para una misma entrada), los modelos probabilísticos (resultados basados en la verosimilitud, como los modelos bayesianos o los grandes modelos de lenguaje (LLM)) o los modelos evolutivos o dinámicos (que aprenden o se adaptan de forma continua)—, pueden surgir riesgos específicos. Puede tratarse de riesgos asociados a: a) una menor transparencia, que limite la capacidad de las personas afectadas para comprender o impugnar los resultados o para obtener reparación por ellos; b) un comportamiento imprevisible o inestable, en el que los resultados puedan variar de formas difíciles de prever o de auditar; c) un desajuste entre la inteligibilidad o accesibilidad del modelo y los requisitos sectoriales, jurídicos o de otra índole, o las expectativas depositadas en su función prevista. Estos riesgos pueden verse amplificados cuando faltan medidas de mitigación adecuadas, por ejemplo, cuando las organizaciones no disponen de capacidad suficiente para ofrecer mecanismos complementarios de explicación (como modelos sustitutivos o análisis de importancia de las características) o cuando no existen salvaguardas que garanticen la reversibilidad o la intervención humana. Tales riesgos pueden repercutir negativamente en los derechos humanos, la democracia y el Estado de Derecho al reducir la rendición de cuentas, la transparencia y la confianza en los resultados del sistema.</i>
B.6.3	¿En qué medida se ha garantizado una evaluación, verificación y validación adecuadas del modelo de IA a lo largo de su ciclo de vida, y cómo podrían las lagunas detectadas afectar al nivel de riesgo asociado a su despliegue o uso? <i>Aunque los mecanismos de evaluación, verificación y validación constituyen en sí mismos medidas de mitigación y control, su ausencia o insuficiencia constituye un factor de riesgo fundamental que incide en la fiabilidad y la seguridad del modelo de IA. Si el modelo de IA no se somete a una supervisión, evaluación, verificación y validación suficientes, existe el riesgo de que los errores, defectos o sesgos de su diseño y funcionamiento pasen inadvertidos. Sin procesos transparentes, incluidas la revisión externa por pares y la evaluación por expertos independientes, la fiabilidad, la equidad y la seguridad del sistema pueden verse comprometidas, lo que daría lugar a impactos dañinos sobre las personas potencialmente afectadas y sus derechos humanos, además de a impactos en la democracia y el Estado de Derecho.</i>
B.6.4	¿En qué medida incluye el sistema de IA procesos de seguimiento y de reevaluación periódica que permitan mantenerse al ritmo de los cambios del mundo real y que puedan provocar derivas de concepto o desplazamientos en las distribuciones de datos subyacentes? ¿Qué riesgos pueden surgir para las personas afectadas si tales procesos no existen o son insuficientes? <i>El seguimiento y la reevaluación funcionan como medidas de control destinadas a mantener el rendimiento del sistema a lo largo del tiempo; es su ausencia o su insuficiencia lo que genera el riesgo de degradación o de inequidad en la práctica. Si el sistema de IA no se somete a seguimiento y a reevaluación periódica que le permitan mantenerse al ritmo de los cambios del mundo real, existe el riesgo de que se aparte progresivamente de su finalidad prevista o de que deje de reflejar las condiciones en las que fue diseñado y validado (deriva de concepto o desplazamientos en las distribuciones de datos subyacentes). Estos cambios pueden degradar con el tiempo la exactitud, la equidad y la fiabilidad del sistema, lo que daría lugar a impactos dañinos sobre las personas potencialmente afectadas.</i>

B.6.5	¿En qué medida las métricas de rendimiento del sistema de IA —más allá de la exactitud (sensibilidad, precisión, especificidad)— se seleccionan, se someten a seguimiento y se comunican de manera adecuada, de forma que reflejen el contexto concreto del caso de uso, minimicen el riesgo de interpretación errónea o de uso indebido por parte de los usuarios y de las partes interesadas y contribuyan a garantizar la transparencia y la rendición de cuentas a lo largo del ciclo de vida del sistema? ¿Qué riesgos pueden surgir para las personas afectadas si no es así? <i>La selección y la comunicación de las métricas de rendimiento funcionan como salvaguardas que respaldan la transparencia y la rendición de cuentas; cuando estos mecanismos son débiles o incompletos, se convierten en sí mismos en puntos de riesgo. Si las métricas de rendimiento del sistema de IA no se seleccionan cuidadosamente, no se definen en función del contexto y no se someten a seguimiento y a comunicación transparente, pueden surgir varios riesgos. Puede tratarse de riesgos asociados a: a) priorizar determinados tipos de error (como los falsos positivos o los falsos negativos) sin tener en cuenta el contexto de uso ni los posibles impactos desproporcionados de unas tasas de error diferenciadas sobre las personas en situación de vulnerabilidad o sobre los grupos con características protegidas; b) limitar la evaluación del rendimiento a la exactitud omitiendo otras medidas pertinentes (como la sensibilidad, la precisión y la especificidad), lo que puede ocultar sesgos o debilidades que afecten a la fiabilidad y la equidad; c) presentar las métricas en formatos excesivamente técnicos, inaccesibles o carentes de explicación contextual, lo que impide a los usuarios y a las partes interesadas interpretar correctamente el rendimiento del sistema; y d) una comunicación irregular o un seguimiento insuficiente que dificulten la detección de variaciones de rendimiento entre grupos de población o a lo largo del tiempo.</i>
B.6.6	Si el sistema fundamenta sustancialmente o adopta decisiones que puedan repercutir en los derechos humanos de tal modo que sean necesarios mecanismos de supervisión, revisión o intervención humanas, ¿se han integrado tales funcionalidades en las fases de diseño e implantación del modelo para garantizar que el sistema pueda sustentar la equidad procesal y sustantiva en su uso futuro? <i>Cuando un sistema de IA pueda fundamentar sustancialmente o adoptar decisiones que afecten a los derechos de las personas o a su acceso a servicios, su diseño e implantación deberían incluir funcionalidades que sustenten la equidad procesal y sustantiva. Estas funcionalidades pueden incluir: a) la capacidad de revisión o de anulación humanas, (tanto ex ante como ex post); b) restricciones operativas integradas que impidan que el sistema exceda de forma autónoma su ámbito previsto; c) mecanismos de trazabilidad, auditabilidad y explicabilidad; y d) una atribución clara de las responsabilidades de supervisión a personas que posean la competencia, la formación y la autoridad necesarias. Estos elementos contribuyen a garantizar que, una vez desplegado, el sistema de IA funcione dentro de un marco que posibilite la equidad, la rendición de cuentas y un control humano efectivo.</i>
B.6.7	¿Qué posibilidades hay de que el sistema de IA se reutilice para otros fines o que se emplee de maneras que pudieran suscitar preocupaciones éticas, jurídicas o reglamentarias adicionales? <i>La reutilización para otros fines (ampliación progresiva del ámbito, doble uso, escalada de capacidades) puede modificar sustancialmente el perfil de riesgo del sistema (también desde la perspectiva de la protección de datos personales) al alterar el contexto, los usos de los datos o las vías de decisión. La posibilidad de tal reutilización suele tener su origen en decisiones de diseño y desarrollo (arquitecturas modulares, modelos transferibles o interfaces abiertas). HUDERIA sigue siendo pertinente cuando un sistema se reutiliza para otros fines o su función evoluciona, pues tales cambios pueden generar impactos nuevos o amplificadas en los derechos humanos, la democracia y el Estado de derecho. En esos casos, la evaluación debería actualizarse o repetirse para tener en cuenta el nuevo contexto.</i> Ejemplo: un sistema de visión por computador utilizado para reconocer los rostros de las personas como forma de autenticación de identidad se reutiliza para la identificación biométrica remota en tiempo real de pequeños delincuentes en un recinto de conciertos.

Recursos COBRA C (Lista de factores de riesgo que surgen en el contexto de despliegue del sistema de IA)

Ejemplos de posibles factores de riesgo	
1. Privacidad y protección de datos personales	C.1.1 ¿Trata el sistema de IA datos personales, incluida la categoría de datos sensibles (como los previstos, por ejemplo, en el artículo 6 del Convenio 108+ o en otros marcos pertinentes)? En caso afirmativo, ¿qué riesgos pueden surgir en relación con los derechos de privacidad y de protección de datos de las personas potencialmente afectadas y con las responsabilidades de quienes desarrollan, despliegan o utilizan el sistema en materia de protección de datos y de gobernanza de la información? <i>Si el sistema de IA trata datos personales, existe el riesgo de que dicho tratamiento no se ajuste a la legislación o a las normas aplicables en materia de protección de datos personales y de privacidad. Entre los riesgos figuran la recogida y el uso ilícitos o desleales, la insuficiencia de las salvaguardas o de la rendición de cuentas en el manejo de los datos personales y la divulgación de datos personales en los resultados.</i>
	C.1.2 ¿Está diseñado el sistema de IA para la selección de contenidos, la elaboración de perfiles, la predicción o la orientación del comportamiento dirigidas a personas concretas? En caso afirmativo, ¿qué riesgos pueden surgir, incluso si las personas potencialmente afectadas no pueden acceder a información suficientemente accesible al respecto? <i>Si el sistema de IA está diseñado para la curación de contenidos, la elaboración de perfiles, la predicción o la orientación del comportamiento dirigidas a personas concretas, existe el riesgo de que las personas afectadas no disfruten del máximo nivel de transparencia y de derechos de información previsto en el marco jurídico aplicable (en particular cuando intervienen datos sensibles, tal como se definen en instrumentos como el Convenio 108+). En concreto, pueden carecer de: a) información clara sobre la recogida y el uso de sus datos personales; b) la motivación de los resultados del sistema, explicada en un lenguaje sencillo y no técnico; c) la finalidad de la selección de contenidos, la elaboración de perfiles, la predicción o la orientación del comportamiento; d) las categorías de personas u organismos a los que pueden comunicarse sus datos, su perfil o los resultados del tratamiento.</i>
2. No discriminación y sesgo	C.2.1 ¿Cómo pueden las modalidades de despliegue —incluidas dónde, cómo y por quién se implanta el sistema de IA— influir en la aparición o la amplificación de sesgos o de efectos discriminatorios detectados en fases anteriores del ciclo de vida? <i>Los sesgos y los patrones discriminatorios detectados durante el diseño o el desarrollo pueden verse reforzados o mitigados en función de cómo se despliegue el sistema de IA.</i>
3. Operadores	C.3.1 ¿Qué posibles riesgos para los derechos humanos de quienes operan el sistema pueden derivarse del despliegue del sistema de IA? <i>El despliegue del sistema de IA puede afectar negativamente a los derechos humanos de quienes lo operan (por ejemplo, al generar una vigilancia excesiva, un estrés indebido, condiciones de trabajo inseguras o la erosión de la autonomía profesional). Tales impactos podrían comprometer sus derechos humanos, pero también la integridad y la rendición de cuentas en el funcionamiento del sistema.</i>
4. Formación	C.4.1 ¿Están los operadores del sistema suficientemente formados para comprender plenamente las limitaciones del sistema y para intervenir de manera eficaz en situaciones de complejidad, incertidumbre, anomalía o fallo? En caso negativo, ¿qué riesgos pueden surgir a este respecto para los derechos humanos, la democracia y el Estado de derecho y para la rendición de cuentas del sistema? <i>Si los operadores del sistema de IA no están suficientemente formados, pueden no llegar a comprender los usos previstos y las limitaciones del sistema, o no reconocer situaciones de complejidad, incertidumbre, anomalías o fallos. Ello podría impedirles ejercer un juicio y una intervención humanos adecuados, lo que aumentaría el riesgo de daño para las personas potencialmente afectadas y menoscabaría la rendición de cuentas, la transparencia y la fiabilidad del sistema.</i>

