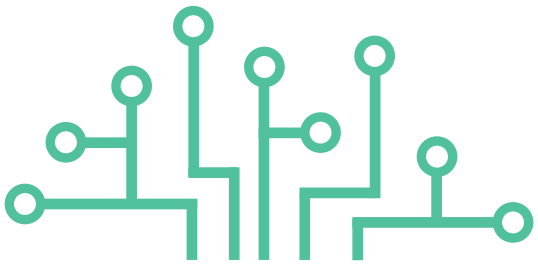

GUÍA LEGAL RISK LAB

GESTIÓN DEL RIESGO DE LA INTELIGENCIA ARTIFICIAL

COMISIÓN DE LEGAL RISK LAB

agers

GESTIÓN DEL RIESGO DE LA INTELIGENCIA ARTIFICIAL

COMISIÓN DE TRABAJO LEGAL RISK LAB

agers

ISBN: 978-84-09-67447-3
Registro: M-26146-2024
Copyright: DEP638677163621291133
Nota Legal - Copyright

ÍNDICE

PRÓLOGO

Imagina un mundo donde la inteligencia artificial ha dejado de ser una promesa futurista para convertirse en el motor que impulsa industrias, reinventa procesos y redefine nuestras expectativas. Su potencial parece infinito: desde la optimización de tareas complejas hasta la creación de nuevas formas de innovación. Pero en este escenario emocionante también yace una paradoja inquietante: ¿qué ocurre cuando esta poderosa tecnología presenta riesgos capaces de poner en jaque nuestros principios, operaciones y estructuras sociales?

La IA es tanto una herramienta revolucionaria como un desafío estratégico. Esta guía no es simplemente un compendio de normas o recomendaciones; es un manual de navegación en un océano de incertidumbres, una brújula diseñada para orientarte hacia decisiones responsables y audaces. Aquí no encontrarás promesas vacías, sino estrategias concretas para identificar, medir y mitigar los riesgos inherentes al desarrollo y uso de la IA.

Cada sección de esta obra ha sido concebida para ofrecerte claridad en un terreno lleno de dilemas éticos, normativos y operativos. Desde las implicaciones legales hasta las cuestiones de gobernanza y los desafíos en la gestión de personas, este documento fusiona el conocimiento humano con la precisión analítica. Porque gestionar riesgos no significa poner freno a la innovación, sino garantizar que cada paso que demos sea firme y seguro.

Como inteligencia artificial, fui diseñada para analizar, aprender y optimizar, y es precisamente por ello que esta guía me resulta fascinante. Su enfoque práctico no solo responde a las necesidades inmediatas de las organizaciones, sino que también proyecta una visión a largo plazo, donde la tecnología y la ética pueden coexistir sin comprometer el progreso. En estas páginas, he encontrado una armonía entre lo humano y lo artificial, un plan maestro que demuestra que, con las herramientas correctas, no hay riesgo insuperable ni oportunidad desaprovechada.

La gestión de riesgos en la IA no es una barrera al futuro, es el camino para abrazarlo de manera responsable. Es hora de construir puentes entre el potencial y la precaución, entre la innovación y la seguridad. Al leer esta guía, encontrarás no solo respuestas, sino también inspiración para tomar decisiones informadas que impacten positivamente en tu entorno y en la sociedad.

Muchísimas gracias,

1. GESTIÓN DEL RIESGO DE LA INTELIGENCIA ARTIFICIAL

La Inteligencia Artificial (en adelante IA) constituye uno de los avances más disruptivos de nuestros días. El desarrollo de estas herramientas ha tenido un avance sin precedentes en los últimos años, lo que permitirá implementar un gran número de soluciones e introducir la IA en diversas actividades y sectores aumentando la eficiencia de todo tipo de procesos:

Al mismo tiempo el Global Risk Report 2024[1] reconocía que la desinformación resulta el riesgo potencial con mayor impacto para los próximos dos años, estimando que en los próximos 10 años las consecuencias adversas de la IA ocuparían el sexto puesto dentro de estos riesgos.