5. Intervención humana en el proceso	C.5.1	¿Presenta el sistema de IA un elevado grado de automatización o de «autonomía» operativa? En caso afirmativo, ¿qué riesgos podrían derivarse de interacciones con las personas afectadas que no estén plenamente mediadas por el control humano? <i>Los sistemas de IA con un elevado grado de automatización o de «autonomía» operativa pueden interactuar con las personas potencialmente afectadas de maneras no plenamente mediadas por el control humano. Ello genera riesgos de menoscabo de la dignidad humana y de la autonomía individual, así como de erosión de la transparencia, la supervisión, la rendición de cuentas y la responsabilidad. Si bien la sustitución completa de los operadores humanos puede producirse solo en determinados contextos, existen numerosos supuestos en los que los operadores humanos siguen interviniendo, pero confían cada vez más en exceso en los resultados de la IA. Ello puede dar lugar a una sensación falsa o «ilusoria» de supervisión humana efectiva, en la que la tarea la ejecuta en gran medida el sistema, aunque exista una supervisión humana nominal. A falta de garantías adecuadas, este tipo de interacciones podrían afectar negativamente a los derechos humanos de las personas potencialmente afectadas.</i>
6. Usos ajenos al ámbito previsto	C.6.1	¿Cómo podrían los fallos de funcionamiento, el uso indebido o la aplicación malintencionada del sistema de IA afectar a los derechos humanos de las personas potencialmente afectadas, a la democracia y al Estado de Derecho? ¿En qué medida están identificados estos riesgos? <i>Existe el riesgo de que el sistema de IA funcione incorrectamente o de que se utilice de forma indebida o abusiva, por ejemplo, por avería o fallo técnico, por un uso que exceda de su ámbito previsto o por una aplicación deliberadamente malintencionada. Tales hechos podrían tener un efecto adverso sobre los derechos humanos de las personas potencialmente afectadas, en particular al menoscabar la transparencia, la supervisión, la rendición de cuentas y la responsabilidad del sistema. Esta categoría de riesgo podría afectar además a la autonomía individual y, por tanto, repercutir en la democracia y el Estado de derecho cuando se aplique a gran escala, por ejemplo, mediante la difusión de información errónea y de desinformación.</i>
7. Proximidad a la decisión o a la actuación	C.7.1	¿En qué medida el sistema de IA adopta o revisa directamente decisiones o actuaciones que puedan afectar a los derechos humanos, a la democracia o al Estado de Derecho, o las fundamenta sustancialmente? ¿Cómo influye este grado de intervención en el nivel potencial de riesgo? <i>El grado de proximidad del sistema de IA con la actividad en cuestión (decisión o actuación) puede amplificar o reducir su impacto potencial sobre los derechos humanos. Cuando el sistema adopta o revisa directamente decisiones, o las fundamenta sustancialmente, el riesgo de impactos adversos sobre los derechos humanos puede ser mayor que en el caso de los sistemas que solo prestan un apoyo periférico o que no ejercen influencia significativa alguna.</i>

Recursos COBRA E (Ámbitos ilustrativos²¹ de posible preocupación desde el punto de vista de los derechos humanos, la democracia y el Estado de Derecho)

Esta sección ofrece una herramienta que puede utilizarse para realizar u orientar la evaluación COBRA. Expone los ámbitos de posible preocupación (columna de la izquierda) con, para cada uno de ellos, una descripción indicativa, ejemplos de posibles riesgos relacionados con la IA, los ámbitos potencialmente pertinentes y las referencias a los instrumentos de derechos humanos potencialmente pertinentes y, en su caso, a otros instrumentos jurídicos.

La lista puede entenderse como un conjunto ilustrativo de elementos que pueden emplearse para facilitar la evaluación de los riesgos relacionados con la IA en distintos contextos. El recurso tiene carácter indicativo, no exhaustivo y abierto, lo que significa que está llamado a evolucionar en función de los avances que se produzcan en los ámbitos de la tecnología y de los regímenes jurídicos

²¹ Las referencias que se hacen en este cuadro a los instrumentos internacionales de derechos humanos se incluyen a título ilustrativo. Dichas referencias solo son aplicables a los Estados que son partes en esos instrumentos. Se espera que cada Estado aplique su propia legislación aplicable de conformidad con sus obligaciones jurídicas internacionales, lo que podría incluir alentar al sector privado a respetar y apoyar los derechos humanos, también en los términos establecidos en los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos.

pertinentes. Se han seleccionado diecisiete ámbitos de posible preocupación. Estos ámbitos representan aspectos fundamentales consagrados en diversos instrumentos internacionales de derechos humanos, y ponen de relieve la protección de la dignidad humana y la autonomía individual.

1. Integridad física y mental y dignidad humana	Descripción ilustrativa
	Las disposiciones de derechos humanos ofrecen diversas protecciones en relación con la integridad física y mental de las personas, y refuerzan la importancia de la autonomía personal e individual y de la disposición sobre uno mismo.
	Muchas de estas protecciones pueden aplicarse, por ejemplo, en situaciones que impliquen el uso de la fuerza por parte de autoridades que ejercen competencias policiales y cuando las personas se hallan bajo custodia.
	En algunos países estas protecciones pueden incluir la investigación de la conducta de las autoridades públicas en situaciones que pongan en peligro la salud o la vida. Las protecciones de los derechos humanos pueden extenderse a otras muchas situaciones en las que pueda estar en juego la integridad física y mental de las personas.
	Ejemplos de posibles riesgos relacionados con la IA
	Los riesgos que la IA plantea para la integridad personal, física y mental, así como para la dignidad humana, pueden derivarse de cuestiones como la falta de transparencia en la toma de decisiones, los defectos en el diseño del sistema (incluidos los casos en que dicho diseño no es adecuado a la edad), el desarrollo o el uso de sistemas de IA y la posibilidad de que los sistemas impulsados por IA deshumanicen y cosifiquen a las personas. Pueden surgir riesgos asimismo de sistemas de IA que manipulan a las personas.
	Obligaciones de derechos humanos potencialmente pertinentes
	Pueden resultar pertinentes las disposiciones jurídicas siguientes:
	PIDCP: ²² artículos 6, 7, 9 y 10
	PIDESC: ²³ artículos 7 (condiciones de trabajo equitativas y satisfactorias) y 12 (salud)
	Convenio Europeo de Derechos Humanos: ²⁴ artículos 2 (derecho a la vida), 3 (prohibición de la tortura) y 13 (derecho a un recurso efectivo);
	CSE: ²⁵ artículos 7 (protección de los niños y adolescentes), 8 (protección de la maternidad), 11 (protección de la salud), 13 (asistencia social y médica), 14 (servicios sociales), 16 (protección de la familia), 17 (protección de los niños), 19 (protección de los trabajadores migrantes y sus familias), 23 (protección social de las personas de edad avanzada), 26 (dignidad en el trabajo), 30 (protección contra la pobreza y la exclusión social) y 31 (vivienda);
	Carta de la UE: ²⁶ artículos 1 (dignidad humana), 3 (derecho a la integridad de la persona) y 4 (prohibición de la tortura y de las penas o los tratos inhumanos o degradantes);
	Pacto de San José: ²⁷ artículos 3 (derecho al reconocimiento de la personalidad jurídica), 4 (derecho a la vida) y 5 (derecho a la integridad personal), y artículo 12 (derecho a la alimentación) del Protocolo de San Salvador

²² El Pacto Internacional de Derechos Civiles y Políticos de las Naciones Unidas (ONU) y sus protocolos facultativos.

²³ El Pacto Internacional de Derechos Económicos, Sociales y Culturales de la ONU y su protocolo facultativo.

²⁴ El Convenio del Consejo de Europa para la Protección de los Derechos Humanos y de las Libertades Fundamentales (STE núm. 5) y sus protocolos adicionales.

²⁵ La Carta Social Europea (STE núm. 35) y sus protocolos, y la Carta Social Europea revisada (STE núm. 163).

²⁶ La Carta de los Derechos Fundamentales de la Unión Europea.

²⁷ La Convención Americana sobre Derechos Humanos y su protocolo adicional.