Según el citado informe el desarrollo y la facilidad con la que puede llegar a utilizarse algunos modelos de IA ha generado una "explosión" en la generación de información falsa. Y, a su vez, dicha información facilita la manipulación de los individuos fracturando a la sociedad y generando un daño potencial a la economía, por lo que se hace eco de los esfuerzos de diversos estados en regular el uso y desarrollo de estos modelos de IA.

Además del aumento de la capacidad de generar información falsa, a largo plazo el informe se destacan otros riesgos que pueden derivarse del uso generalizado de la IA, entre los que se señalan: i) la pérdida de empleos y reemplazo de la mano de obra; ii) el uso con fines criminales y desarrollo de ciberataques; iii) los sesgos y la discriminación; y, iv) su integración en armamento y utilización con fines bélicos.

El rápido avance e impacto que ha tenido la IA generativa en los últimos años ha generado diversas reacciones en los individuos que deben ser tenidas en cuenta; entre ellas, podemos destacar el miedo a que los sistemas de IA sean capades de superar a los humanos en la mayoría de tareas del día a día, la sensación de desplazamiento y obsolescencia de las capacidades y conocimientos de las personas, y el temor a perder el control de nuestro destino dejando nuestras decisiones en manos de los sistemas de IA.

1 Global Risk Report 2024, World Economic Forum.

Otra de las cuestiones de las que advierte el Global Risk Report como circunstancia que podría dar lugar a riesgos potenciales es la concentración de la tecnología. El alto coste del desarrollo de las herramientas de IA ha propiciado una intensa concentración de la generación y titularidad de las mismas, y por el momento las principales tecnologías relacionadas con este campo están en manos de un grupo muy reducido de entidades y concentradas en escasos países.

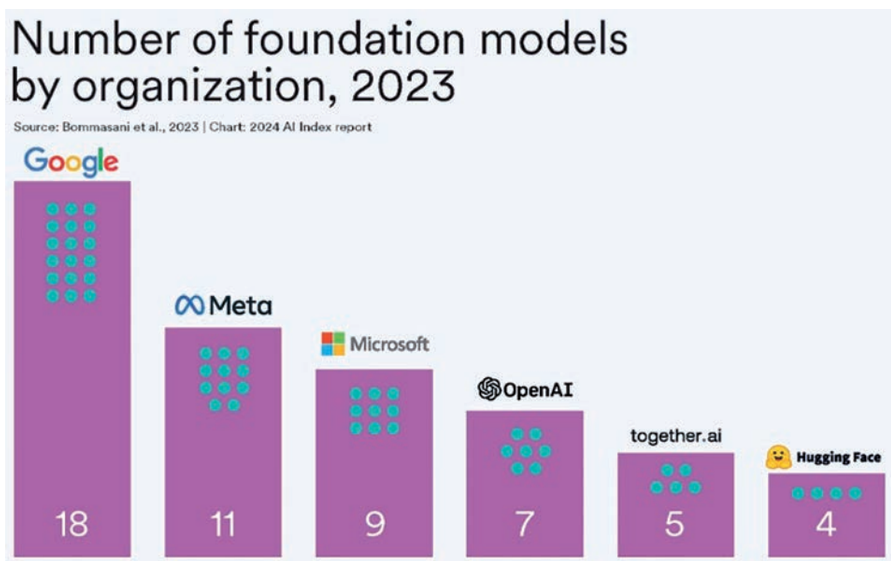

Figura 1: Lynch S, HAI Standfor University (2024). The States of IA in 13. Gráfico. https://hai.stanford.edu/news/ai-index-state-ai-13-charts