	Descripción ilustrativa
	Pueden plantearse cuestiones de derechos humanos y de Estado de Derecho en relación con el trato físico de las personas en situaciones de detención, privación de libertad, sanción o trata de seres humanos. Existen protecciones frente a la detención o privación de libertad arbitrarias, la tortura y las penas o los tratos crueles, inhumanos o degradantes, las restricciones injustificadas de la libertad de circulación y la trata de seres humanos, así como para garantizar el acceso a recursos y a la justicia.
	Ejemplos de posibles riesgos relacionados con la IA
2. Libertad y seguridad físicas, circulación y residencia	La recogida de datos personales o el uso de herramientas de IA pueden permitir inferir comportamientos o actividades, o dar lugar a tales inferencias, que se utilicen para justificar o facilitar actuaciones como la detención o la privación de libertad, o restricciones de la circulación y la residencia, aun cuando esas inferencias sean incorrectas; las decisiones incorrectas adoptadas o fundamentadas por herramientas de IA sobre las que se actúa sin un examen ulterior suficiente pueden dar lugar a un trato indebido o menoscabar la capacidad de la persona acusada de impugnar la legalidad de su privación de libertad. Los sistemas de IA pueden utilizarse para facilitar o permitir la trata de personas, por ejemplo, generando contenidos que sirvan para captar o atraer a personas hacia situaciones de explotación. Al igual que ocurre con la trata de seres humanos en general, las mujeres y los niños son quienes corren mayor riesgo. La IA puede utilizarse además para llevar a cabo una vigilancia masiva e identificar a personas sin un objetivo justificable o razonablemente previsible.
	Obligaciones de derechos humanos potencialmente pertinentes
3. Justicia y administración de justicia	Pueden resultar pertinentes las siguientes disposiciones jurídicas: PIDCP: artículos 7, 8, 9, 10, 11, 12, 13 y 14 Convenio 105 de la Organización Internacional del Trabajo (OIT) (abolición del trabajo forzoso) Convenio Europeo de Derechos Humanos: artículos 3 (prohibición de la tortura), 4 (prohibición de la esclavitud y del trabajo forzoso), 5 (derecho a la libertad y a la seguridad) y 13 (derecho a un recurso efectivo), y artículo 2 (libertad de circulación) del Protocolo n.º 4 CSE: artículo 1.2 (derecho al trabajo); Carta de la UE: artículos 3 (derecho a la integridad de la persona), 4 (prohibición de la tortura y de las penas o los tratos inhumanos o degradantes), 5 (prohibición de la esclavitud y del trabajo forzado), 6 (derecho a la libertad y a la seguridad) y 45 (libertad de circulación y de residencia) Pacto de San José: artículos 5 (derecho a la integridad personal), 6 (prohibición de la esclavitud y servidumbre), 7 (derecho a la libertad personal) y 22 (derecho de circulación y de residencia).
	Descripción ilustrativa
	El acceso a la justicia y un conjunto complejo de reglas mínimas que regulan el acceso a la justicia en el ámbito más amplio de la justicia y la administración pública. Las protecciones de los derechos humanos en este ámbito abarcan diversos derechos y garantías relativos al acceso a recursos, a la calidad del examen de los asuntos por los tribunales (por ejemplo, la exigencia de que la sentencia se pronuncie públicamente), a la calidad de la participación en el proceso judicial (por ejemplo, la exigencia de una vista equitativa y pública) y a los requisitos relativos a la composición de los tribunales (por ejemplo, la exigencia de un tribunal competente, independiente e imparcial). No obstante, en muchos países un amplio abanico de asuntos se resuelve por órganos de la administración pública, y los ciudadanos acceden a los tribunales en una fase posterior del procedimiento respecto de determinadas materias. La administración pública está dotada de potestades decisorias que le permiten influir de manera significativa en las personas y en sus intereses vitales importantes a este respecto.

4. Privacidad y protección de datos	Ejemplos de posibles riesgos relacionados con la IA El uso de la IA en el sistema judicial introduce riesgos como el sesgo nocivo, la falta de transparencia o la insuficiencia de la supervisión humana, el uso de datos defectuosos (algo especialmente problemático en el contexto de la administración de justicia) y la erosión de la independencia judicial y la pérdida de confianza pública en las instituciones judiciales. Suscita además preocupación por las vulneraciones de las garantías procesales que podrían dar lugar a violaciones de los derechos humanos.
	Obligaciones de derechos humanos potencialmente pertinentes Pueden resultar pertinentes las siguientes disposiciones jurídicas: PIDCP: artículos 2.3, 9, 14 y 26 Convenio Europeo de Derechos Humanos: artículos 5 (derecho a la libertad y a la seguridad), 6 (derecho a un proceso equitativo), 8 (derecho al respeto a la vida privada y familiar), 13 (derecho a un recurso efectivo) y artículo 3 del Protocolo n.º 7 (indemnización en caso de error judicial) Carta de la Unión Europea: artículos 41 (derecho a una buena administración), 42 (derecho de acceso a los documentos), 47 (derecho a la tutela judicial efectiva y a un juez imparcial) y 48 (presunción de inocencia y derechos de la defensa) Pacto de San José: artículos 8 (garantías judiciales), 10 (derecho a indemnización), 24 (igualdad ante la ley) y 25 (protección judicial)
	Descripción ilustrativa Los derechos de privacidad y de protección de datos garantizan que existan límites a la injerencia externa en los asuntos privados y que las personas puedan evitar que sus datos personales, comportamientos y decisiones sean divulgados o vigilados sin su consentimiento, con lo que se salvaguardan la autonomía personal y la dignidad. En el contexto específico del sector público, la legislación sobre privacidad y protección de datos protege la información personal de las personas al garantizar que las autoridades públicas o las entidades que actúan por cuenta de estas traten únicamente la información personal para la que están autorizadas. Las protecciones de la privacidad permiten además a las personas mantener el control sobre el modo en que desarrollan, definen y presentan su identidad personal. Estas protecciones amparan frente al uso no autorizado o la tergiversación de los atributos personales de una persona, como su nombre, su imagen u otras características distintivas. Existen asimismo aspectos de la privacidad y de la autonomía individual que guardan relación con la interacción con los demás. Las protecciones de los derechos humanos en este ámbito abarcan diversos derechos y garantías relativos a la vida familiar, el matrimonio y, de manera más general, las relaciones con los demás, tanto en el ámbito privado como en el laboral.
	Ejemplos de posibles riesgos relacionados con la IA La recogida de ingentes cantidades de datos personales (incluso cuando esos datos se recogen de manera transparente), así como el uso de técnicas de IA para inferir información personal a partir de ellos, pueden generar riesgos graves tanto para la privacidad como para la identidad. Desde la vigilancia invasiva y la reidentificación de datos anonimizados hasta el uso indebido de datos biométricos, la creación de ultrafalsificaciones (deepfakes), la usurpación de identidad y la manipulación del comportamiento, las tecnologías de IA pueden menoscabar la autonomía personal y exponer información sensible.
	Obligaciones de derechos humanos potencialmente pertinentes

5. Igualdad y no discriminación ³⁰	Pueden resultar pertinentes las disposiciones jurídicas siguientes: PIDCP: artículo 17 CDN:²⁸ artículo 16 Convenio Europeo de Derechos Humanos: artículos 8 (derecho al respeto a la vida privada y familiar), 10 (libertad de expresión) y 12 (derecho a contraer matrimonio) Carta de la Unión Europea: artículos 7 (respeto de la vida privada y familiar), 8 (protección de datos de carácter personal) y 9 (derecho a contraer matrimonio y derecho a fundar una familia) Convención 108+:²⁹ artículos 3, 5, 6 y 9 Pacto de San José: artículos 11 (protección de la honra y de la dignidad), 17 (protección a la familia), 18 (derecho al nombre) y 20 (derecho a la nacionalidad)
	Descripción ilustrativa Los derechos a la igual protección de la ley y a no sufrir discriminación en el disfrute de los derechos y libertades complementan las demás disposiciones sustantivas de diversos instrumentos internacionales de derechos humanos. La discriminación puede ser directa o indirecta, derivarse de una relación de asociación, basarse en uno o varios motivos, o fundarse en una acción o en una omisión. A título de referencia general, las características protegidas mencionadas en el cuerpo del derecho vigente de los derechos humanos —integrado por instrumentos jurídicos internacionales (tanto de alcance mundial como regional) y nacionales, según sean aplicables en cada Estado— pueden incluir, entre otras, las siguientes: (1) El sexo (2) La "Raza"³¹ y color (3) La lengua (4) La religión (5) Las opiniones políticas o de otro tipo (6) El origen nacional o social (7) La pertenencia a una minoría social (8) La riqueza (9) El nacimiento (10) La edad (11) La identidad y la expresión de género (12) La orientación sexual (13) Las características sexuales (14) La salud y la discapacidad (15) La situación parental y el estado civil (16) La situación migratoria (17) La situación relacionada con el empleo.

²⁸ La Convención de las Naciones Unidas sobre los Derechos del Niño y sus protocolos facultativos.

²⁹ El Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal, en su versión modificada (STE núm. 108, STCE núm. 223), y sus protocolos.

³⁰ Las disposiciones sobre igualdad y no discriminación complementan otras disposiciones sustantivas de diversos instrumentos internacionales de derechos humanos, por lo que estas cuestiones resultan pertinentes respecto de todos los ámbitos de posible preocupación recogidos en este cuadro.

³¹ Dado que todos los seres humanos pertenecen a la misma especie, se rechazan las teorías basadas en la existencia de distintas «razas». No obstante, el término «raza» se emplea para garantizar que las personas a las que general y erróneamente se percibe como «pertenecientes a otra raza» no queden excluidas de la protección dispensada.

5. Igualdad y no discriminación	El respeto de la igualdad puede ir más allá de la prohibición de un trato menos favorable, sin justificación, basado en una o varias características protegidas.
	Ejemplos de posibles riesgos relacionados con la IA
	Las tendencias discriminatorias pueden crearse o agravarse por: la finalidad de un sistema, si esa finalidad es en sí misma discriminatoria; conjuntos de datos que no son suficientemente representativos o que codifican desigualdades históricas; sistemas algorítmicos que no tienen suficientemente en cuenta los patrones de sesgo³² presentes en los conjuntos de datos o que generan resultados sesgados de otro modo; unas pruebas y una evaluación inadecuadas en materia de sesgo; la falta de representación o de consideración (con independencia de que se produzca por negligencia o de manera intencionada) de los puntos de vista de grupos diversos durante el desarrollo de la IA u otras actividades relacionadas con ella; una formación insuficiente y otras medidas de gobernanza deficientes (como unos procesos de supervisión inadecuados) para los desarrolladores o los usuarios de sistemas de IA en relación con la posible discriminación. Al mismo tiempo, un sistema entrenado con datos teóricamente exentos por completo de sesgo puede aun así producir impactos injustos aguas abajo por razones relacionadas con el modo en que se despliega.
6. Descripción ilustrativa	Obligaciones de derechos humanos potencialmente pertinentes
	Pueden resultar pertinentes las siguientes disposiciones jurídicas: PIDCP: artículos 2, 3, 23, 24 y 26 PIDESC: artículos 2.2 y 3 Convenios de la OIT núm. 111 (discriminación), núm. 190 (violencia y acoso en el mundo del trabajo) Convenio Europeo de Derechos Humanos: artículo 14 (prohibición de discriminación), Protocolo núm. 12, artículo 1 CSE: artículos 1.2, 4.3, 19.4-5 y 20 Carta de la UE: diversas disposiciones del capítulo III sobre la igualdad (artículos 20 a 26), artículos 14 (derecho a la educación), 15 (libertad profesional y derecho a trabajar) y 16 (libertad de empresa) Pacto de San José: artículo 24 (igualdad ante la ley) e instrumentos jurídicos específicos como la ICERD,³³ la CEDAW,³⁴ la CDN y la CDPD³⁵

³² En el contexto de la IA, el sesgo designa un resultado desviado de manera injusta. Es probable que el sesgo dé lugar a resultados discriminatorios siempre que a) guarde relación con características protegidas y b) se emprenda algún tipo de actuación a partir del resultado sesgado de un sistema de IA (con independencia de que el sesgo se haya introducido en el sistema a través de los datos de entrenamiento o de ajuste fino, del propio modelo o del modo en que una persona interpreta el resultado del modelo).

³³ La Convención Internacional de las Naciones Unidas sobre la Eliminación de Todas las Formas de Discriminación Racial.

³⁴ La Convención de las Naciones Unidas sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer y su protocolo facultativo.

³⁵ La Convención de las Naciones Unidas sobre los Derechos de las Personas con Discapacidad y su protocolo facultativo.

Las disposiciones de derechos humanos protegen la facultad de la persona de profesar, practicar y expresar sus creencias religiosas, o la ausencia de estas, sin injerencia ni discriminación. Comprenden la protección de la facultad de rendir culto, de cambiar de religión o de convicciones y de observar prácticas religiosas (como la vestimenta, las restricciones alimentarias o las festividades religiosas) en público o en privado. Amparan asimismo frente a la coacción para adoptar o renunciar a una religión y garantizan la igualdad de trato con independencia de la afiliación religiosa, sea esta real o supuesta.

La libertad de conciencia protege la facultad de sostener convicciones profundas sobre lo que está bien y lo que está mal, no necesariamente basadas en sistemas religiosos, y de actuar conforme a ellas. Cabe citar como ejemplos el ateísmo, el vegetarianismo o la objeción de conciencia al servicio militar.

Aunque existe cierto debate sobre el contenido de la libertad de pensamiento, esta podría comprender el derecho a no tener que revelar los propios pensamientos, a no ser sancionado por ellos y a la protección frente a una alteración inadmisible del pensamiento.

Ejemplos de posibles riesgos relacionados con la IA

Los riesgos que la IA plantea para los derechos religiosos pueden derivarse de algoritmos que crean o perpetúan un sesgo nocivo que afecta a las personas por sus creencias religiosas, reales o percibidas, mediante violaciones de la privacidad o facilitando la elaboración de perfiles religiosos, la supresión de la expresión religiosa y la manipulación de las creencias. Estos riesgos amenazan la facultad de las personas de practicar y expresar su religión sin temor a la discriminación, la vigilancia o la coacción.

Además, los sistemas de IA, como los sistemas de reconocimiento de emociones y los que perpetúan la desinformación, podrían interferir en la libertad de pensamiento o de conciencia y producir un efecto disuasorio que podría interferir en dichas libertades.

El uso de sistemas de IA puede comprometer la autonomía individual en materia de convicciones al configurar sutilmente las percepciones, influir en los procesos de pensamiento y restringir potencialmente la libertad de formar y expresar convicciones religiosas o espirituales sin injerencias externas indebidas.

Obligaciones de derechos humanos potencialmente pertinentes

Pueden resultar pertinentes las siguientes disposiciones jurídicas

PIDCP: artículos 18, 19 y 20

Convenio Europeo de Derechos Humanos: artículos 9 (libertad de pensamiento, de conciencia y de religión), 10 (libertad de expresión) y 14 (prohibición de discriminación)

Carta de la UE: artículos 10 (libertad de pensamiento, de conciencia y de religión) y 11 (libertad de expresión y de información), y capítulo III sobre la igualdad

Pacto de San José: artículos 12 (libertad de conciencia y de religión) y 13 (libertad de pensamiento y de expresión)

CDN: artículo 14

7. Opiniones, expresión e información	Descripción ilustrativa
	La protección de las opiniones y de la libre expresión se aplica a profesar, buscar, difundir y recibir informaciones e ideas por cualquier medio. Las formas de expresión pueden ser de naturaleza política, artística —incluida la producción de obras teatrales y espectáculos—, personal o comercial, y pueden comprender la información periodística, las fotografías y determinadas conductas, como los boicots o las campañas, la vestimenta, los símbolos, etc. Estas protecciones pueden aplicarse asimismo a determinadas relaciones regidas por el derecho privado (como las relaciones laborales) y a las manifestaciones vertidas en correspondencia privada o en reuniones a puerta cerrada.
	Ejemplos de posibles riesgos relacionados con la IA
	Los riesgos que la IA plantea para las opiniones, la expresión y el acceso a la información pueden incluir la posible censura, el sesgo de los sistemas algorítmicos empleados para la selección de contenidos y la creación y difusión de información inexacta. Además, la vigilancia facilitada por la IA puede coartar la libertad de expresión, mientras que el uso de algoritmos puede reducir el acceso a todos los puntos de vista. El uso de sistemas de IA puede menoscabar la autonomía individual en la expresión al influir sutilmente en las opiniones, configurar el discurso y limitar la capacidad de formar, articular y compartir libremente ideas sin manipulación ni constricción encubiertas. A partir del modo en que una persona se expresa públicamente, un sistema de IA puede realizar inferencias incorrectas, y ello podría alterar la información o el acceso de los que de otro modo esa persona podría beneficiarse (por ejemplo, una persona que viste de determinada manera podría ser clasificada como perteneciente a un grupo al que no pertenece).
	Obligaciones de derechos humanos potencialmente pertinentes
	Pueden resultar pertinentes las siguientes disposiciones jurídicas PIDCP: artículos 19 y 20 CDN: artículo 13 Convenio Europeo de Derechos Humanos: artículo 10 (libertad de expresión) Carta de la UE: artículos 10 (libertad de pensamiento, de conciencia y de religión), 11 (libertad de expresión y de información) y 42 (derecho de acceso a los documentos) Pacto de San José: artículos 13 (libertad de pensamiento y de expresión) y 14 (derecho de rectificación o respuesta)
8. Reunión pacífica y asociación	Descripción ilustrativa
	Existen protecciones de los derechos humanos respecto de las concentraciones y los desplazamientos no violentos, incluidas las reuniones en lugares privados y públicos.
	Existen protecciones para las agrupaciones voluntarias con un fin común y para la posibilidad de constituir un grupo u organización que persiga determinados objetivos, o de afiliarse a él. Son ejemplos destacados de tales asociaciones los partidos políticos, las asociaciones minoritarias y religiosas y los sindicatos.
	Ejemplos de posibles riesgos relacionados con la IA
	Entre los posibles riesgos de la IA en este ámbito pueden figurar la vigilancia arbitraria o discriminatoria de los miembros de una asociación y de los participantes en reuniones, la vigilancia policial predictiva, la censura en las redes sociales, la información errónea y el acoso selectivo, factores todos ellos que pueden disuadir a las personas de ejercer sus derechos de reunión pacífica y de libre asociación y de participar en los asuntos públicos.
	Obligaciones de derechos humanos potencialmente pertinentes

	Pueden resultar pertinentes las disposiciones jurídicas siguientes: PIDCP: artículo 21 CDN: artículo 15 Convenio Europeo de Derechos Humanos: artículo 11 (libertad de reunión y de asociación) CSE: artículos 5, 6.1-2, 21, 28 y 29 Carta de la UE: artículos 12 (libertad de reunión y de asociación) y 28 (derecho de negociación y de acción colectiva) Pacto de San José: artículos 15 (derecho de reunión) y 16 (libertad de asociación)
9. Propiedad	Descripción ilustrativa
	Las protecciones de los derechos humanos en este ámbito no se limitan por lo general a la titularidad de bienes materiales, sino que pueden extenderse a otros derechos e intereses, incluidos los derechos de propiedad intelectual, que constituyen activos que también pueden considerarse "derechos de propiedad".
	Ejemplos de posibles riesgos relacionados con la IA
	Los derechos de propiedad pueden verse afectados negativamente por los sistemas de IA a través de cuestiones como: ▶ el uso de algoritmos defectuosos en el sector financiero e inmobiliario para adoptar decisiones automatizadas sobre el valor de los bienes, los préstamos o el crédito; ▶ los contenidos generados por IA, como obras plásticas, música o textos, y el desarrollo de modelos de IA, que pueden suscitar cuestiones complejas en materia de derechos de propiedad intelectual; ▶ sin supervisión humana, los rápidos avances de la tecnología de IA podrían superar el ritmo de los marcos jurídicos existentes y dejar lagunas en la protección de los derechos de propiedad; ▶ el uso de propiedad intelectual para entrenar modelos allí donde la legislación no es clara o se aplica de forma dispar en cuanto a si tal uso está permitido. Estos riesgos pueden menoscabar el control de las personas sobre sus bienes y sus creaciones intelectuales y dar lugar a posibles situaciones de explotación y discriminación.
	Obligaciones de derechos humanos potencialmente pertinentes
	Pueden resultar pertinentes las disposiciones jurídicas siguientes: Convenio Europeo de Derechos Humanos: artículo 1 del Protocolo n.º 1 (protección de la propiedad) Carta de la UE: artículo 17 (derecho a la propiedad); Pacto de San José: artículo 21 (derecho a la propiedad privada)
10. Educación	Descripción ilustrativa
	Las protecciones de los derechos humanos en este ámbito se refieren al acceso de las personas a una educación de calidad y a la prestación de esta sin discriminación. Las protecciones pueden abarcar no solo el acceso a la enseñanza primaria, sino también a la secundaria, la superior, la formación profesional y las oportunidades de aprendizaje a lo largo de la vida.
	Ejemplos de posibles riesgos relacionados con la IA
	Entre los riesgos relacionados con la IA figura el sesgo nocivo que producen los algoritmos en ámbitos como la admisión, la calificación y el aprendizaje personalizado. Otros riesgos relacionados con la IA (incluidos los derivados de la falta de alfabetización digital) pueden menoscabar la eficacia y la equidad de los sistemas educativos, de formación profesional y de capacitación, con repercusión en la experiencia global de aprendizaje del alumnado o en sus oportunidades en el mercado de trabajo.
	Obligaciones de derechos humanos potencialmente pertinentes