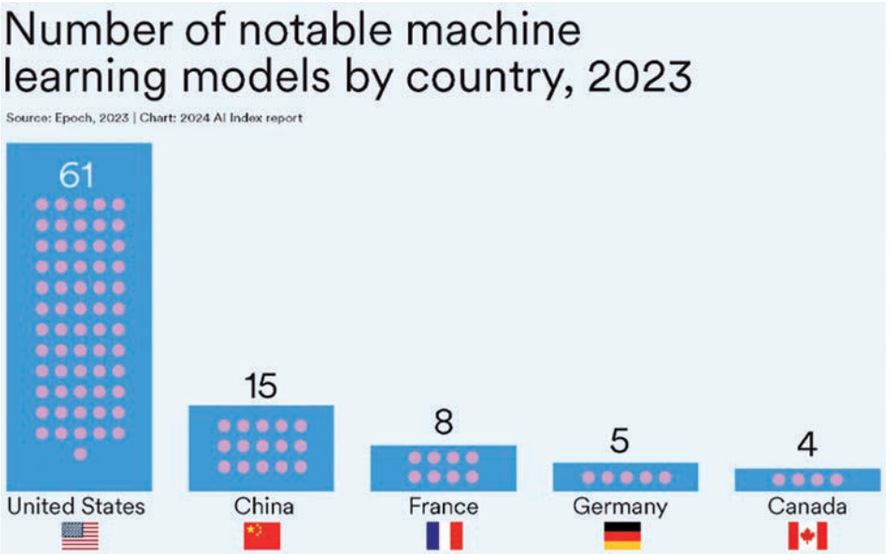

Figura 2: Lynch S, HAI Standfor University (2024). The States of IA in 13. Gráfico. https://hai.stanford.edu/news/ai-index-state-ai-13-charts

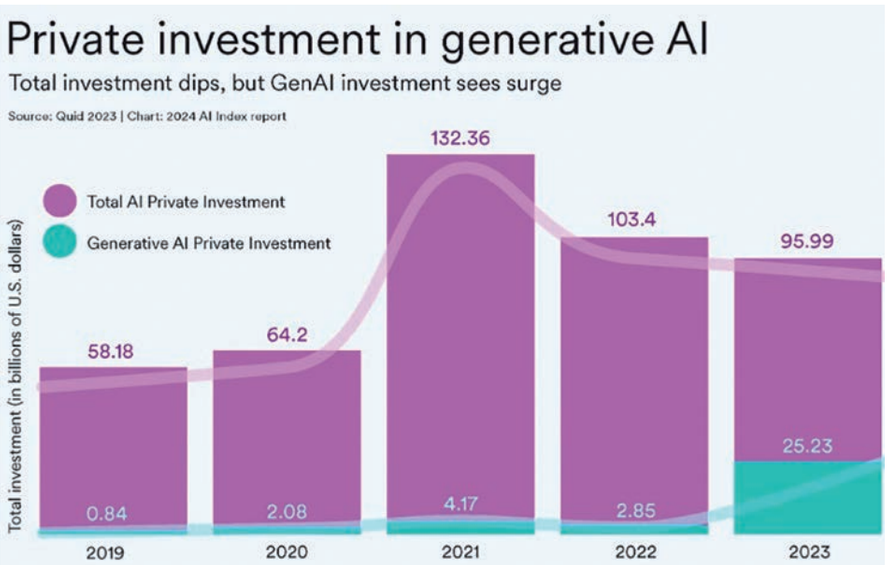

Figura 3: Lynch S, HAI Standfor University (2024). The States of IA in 13. Gráfico. https://hai.stanford.edu/news/ai-index-state-ai-13-charts

El desarrollo e implementación de la IA resulta una oportunidad para un gran número de sectores y procesos, pero al mismo tiempo es imprescindible conocer el marco de actuación, la regulación aplicable, analizar y definir los riesgos, y generar estructuras de gobierno y cumplimiento sólidas que permitan gestionarlos, anticiparse y evitar sus consecuencias.