11. Artes, ciencias, cultura y lengua	Pueden resultar pertinentes las disposiciones jurídicas siguientes: PIDESC: artículo 13 CDN: artículo 28 Convenio Europeo de Derechos Humanos: artículo 2 del Protocolo n.º 1 (derecho a la instrucción) CSE: artículos 1.4, 7.6, 9, 10, 15, 17, 19 (11-12), 20.b y 30 Carta de la UE: artículo 14 (derecho a la educación) Pacto de San José: artículo 13 (derecho a la educación) del Protocolo de San Salvador
	Descripción ilustrativa Las protecciones de los derechos humanos en esta esfera permiten a todas las personas el acceso a diversos recursos culturales, el disfrute del progreso científico y de sus aplicaciones y de los beneficios que de ellos resulten, y la protección de los intereses morales y materiales derivados de las producciones científicas, literarias o artísticas de que sean autoras. Estas protecciones amparan asimismo la libertad de investigación científica y de actividad creadora. Protegen además la facultad de las personas, incluidos los miembros de minorías étnicas, religiosas o lingüísticas, de disfrutar de su cultura, practicar su religión y emplear su lengua.
	Ejemplos de posibles riesgos relacionados con la IA Las consideraciones relativas a la IA en el ámbito de la cultura, las artes y las ciencias pueden incluir cuestiones complejas en materia de propiedad intelectual, problemas de exactitud de los datos y de control de calidad, fuentes de datos poco fiables y riesgos para el pensamiento crítico y para las prácticas éticas de investigación. Estos riesgos pueden menoscabar la riqueza y la diversidad de la expresión humana, la creatividad y el avance del conocimiento.
	Obligaciones de derechos humanos potencialmente pertinentes Pueden resultar pertinentes las siguientes disposiciones jurídicas PIDCP: artículo 27 PIDESC: artículo 15 CSE: artículos 15.3, 22.c, 23 y 30 Carta de la UE: artículos 13 (libertad de las artes y de las ciencias) y 22 (diversidad cultural, religiosa y lingüística) Pacto de San José: artículo 14 (derecho a los beneficios de la cultura) del Protocolo de San Salvador
	Descripción ilustrativa Los derechos laborales protegen a los trabajadores en el lugar de trabajo en lo relativo al trato justo, a unos entornos de trabajo seguros y saludables, a unas condiciones de trabajo saludables, seguras y dignas, a unos salarios razonables, a la privacidad en el lugar de trabajo, a un nivel de vida digno, a no sufrir discriminación, a un trato justo y equitativo y a la facultad de constituir sindicatos y afiliarse a ellos. Los derechos laborales pueden comprender, por ejemplo, las normas mínimas en materia de salarios y de nivel de vida digno, los límites a la jornada de trabajo, la protección frente al trabajo forzoso, las cuestiones relativas al trabajo infantil, la libertad sindical y el reconocimiento efectivo del derecho de negociación colectiva y la eliminación de la discriminación.
12. Trabajo y empleo	Ejemplos de posibles riesgos relacionados con la IA La IA puede tener impactos transformadores de diversa índole en el trabajo y el empleo. Entre los riesgos relacionados con la IA figura su uso para vigilar y gestionar a los trabajadores, lo que suscita preocupación por la privacidad y genera entornos estresantes debido al seguimiento constante del rendimiento, el comportamiento y la productividad, así como el posible desplazamiento de puestos de trabajo por efecto de la IA. Los sistemas de IA utilizados para fundamentar o adoptar decisiones sobre contratación, despido o promoción entrañan el riesgo de dar lugar a prácticas laborales sesgadas, opacas o injustas, incluidas las que afectan a las personas que trabajan en el ámbito de la IA. La IA puede utilizarse para automatizar y optimizar las cadenas de suministro, lo que podría aumentar

	de forma involuntaria la demanda de mano de obra barata, incluido el trabajo infantil en determinados sectores. El uso de la IA en la vigilancia y la gestión del trabajo puede reducir la supervisión humana, con lo que podrían pasar inadvertidas prácticas de explotación, en particular en lo relativo al trabajo infantil.
	Obligaciones de derechos humanos potencialmente pertinentes Pueden resultar pertinentes las disposiciones jurídicas siguientes: PIDESC: artículos 6, 7, 8 y 10 CDN: artículo 32 CSE: artículos 1 a 10, 21, 22 a 26, 28 y 29 Carta de la UE: artículos 8 (protección de datos de carácter personal), 12 (libertad de reunión y de asociación), 15 (libertad profesional y derecho a trabajar), 27 (derecho a la información y consulta de los trabajadores en la empresa), 28 (derecho de negociación y de acción colectiva), 29 (derecho de acceso a los servicios de colocación), 30 (protección en caso de despido injustificado), 31 (condiciones de trabajo justas y equitativas), 32 (prohibición del trabajo infantil y protección de los jóvenes en el trabajo) y 33 (vida familiar y vida profesional) Pacto de San José: artículos 7 (condiciones justas, equitativas y satisfactorias de trabajo) y 8 (derechos sindicales) del Protocolo de San Salvador Además, pueden resultar pertinentes en este contexto los siguientes convenios de la Organización Internacional del Trabajo (OIT): 1. Convenio núm. 87 – Libertad sindical y protección del derecho de sindicación (1948) 2. Convenio núm. 98 – Derecho de sindicación y de negociación colectiva (1949) 3. Convenio núm. 111 – Discriminación (empleo y ocupación) (1958) 4. Convenio núm. 138 – Edad mínima (1973) 5. Convenio núm. 155 – Seguridad y salud de los trabajadores (1981) 6. Convenio núm. 190 – Violencia y acoso (2019)
12. Trabajo y empleo	
	Descripción ilustrativa El sector de la salud comprende el acceso a los servicios médicos, las iniciativas de salud pública y las infraestructuras sanitarias destinadas a garantizar el bienestar físico y mental de las personas. Las protecciones en este ámbito guardan relación por lo general con la posibilidad de que las personas accedan a la asistencia sanitaria para disfrutar del más alto nivel posible de salud física y mental. El sector de la seguridad social facilita el acceso a apoyo y protección económicos a las personas que se enfrentan al desempleo, la discapacidad, la vejez o dificultades económicas, mediante prestaciones y programas de bienestar social. Las protecciones en este ámbito guardan relación por lo general con la posibilidad de que las personas reciban esa asistencia. La protección social, la asistencia social y la protección contra la pobreza y la exclusión social tienen por objeto suprimir o reducir la pobreza y la desigualdad y promover la dignidad humana.
13. Salud, asistencia sanitaria y seguridad social	
	Ejemplos de posibles riesgos relacionados con la IA