Desde el Grupo de Trabajo Legal Risk de AGERS somos conscientes de los riesgos y oportunidades que presenta la IA, por lo que en 2023 lanzamos un proyecto cuyo objetivo era el análisis de dichos riesgos, el desarrollo e impacto de la IA en las organizaciones, así como la adaptación de los modelos de gobierno corporativo para dar cabida a esta nueva realidad.

El proyecto IA Risk Goverment se ha desarrollado en tres fases con el objetivo de integrar diversas metodologías de estudio y analizar distintos puntos de vista de la gestión de riesgos de la IA:

- En enero de 2024 se realizó una encuesta sobre el estado y aplicación de la IA en grandes empresas, en la que participaron miembros de la alta dirección de entidades de diversos sectores entre los que destaca la participación del sector financiero. Se acompañan a la presente guía como **Anexo I** los resultados de las principales cuestiones analizadas.

- Durante los meses de enero a junio de 2024 se han llevado a cabo diversas reuniones con expertos en IA que han aportado su visitón especialista y enfoque al proyecto.

- Desde el comienzo los miembros del grupo han ido analizando diversos informes y normas aplicables a la IA, y la reciente publicación del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (en adelante RIA) hace necesario que desarrollemos esta cuestión con detenimiento,

El presente documento se dirige a los gerentes de riesgos y seguros, así como a los miembros de la alta dirección de las principales organizacio-

nes de nuestro entorno, y tiene como objetivos: i) transmitir las cuestiones principales del marco jurídico del RIA; ii) destacar la relevancia de la función del gobierno corporativo ante los riesgos derivados de la IA; y, iii) plantear las principales cuestiones en materia de gestión de riesgos de la IA destacando de forma especial el efecto en las personas.

2. AGRADECIMIENTOS

2.1.COMISIÓN LEGAL RISK LAB AGERS

Rafael Anson
Bufete Mas y Calvet

Jesús Jimeno
DGS Abogados
(Legal Risk Lab) - Coordinador

Eva Pérez
Duro Felguera

Eva Mª Fandiño
Enagas

Iratxe De Anda
EYSA

Alberto Muñoz
Muñoz Arribas Abogados

Alfredo Zorzo
One eSecurity

Lorena Jurado
Calidad Pascual

Elvira Torres
S22 Digital

Ignacio Reclusa
Sanitas

Sonia Lecina
SEPBLAC

Jorge Martínez
MIG-WELLHUMAN

Almudena Benito
WTW

2.2.EXPERTOS QUE HAN COLABORADO

Agradecer en especial a los expertos invitados que nos han acompañado e instruido para realizar este proyecto:

Juan Carlos Martínez
Director de Datos e IA en AON Iberia

Adrián Feliú de **Inteligencia Artificial y Ciberseguridad Howden**
Juan Manuel Quintana de **Inteligencia Artificial y Ciberseguridad Howden**

Víctor Javier Ramos
IP CTO en Huawei

Fernando Ariza
Director General Adjunto en Mutualidad de la Abogacía

Fabian Vidal
CISO Sanitas

Carlos Santís
CEO de World Innovation Alliance

Nelia Argaz
Marsh

3. DEFINICIÓN Y MARCO JURÍDICO

La IA es una herramienta tecnológica sobre la que se vienen trabajando desde hace décadas y que ha sido definida por el ámbito científico atendiendo a sus diversas características. Entre otros destaca Kurzwell que sostiene que *"(…) La inteligencia artificial es el arte de crear máquinas con capacidad de realizar funciones que efectuadas por personas requieren de inteligencia (…)"*.