	El uso de sistemas de IA puede dar lugar a datos sesgados de forma nociva que provoquen un acceso desigual a la asistencia sanitaria o a las prestaciones sociales, en particular para los grupos marginados. El uso de la IA en el tratamiento de datos personales sensibles de salud y de carácter social plantea riesgos significativos para la privacidad y la seguridad (incluidas cuestiones relativas al consentimiento y a la autonomía individual), como las violaciones de la seguridad de los datos o su uso indebido. Una confianza excesiva en la IA para los diagnósticos médicos o las decisiones sobre prestaciones sociales podría dar lugar a errores, reducir la supervisión humana y llegar a perjudicar los resultados de los pacientes o a denegar indebidamente prestaciones. Por lo que respecta a la protección social y la seguridad social, el uso de la IA puede plantear riesgos cuando se emplea para determinar si las personas cumplen los requisitos para acceder a prestaciones sociales o para detectar los casos en que estas se abonaron indebidamente y la autoridad tiene derecho a reclamar su reintegro: las consecuencias de los errores (posiblemente provocados por sesgos) pueden ser graves y a menudo afectan en particular a personas en situación de vulnerabilidad.
	Obligaciones de derechos humanos potencialmente pertinentes Pueden resultar pertinentes las disposiciones jurídicas siguientes: PIDESC: artículos 9, 10, 11 y 12 CSE: artículos 2, 3, 7, 8, 11, 12, 13, 14, 15, 17.1, 19.2, 22.b, 23, 27.1.b y 30 Código Europeo de Seguridad Social (CESS): artículos 7 a 12 (asistencia médica), 13 a 18 (prestaciones monetarias de enfermedad), 19 a 24 (prestaciones de desempleo), 25 a 30 (prestaciones de vejez), 31 a 38 (prestaciones en caso de accidente del trabajo y de enfermedad profesional), 46 a 52 (prestaciones de maternidad), 53 a 58 (prestaciones de invalidez) y 59 a 64 (prestaciones de supervivencia) CDN: artículo 24 Carta de la UE: artículos 8 (protección de datos de carácter personal), 34 (seguridad social y ayuda social) y 35 (protección de la salud) Pacto de San José: artículos 9 (derecho a la seguridad social) y 10 (derecho a la salud) del Protocolo de San Salvador.
14. Infancia	Descripción ilustrativa Respecto de este grupo existen tanto protecciones generales de derechos humanos como protecciones específicas de la infancia. Su finalidad es garantizar que los niños puedan crecer, aprender, jugar, desarrollarse y prosperar con dignidad, y que sus vulnerabilidades y necesidades específicas se tengan debidamente en cuenta. En todas las medidas concernientes a los niños, su interés superior debe ser una consideración primordial. Teniendo presente lo anterior, muchos de los ámbitos de posible preocupación que se recogen en este cuadro resultarán pertinentes para la infancia.
	Ejemplos de posibles riesgos relacionados con la IA Los sistemas de IA podrían plantear riesgos para los niños si se emplean para adoptar decisiones que no tengan en cuenta sus necesidades o derechos, así como mediante la creación de imágenes o vídeos falsos pero realistas que puedan confundir, inducir a error o dañar a los niños. Los contenidos generados, seleccionados o recomendados por sistemas de IA (por ejemplo) en redes sociales o en plataformas de vídeo pueden exponer a los niños a contenidos ilícitos, nocivos o no adecuados a su edad. La publicidad selectiva y los contenidos basados en IA pueden explotar las vulnerabilidades de los niños e influir en su comportamiento y sus preferencias, incluso con ánimo de lucro. Las plataformas de redes sociales impulsadas por IA pueden contribuir a la ansiedad, la depresión y una autoestima negativa en los niños a través del ciberacoso, la adicción y las comparaciones nocivas. Los niños son especialmente vulnerables a los daños vinculados a la IA generativa. Los niños tienen menos capacidad para distinguir los contenidos sintéticos de los auténticos, para detectar la información inexacta y los momentos en que están siendo manipulados mediante patrones oscuros, y para comprender que están interactuando con una máquina y no con un ser humano. La IA se utiliza cada vez más para generar material de abuso sexual infantil y para explotar a los niños. La IA generativa puede posibilitar la creación de material sintético de abuso sexual infantil y causar daños a los niños y a las familias, así como a la sociedad. Los depredadores pueden además servirse de la IA generativa para captar, extorsionar y explotar a las víctimas.

15. Medio ambiente	Obligaciones de derechos humanos potencialmente pertinentes
	PIDESC: artículos 10.3 y 12 CSE: artículos 7 y 17 CSE: artículos 39 a 45 (prestaciones familiares) Carta de la UE: artículos 7, 8, 24 y 32 Pacto de San José: artículo 19 (derechos del niño) Además de las disposiciones de derechos humanos contenidas en los tratados internacionales de derechos humanos, pueden resultar pertinentes los siguientes instrumentos especializados de derechos humanos: 1) la Convención de las Naciones Unidas sobre los Derechos del Niño, de 1989 (CDN), y sus protocolos facultativos; 2) la Observación general núm. 25 relativa a los derechos de los niños en relación con el entorno digital; 3) el Convenio del Consejo de Europa para la protección de los niños contra la explotación y el abuso sexual, de 2007 (STCE núm. 201, Convenio de Lanzarote).
	Descripción ilustrativa
	En algunos ordenamientos, las disposiciones de derechos humanos pueden reconocer cierto grado de responsabilidad de los poderes públicos en la protección de las personas frente a los daños ambientales.
	Ejemplos de posibles riesgos relacionados con la IA
	Los sistemas de IA, en particular los que emplean recursos computacionales considerables para el entrenamiento y la inferencia, consumen grandes cantidades de energía. Ello contribuye a las emisiones de carbono, especialmente cuando se alimentan de fuentes de energía no renovable. Del mismo modo, la fabricación del hardware necesario para entrenar y ejecutar los sistemas de IA requiere numerosas tierras raras, cuya extracción tiene en algunos casos efectos contaminantes de mayor alcance. Las aplicaciones de la IA en la agricultura, la economía, la gestión de recursos o el seguimiento de la fauna silvestre, aun ofreciendo numerosos beneficios ambientales potenciales, pueden tener también impactos ambientales imprevistos si no se gestionan con cuidado, como la alteración de los ecosistemas, la sobreexplotación de los recursos naturales, la contaminación o el daño a la biodiversidad.
	Obligaciones de derechos humanos potencialmente pertinentes
	Desde una óptica de derechos humanos, las cuestiones relativas a la protección del medio ambiente podrían abordarse al amparo de las disposiciones siguientes: PIDESC: artículo 12 Convenio Europeo de Derechos Humanos: artículos 2 (derecho a la vida), 3 (prohibición de la tortura), 6 (derecho a un proceso equitativo), 8 (derecho al respeto a la vida privada y familiar), 10 (libertad de expresión) y 11 (libertad de reunión y de asociación) Carta de la UE: artículo 37 (protección del medio ambiente) Pacto de San José: artículo 11 (derecho a un medio ambiente sano) del Protocolo de San Salvador Convenio de Aarhus³⁶

³⁶ El Convenio de la CEPE sobre el acceso a la información, la participación del público en la toma de decisiones y el acceso a la justicia en materia de medio ambiente.

	Descripción ilustrativa
16. Democracia	Las disposiciones de derechos humanos protegen la facultad de las personas de participar en la dirección de los asuntos públicos, de votar en elecciones libres y justas, de presentarse a cargos públicos y de acceder a los servicios públicos. Un rasgo importante de los sistemas democráticos de gobierno es el pluralismo político, que se garantiza en gran medida mediante la protección de los derechos humanos cuyo respeto resulta esencial para una democracia próspera, como la libertad de expresión, la libertad de asociación y la libertad de reunión pacífica; y mediante la existencia de unos medios de comunicación plurales e independientes y de un abanico de partidos políticos que representen distintos intereses y puntos de vista, un acceso equitativo al debate público y a la toma de decisiones públicas y una participación efectiva en ellos, y el acceso a información exacta y fiable. Los derechos que rigen la participación efectiva en los procesos públicos de toma de decisiones y en los asuntos de gobierno garantizan que las personas puedan intervenir en el proceso político, incluso votando, presentándose a cargos públicos y expresando opiniones sobre las políticas públicas. El fomento de la gobernanza democrática, la rendición de cuentas y la representación permite a los ciudadanos influir en las decisiones que afectan a sus vidas y a sus comunidades.
	Ejemplos de posibles riesgos relacionados con la IA
	Los sistemas de IA que desempeñan un papel en la influencia sobre los procesos democráticos o en su fundamentación podrían afectar negativamente al acceso equitativo de las personas al debate público y a unas elecciones libres y justas, y a su participación en ellos, así como a su capacidad de formarse una opinión libremente, a través de los siguientes factores, que constituyen una lista no exhaustiva: a) el engaño, la información errónea o la desinformación en los planos local, nacional o mundial provocados por el despliegue del sistema; b) la manipulación en los planos local, nacional o mundial posibilitada por el despliegue del sistema; c) la intimidación o el control del comportamiento, en los planos local, nacional o mundial, posibilitados por el despliegue del sistema. La IA podría utilizarse de maneras que interfieran en la participación en los asuntos de gobierno, por ejemplo, mediante el empleo de algoritmos para cribar a los candidatos políticos o a los participantes en procesos públicos, o para suprimir manifestaciones sobre cuestiones de política pública.
	Obligaciones de derechos humanos potencialmente pertinentes
	Pueden resultar pertinentes las disposiciones jurídicas siguientes: PIDCP: artículos 19, 20, 21 y 25 Convenio Europeo de Derechos Humanos: artículos 10 (libertad de expresión) y 11 (libertad de reunión y de asociación), y artículo 3 del Protocolo n.º 1 (derecho a elecciones libres) CSE: artículos 5 y 6.1-2 Carta de la UE: artículos 11 (libertad de expresión y de información), 12 (libertad de reunión y de asociación), 39 (derecho de sufragio activo y pasivo en las elecciones al Parlamento Europeo), 40 (derecho de sufragio activo y pasivo en las elecciones municipales) y 42 (derecho de acceso a los documentos) Pacto de San José: artículos 13 (libertad de pensamiento y de expresión), 14 (derecho de rectificación o respuesta), 15 (derecho de reunión), 16 (libertad de asociación) y 23 (derechos políticos)
	Descripción ilustrativa
17. Estado de Derecho	El Estado de derecho es un principio de gobernanza según el cual todas las personas, instituciones y entidades, públicas y privadas, incluido el propio Estado, están sometidas a leyes promulgadas públicamente, aplicadas por igual y sujetas a control judicial independiente, y compatibles con el respeto de los derechos humanos internacionales. La independencia judicial es un fundamento de la democracia y de la confianza pública en la administración de justicia. Las instituciones del Estado de derecho son organizaciones (el poder judicial, el legislativo y el ejecutivo, la abogacía y órganos semiindependientes como los organismos anticorrupción, las autoridades de protección de datos, los defensores del pueblo, etc.) y sistemas (los ordenamientos jurídicos) que garantizan que las leyes se apliquen de manera justa, coherente y transparente, protegiendo los derechos de las personas y manteniendo el orden en la sociedad. En conjunto, estas instituciones mantienen el Estado de derecho al garantizar que sea la ley, y no las decisiones arbitrarias, la que rija la

sociedad, y que todas las personas reciban el mismo trato ante la ley.

Ejemplos de posibles riesgos relacionados con la IA

El uso de la IA plantea riesgos (por la posible falta de transparencia y de una supervisión humana suficiente, por el uso de sistemas que dan lugar a consecuencias no previstas, por el sesgo nocivo o por las amenazas a la ciberseguridad, entre otros factores) para el Estado de derecho, la equidad procesal, la transparencia y la rendición de cuentas en contextos como:

- (a) la integridad de las instituciones y los procesos democráticos, incluido el principio de separación de poderes, y el respeto de la independencia judicial;
- (b) el acceso a la justicia;
- (c) los mecanismos de rendición de cuentas (control del poder ejecutivo), y los organismos y políticas anticorrupción.