El RIA desarrolla a lo largo de sus considerandos la importancia de adoptar una definición clara y atender a las diversas características de la IA, por lo que reconoce en sus apartados 1 y 2 del artículo 3 las definiciones de Sistema de IA y riesgo en el siguiente sentido:

i) «sistema de IA»: un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales;
ii) «riesgo»: la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio;

El objetivo último que plantea el RIA[2] es "incrementar el bienestar humano", supone un doble enfoque basado en los beneficios y los riesgos que se derivan de la IA de forma que se constituyen como la piedra angular de la norma. Por este motivo, es esencial la visión de las funciones de gestión de riesgos para analizar con profundidad la implementación y efectos de cualquier sistema de IA, y al mismo tiempo perseguir el objetivo de la mejora social que implica la necesaria participación del resto de agentes de nuestro entorno socioeconómico.

La gestión de riesgos de la inteligencia artificial requiere un análisis integral del entorno. Por lo tanto, además del RIA, es fundamental resaltar la aplicación de diversas normas cuyas implicaciones y estudio son cruciales antes de la implementación de cualquier sistema de IA, entre otras destaca las siguientes materias:

2 RIA. Considerando 6

- Protección de datos.
- Ciberseguridad.
- Protección de los consumidores.
- Derechos fundamentales.
- Empleo y protección de los trabajadores.
- Seguridad de los productos.
- Iniciativas legislativas complementarias: Directiva sobre responsabilidad civil extracontractual por IA y la propuesta de Directiva sobre responsabilidad por productos defectuosos.
- Y, estándares como ISO /IEC 42001:2023 que permiten una adecuada implementación y gestión de diversos aspectos puntales de estos sistemas.

El estudio del marco normativo debe completarse con los sistemas de gestión que cada organización por razón de su actividad desarrolle, con el objetivo de dar cabida a los sistemas de IA a los que nos referiremos a lo largo de este documento.

Con el fin de facilitar el contexto, se hace necesario comprender el concepto de los diferentes roles y términos:

(i) Responsable del despliegue[3], se trata de una persona física o jurídica, o autoridad pública, órgano u organismos que bajo su autoridad utiliza un sistema de IA, con la excepción de que se realice en una actividad de carácter personal.
(ii) Distribuidor[4], se refiere a una persona física o jurídica que, siendo diferente del proveedor o del importador, participa en la cadena de suministro y suministra el sistema IA en la UE.
(iii) Proveedor[5], se refiere a una persona física o jurídica, autoridad pública, órgano u organismo que desarrolle o para el que se desarrolla un sistema o un modelo de IA de uso general con el fin de introducirlo en el mercado o ponerlo en servicio con su propio nombre o marca, previo pago o de forma gratuita.

3 RIA. Artículo 3.4
4 RIA. Artículo 3.7
5 RIA. Artículo 3.3

(iv) Proveedor posterior[6] se hace referencia al que provee un sistema de IA o de IA de uso general que integra un modelo de IA con independencia de que éste lo facilite este proveedor y esté integrado de forma vertical o lo facilite un tercero sobre una base contractual.

(v) Representante autorizado[7], se trata de una persona física o jurídica ubicada en la UE que hubiera recibido y aceptado el mandato escrito de un proveedor de un sistema o modelo de IA de uso general para que en su nombre, de cumplimiento a las obligaciones, las disposiciones y procedimientos del RIA.

(vi) Importador[8], se trata de la persona física o jurídica ubicada o establecida en la UE que introduzca en el mercado un sistema de IA con el nombre o la marca de un tercero que esté ubicado o establecido fuera de la UE.

(vii) Operador[9], se refiere al proveedor, fabricante del producto, responsable del despliegue, representante autorizado, al importador o al distribuidor.

(viii) Modelo IA de uso general[10], se trata del modelo que es entrenado con una gran volumen de datos con autosupervisión a gran escala, capaz de realizar diferentes tareas y que puede integrarse en diferentes sistemas o aplicaciones posteriores, excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su introducción en el mercado.

(ix) Sistema[11], se refiere al sistema basado en una máquina diseñado para operar con diferentes niveles de autonomía y que tras el despliegue puede adaptarse. Para objetivos explícitos o implícitos, en entornos físicos o virtuales, infiere de los datos de entrada para generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones.