Obligaciones de derechos humanos potencialmente pertinentes

El ejercicio y la protección efectivos de estos derechos son esenciales para garantizar el respeto del Estado de derecho.

PIDCP: artículos 2.3, 4 y 14

Convenio Europeo de Derechos Humanos: artículos 6 (derecho a un proceso equitativo) y 13 (derecho a un recurso efectivo)

Carta de la UE: artículos 41 (derecho a una buena administración), 42 (derecho de acceso a los documentos), 47 (derecho a la tutela judicial efectiva y a un juez imparcial) y 48 (presunción de inocencia y derechos de la defensa)

Pacto de San José: artículos 8 (garantías judiciales), 10 (derecho a indemnización) y 25 (protección judicial)

Recursos COBRA F (Lista ilustrativa³⁷ de sectores/esferas y ámbitos de posible preocupación)

Este recurso enumera (sin orden de prioridad) los sectores y esferas de posible preocupación desde la perspectiva de los derechos humanos, la democracia y el Estado de derecho (columnas primera y segunda) y está concebido para utilizarse junto con los **recursos COBRA E**, con el fin de llevar a cabo la cartografía de impactos como parte de COBRA. Del mismo modo, no es exhaustivo y debería revisarse periódicamente.

1. Administración pública	Esferas	Ámbitos de posible preocupación según los recursos COBRA E
	a) Asistencia sanitaria , incluidas, entre otras, cuestiones como el acceso a los servicios sanitarios, el diagnóstico, el pronóstico y la atención preventiva, la administración de tratamientos de soporte vital, el tratamiento de enfermedades potencialmente mortales, los servicios de urgencias, el asesoramiento y el tratamiento en salud mental y las decisiones sobre el final de la vida.	(1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (13) Salud, asistencia sanitaria y seguridad social/protección social; (14) Infancia; (15) Medio ambiente.
	b) Vida familiar y atención social , incluidas, entre otras, cuestiones como el disfrute mutuo entre padres e hijos, la custodia, el régimen de visitas, los derechos de comunicación, el acogimiento público, las familias de acogida, la adopción y los servicios reproductivos, y el acceso a las prestaciones públicas y su concesión.	(1) Integridad física y mental y dignidad humana; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (9) Propiedad; (13) Salud, asistencia sanitaria y seguridad social/protección social; (14) Infancia.
	c) Migración y control de fronteras , incluidas, entre otras, cuestiones como la expulsión, la extradición, la devolución, los cambios de estatuto, la denegación del derecho de entrada, la notificación de derechos, los servicios de traducción e interpretación, la elaboración de transcripciones, la obtención y valoración de pruebas y las condiciones y modalidades de entrada en el territorio del Estado y de salida de él.	(1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (12) Trabajo y empleo; (13) Salud, asistencia sanitaria y seguridad social/protección social; (14) Infancia; (15) Medio ambiente; (17) Estado de Derecho.
	d) Desarrollo y mantenimiento de infraestructuras , incluidas, entre otras, cuestiones como la seguridad sanitaria y el disfrute del espacio público, el transporte público y la movilidad, la gestión de los riesgos ambientales, la ordenación del territorio y el urbanismo, la vivienda, las infraestructuras digitales y la gestión y el consumo de energía.	(1) Integridad física y mental y dignidad humana; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (9) Propiedad; (13) Salud, asistencia sanitaria y seguridad social/protección social; (14) Infancia; (15) Medio ambiente.
	e) Servicios de emergencia , incluidas, entre otras, cuestiones como la gestión de las operaciones de rescate, las infraestructuras de comunicaciones de emergencia y la gestión de las consecuencias de las catástrofes.	(1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (14) Infancia; (15) Medio ambiente.

³⁷ Las referencias que se hacen en este cuadro a los instrumentos internacionales de derechos humanos se incluyen a título ilustrativo. Dichas referencias solo son aplicables a los Estados que son partes en esos instrumentos. Se espera que cada Estado aplique su propia legislación aplicable de conformidad con sus obligaciones jurídicas internacionales, lo que podría incluir alentar al sector privado a respetar y apoyar los derechos humanos, también en los términos establecidos en los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos.

f) Educación pública, incluidas, entre otras, cuestiones como el acceso a los centros educativos y las evaluaciones educativas, y el reconocimiento oficial de los estudios.	(5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (10) Educación; (14) Infancia; (11) Artes, ciencias, cultura y lengua.
g) Empleo, incluidas, entre otras, cuestiones como la selección de personal, el acceso al empleo, la gestión del rendimiento y las políticas de personal, y la accesibilidad y los ajustes razonables para las personas con discapacidad.	1) Integridad física y mental y dignidad humana; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (8) Reunión pacífica y asociación; (12) Trabajo y empleo; (13) Salud, asistencia sanitaria y seguridad social/protección social.

	Esferas	Ámbitos de posible preocupación según los recursos COBRA E
2. Aplicación de la ley y seguridad	a) Policía y servicios asimilados, incluidas, entre otras, cuestiones como el uso de la fuerza letal, el empleo de la fuerza física durante las detenciones, los controles de identidad y la identificación de personas con fines policiales, los programas de protección de personas en peligro (como las víctimas de violencia doméstica o los testigos protegidos), las detenciones y privaciones de libertad, la gestión de los programas de evaluación de idoneidad de los funcionarios, la gestión de las operaciones de rescate y de liberación de rehenes, el control de multitudes durante actos públicos, la vigilancia policial predictiva, el análisis de emociones y sentimientos, y la vigilancia y las restricciones aplicadas por la policía y otros organismos encargados de la aplicación de la ley.»	(1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (3) Justicia y administración de justicia; (4) Vida privada y protección de datos; (5) Igualdad y no discriminación; (7) Opiniones, expresión e información; (8) Reunión pacífica y asociación; (9) Propiedad; (14) Infancia; (15) Medio ambiente; (17) Estado de Derecho
	b) Acción penal, incluidas, entre otras, cuestiones como la obtención y valoración de pruebas.	(1) Integridad física y psíquica y dignidad humana; (3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (7) Opiniones, expresión e información; (14) Infancia; (15) Medio ambiente; (17) Estado de Derecho.
3. Administración de justicia	a) Tribunales y justicia, incluidas, entre otras, cuestiones como las detenciones, las privaciones de libertad, las decisiones sobre fianza, libertad condicional, excarcelación condicional y uso de dispositivos electrónicos de control, la notificación de derechos y resoluciones, los servicios de traducción e interpretación, la elaboración de transcripciones, la obtención y valoración de pruebas (incluida la valoración de la credibilidad de los testigos y de las pruebas), la concesión de asistencia jurídica gratuita, la determinación de cualquier acusación penal, la determinación de derechos y obligaciones de carácter civil, las decisiones sobre recusación de jueces o de miembros del jurado, las decisiones sobre el acceso a la segunda instancia, la imposición de penas y los procedimientos automatizados.	(2) Libertad y seguridad físicas, circulación y residencia; (3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (7) Opiniones, expresión e información; (9) Propiedad; (14) Infancia; (17) Estado de derecho.
	b) Aspectos institucionales de la organización del poder judicial, a saber, el respeto de la independencia judicial e, incluidas, entre otras, cuestiones como la gestión del proceso de evaluación de idoneidad, el nombramiento y el cese de jueces y funcionarios judiciales, la atribución de asuntos a jueces o funcionarios judiciales concretos y la gestión de los asuntos en el procedimiento judicial.	(3) Justicia y administración de justicia; (5) Igualdad y no discriminación; (9) Propiedad; (12) Trabajo y empleo; (17) Estado de Derecho.

4. Procesos democráticos	a) Sistema electoral , incluidas, entre otras, cuestiones como las condiciones y modalidades de ejercicio del derecho de voto, la edad electoral, las reglas de exclusión, las condiciones y modalidades de votación, los métodos y procedimientos de votación, las condiciones y modalidades de escrutinio, el derecho a presentarse a las elecciones, la organización de elecciones y referendos, la delimitación de circunscripciones, la gestión de los contenciosos electorales y los recursos efectivos al respecto, y la difusión de información electoral.	(4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (7) Opiniones, expresión e información; (8) Reunión pacífica y asociación; (16) Democracia.
	b) Instituciones y procesos políticos , incluidas, entre otras, cuestiones como la supremacía de la Constitución, el papel del poder judicial en el equilibrio de poderes, la independencia judicial y la delegación de la función legislativa.	(3) Justicia y administración de justicia; (5) Igualdad y no discriminación; (7) Opiniones, expresión e información; (16) Democracia; (17) Estado de derecho.

	Esferas	Ámbitos de posible preocupación según los recursos COBRA E
4. Procesos democráticos	c) Opiniones y discurso público , incluidas, entre otras, cuestiones como la expresión de manifestaciones protegidas en sus diversas formas, la protección de las fuentes periodísticas, las actividades de recopilación de información, el acceso, la recogida y el tratamiento automatizado de datos, la propiedad, las actividades de investigación e indagación, el régimen de divulgación de la información recibida con carácter confidencial, la protección de los denunciantes de irregularidades, y la democracia participativa y las consultas públicas, incluidas las cuestiones relativas a la organización de reuniones de comisiones.	(3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (7) Opiniones, expresión e información; (8) Reunión pacífica y asociación; (10) Educación; (11) Artes, ciencias, cultura y lengua; (14) Infancia; (16) Democracia; (17) Estado de derecho.
	d) Reunión pacífica y asociación , incluidas, entre otras, cuestiones como el momento, el lugar y la forma de celebración de las reuniones, las condiciones y modalidades para constituir un grupo u organización que persiga determinados objetivos o afiliarse a él, la vigilancia de las reuniones y la identificación de los participantes, y la participación de los ciudadanos en la vida pública.	(5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (7) Opiniones, expresión e información; (8) Reunión pacífica y asociación; (9) Propiedad; (11) Artes, ciencias, cultura y lengua; (14) Infancia; (16) Democracia.
	e) Acceso a la información , incluidas, entre otras, cuestiones como el acceso a la información personal, a la información financiera y a la información sobre las operaciones mercantiles de las personas, el deber de facilitar información fiable y precisa, las responsabilidades en materia de verificación y transmisión de la información, y el acceso a la información en poder del Estado.	(3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (7) Opiniones, expresión e información; (9) Propiedad; (10) Educación; (11) Artes, ciencias, cultura y lengua; (14) Infancia; (16) Democracia; (17) Estado de derecho.

	f) Medios de comunicación , incluidas, entre otras, cuestiones como la transparencia respecto de la propiedad de los medios, el pluralismo mediático, la libertad de expresión durante las elecciones (fuera de línea y en línea), los deberes y responsabilidades de los portales de noticias en internet, la generación automatizada de noticias, las plataformas de medios, la información errónea y la desinformación, y la moderación de contenidos en línea.	(5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (7) Opiniones, expresión e información; (8) Reunión pacífica y asociación; (10) Educación; (11) Artes, ciencias, cultura y lengua; (14) Infancia; (16) Democracia
5. Prisiones y libertad vigilada	a) Gestión de prisiones y centros de privación de libertad , incluidas, entre otras, cuestiones como la elaboración de perfiles de los reclusos, el cribado psicológico de los internos potencialmente vulnerables, la gestión de los reclusos peligrosos, la gestión de la población penitenciaria, los registros de visitantes e internos y la vigilancia de las comunicaciones.	(1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (7) Opiniones, expresión e información; (13) Salud, asistencia sanitaria y seguridad social/protección social; (15) Medio ambiente.
	b) Servicios de libertad condicional y de libertad vigilada , incluidas, entre otras, cuestiones como la libertad condicional, la excarcelación condicional, el seguimiento de las personas y los dispositivos electrónicos que estas puedan portar.	(1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación.
6. Servicios esenciales prestados por el sector	a) Servicios de comunicación	(4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (7) Opiniones, expresión e información; (9) Propiedad; (10) Educación; (11) Artes, ciencias, cultura y lengua; (14) Infancia.
	b) Educación y formación profesional	(5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (7) Opiniones, expresión e información; (10) Educación; (11) Artes, ciencias, cultura y lengua; (12) Trabajo y empleo; (14) Infancia.

6. Servicios esenciales prestados por el sector privado	Esferas	Ámbitos de posible preocupación según los recursos COBRA E
	c) Aplicaciones biomédicas, ciencias de la vida, epidemiología y asistencia sanitaria	(1) Integridad física, psíquica y moral, y (1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (9) Propiedad; (13) Salud, asistencia sanitaria y seguridad social/protección social; (14) Infancia; (15) Medio ambiente

d) Gestión ambiental y de residuos	(1) Integridad física y mental y dignidad humana; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (13) Salud, asistencia sanitaria y seguridad social/protección social; (15) Medio ambiente.
e) Gestión de la energía	(5) Igualdad y no discriminación
f) Infraestructuras y planificación urbanas	(4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (9) Propiedad; (14) Infancia; (15) Medio ambiente.
g) Fabricación y automatización industrial	(1) Integridad física y mental y dignidad humana; (12) Trabajo y empleo; (15) Medio ambiente.
h) Construcción y edificación	(4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (12) Trabajo y empleo; (15) Medio ambiente.
i) Seguridad y protección pública	(1) Integridad física y mental y dignidad humana; (2) Libertad y seguridad físicas, circulación y residencia; (3) Justicia y administración de justicia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (8) Reunión pacífica y asociación; (15) Medio ambiente.
(j) Domótica (tecnologías del hogar inteligente)	(4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (14) Infancia.
k) Vivienda y provisión de alojamiento social	(1) Integridad física y mental y dignidad humana; (5) Igualdad y no discriminación; (13) Salud, asistencia sanitaria y seguridad social/protección social.
l) Empleo, recursos humanos y gestión laboral , incluidas, entre otras, cuestiones como la selección de personal, el acceso al empleo, la gestión del rendimiento y las políticas de personal, y la accesibilidad y los ajustes razonables para las personas con discapacidad.	(2) Libertad y seguridad físicas, circulación y residencia; (4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (10) Educación; (12) Trabajo y empleo.
m) Servicios financieros	(4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (9) Propiedad.
n) Tecnologías de la información y redes	(4) Privacidad y protección de datos; (5) Igualdad y no discriminación; (7) Opiniones, expresión e información; (9) Propiedad; (10) Educación; (11) Artes, ciencias, cultura y lengua; (14) Infancia.
o) Fabricación de vehículos e infraestructuras de transporte	(2) Libertad y seguridad físicas, circulación y residencia; (5) Igualdad y no discriminación; (15) Medio ambiente.
p) Agricultura y abastecimiento alimentario	(1) Integridad física y mental y dignidad humana; (5) Igualdad y no discriminación; (6) Pensamiento, conciencia, religión y convicciones; (12) Trabajo y empleo; (13) Salud, asistencia sanitaria y seguridad social/protección social; (15) Medio ambiente.

Agentes de ventas de las publicaciones del Consejo de Europa

BÉLGICA/BELGIQUE

La Librairie Européenne - The European Bookshop Rue de l'Orme, 1

BE-1040 BRUSELAS Tel.: + 32 (0)2 231 04 35

Fax: + 32 (0)2 735 08 60

Correo electrónico: info@libeurop.eu

<http://www.libeurop.be>

Jean De Lannoy/DL Services c/o Michot Warehouses

Bergense steenweg 77 Chaussée de Mons

BE-1600 SINT PIETERS LEEUW Fax: + 32 (0)2 706 52 27

Correo electrónico: jean.de.lannoy@dl-servi.com

<http://www.jean-de-lannoy.be>

CANADÁ

Renouf Publishing Co. Ltd. 22-1010 Polytek Street CDN-

OTTAWA, ONT K1J 9J1 Tel.: + 1 613 745 2665

Fax: + 1 613 745 7660

Tel. gratuito: (866) 767-6766

Correo electrónico: order.dept@renoufbooks.com

<http://www.renoufbooks.com>

FRANCIA

Póngase en contacto directamente /

Merci de contacter directement Council of Europe

Publishing Éditions du Conseil de l'Europe F-67075

STRASBOURG Cedex Tel.: + 33 (0)3 88 41 25 81

Correo electrónico: publishing@coe.int

<http://book.coe.int>

Librairie Kléber

1, rue des Francs-Bourgeois F-67000 ESTRASBURGO

Tel.: + 33 (0)3 88 15 78 88

Fax: + 33 (0)3 88 15 78 80

Correo electrónico: librairie-kleber@coe.int

<http://www.librairie-kleber.com>

NORUEGA/NORVÈGE

Akademika

Postboks 84 Blindern NO-0314 OSLO

Tel.: + 47 2 218 8100

Fax: + 47 2 218 8103

Correo electrónico: support@akademika.no

<http://www.akademika.no>

POLONIA/POLOGNE

Ars Polona JSC

25 Obroncow Street

PL-03-933 WARSZAWA Tel.: + 48 (0)22 509 86 00

Fax: + 48 (0)22 509 86 10

Correo electrónico: arspolona@arspolona.com.pl

<http://www.arspolona.com.pl>

PORTUGAL

Marka Lda

Rua dos Correeiros 61-3 PT-1100-162 LISBOA Tel: 351 21

3224040

Fax: 351 21 3224044

Correo electrónico: apoio.clientes@marka.pt

www.marka.pt

SUIZA/SUISSE

Planetis Sàrl

16, chemin des Pins CH-1273 ARZIER

Tel.: + 41 22 366 51 77

Fax: + 41 22 366 51 78

Correo electrónico: info@planetis.ch

REINO UNIDO/ROYAUME-UNI

Williams Lea TSO 18 Central Avenue

St Andrews Business Park Norwich

NR7 0HR

Reino Unido

Tel. +44 (0)333 202 5070

Correo electrónico: customer.services@tso.co.uk

<http://www.tsoshop.co.uk>

ESTADOS UNIDOS y CANADÁ/ ÉTATS-UNIS et CANADA

Manhattan Publishing Co 670 White Plains Road

USA-10583 SCARSDALE, NY Tel: + 1 914 472 4650

Fax: + 1 914 472 4316

Correo electrónico: coe@manhattanpublishing.com

<http://www.manhattanpublishing.com>

Ediciones del Consejo de Europa/Éditions du Conseil de l'Europe

F-67075 Estrasburgo Cedex

Tel.: + 33 (0)3 88 41 25 81 – Correo electrónico: publishing@coe.int – Sitio web: <http://book.coe.int>

Elaborada por el **Comité sobre Inteligencia Artificial (CAI) del Consejo de Europa**, esta publicación reúne la **metodología HUDERIA** y el **modelo HUDERIA: recursos de análisis de riesgos basado en el contexto (COBRA)** como recurso práctico para quienes deseen evaluar y gestionar los riesgos e impactos de los sistemas de IA en los **derechos humanos, la democracia y el Estado de derecho**. Explica los conceptos básicos y los pasos del proceso HUDERIA y ofrece herramientas adaptables y materiales de apoyo que ayudan a trasladar los principios a la práctica cotidiana a lo largo del ciclo de vida de la IA. Concebida para ser **tecnológicamente neutral** y utilizable en distintos sectores, ordenamientos jurídicos y niveles de capacidad, HUDERIA puede ayudar a las autoridades públicas y a los agentes privados a construir una IA de confianza, responsable y respetuosa con los derechos. **Como instrumento no jurídicamente vinculante, HUDERIA puede además ayudar a las Partes en el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho a desarrollar enfoques de evaluación de riesgos e impactos acordes con el Convenio Marco.**

ESP

www.coe.int

El Consejo de Europa es la principal organización de derechos humanos del continente. Está integrado por 46 Estados miembros, incluidos todos los miembros de la Unión Europea. Todos los Estados miembros del Consejo de Europa han firmado el Convenio Europeo de Derechos Humanos, un tratado concebido para proteger los derechos humanos, la democracia y el Estado de derecho. El Tribunal Europeo de Derechos Humanos supervisa la aplicación del Convenio en los Estados miembros.



<http://book.coe.int>
ISBN 978-92-871-9693-4 (PDF)
€9.50/US\$19