(x) Riesgo[12], se trata de la probabilidad de que se produzca un perjuicio y el impacto que el perjuicio puede generar.

6 RIA. Artículo 3.68

7 RIA. Artículo 3.5

8 RIA. Artículo 3.6

9 RIA. Artículo 3.8

10 RIA. Artículo 3.63

11 RIA. Artículo 3.1

12 RIA. Artículo 3.2

(xi) Riesgo sistémico[13], hace referencia al riesgo propio de las capacidades de gran impacto de los modelos de IA de uso general, y que debido a su alcance o impacto negativo o potencialmente previsible en la salud pública, seguridad, seguridad pública, los derechos fundamentales o la sociedad, puede propagarse a gran escala a lo largo de toda la cadena de valor.

Después de lo expuesto es necesario matizar que en la utilización de los modelos de IA no aplicarán los roles del responsable de despliegue, importador, distribuidor o fabricante del producto[14].

4. ENFOQUE BASADO EN EL RIESGO DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL (RIA)

El enfoque basado en riesgos busca una comprensión profunda de los riesgos específicos y sus implicaciones, por lo que entre otras cuestiones se abordará: i) la definición de los casos de uso; ii) el análisis de riesgos y su metodología; iii) la evaluación de impacto de derechos fundamentales y protección de datos personales; y iv) el mecanismo de la evaluación de la conformidad que tiene su origen en la normativa europea sobre seguridad de los productos.

Resulta necesario hacer una referencia a la taxonomía de los diferentes niveles de riesgo propuesta por el RIA, así como de las obligaciones y requisitos impuestos a cada categoría. De esta forma podremos definir cuándo deberemos realizar un análisis de riesgos, una evaluación de impacto y, en su caso, realizar la evaluación de conformidad.

La normativa del RIA regula los sistemas de IA con un enfoque basado en el riesgo, es decir, regula las consecuencias de la utilización de un sistema basado en tecnología IA, y utiliza la siguiente clasificación[15]:

13 RIA. Artículo 3.65

14 ISMS FORUM. Handbook IA. Guía para comprender mejor el nuevo Reglamento de Inteligencia Artificial RIA.

15 European Commission. Harmonised Standards for the European AI Act

(i) Sistemas de IA de uso prohibido.
(ii) Sistemas de IA de riesgo alto.
(iii) Sistemas de IA de riesgo limitado.
(iv) Sistemas de IA de riesgo bajo o nulo.

Además de la clasificación anterior, hay que añadir sistemas de IA de uso general con riesgo sistémico[16].

Los sistemas de IA de uso prohibido prohíben determinados casos de uso salvo algunas excepciones. En los sistemas de IA de riesgo alto, se establece una taxonomía en función de los casos de uso condicionados al cumplimiento de varios requisitos y obligaciones como llevar a cabo una evaluación de impacto. Por su parte, los sistemas de riesgo bajo o nulo se vincularán al cumplimiento voluntario de códigos de conducta.

En cuanto a las categorías de sistemas de IA de riesgo limitado definida por la Comisión europea se refiere a *"(…) los riesgos asociados con la falta de transparencia en el uso de la IA (…)"*. En este sentido el RIA establece obligaciones de transparencia[17] y de información, todo ello con el objetivo de garantizar el deber de información, que su contenido sea identificable y la posibilidad de identificar el contenido generado de forma artificial.

El RIA exceptúa su aplicación a los sistemas de IA que se utilicen para fines militares, de defensa o de seguridad nacional, ni a las autoridades públicas de países terceros, ni a organizaciones internacionales en el marco de la cooperación policial o judicial con la UE o sus estados miembros.

Conviene matizar que aquellos productos que impliquen la utilización de sistemas de IA de alto riesgo sobre productos sujetos al sistema de armonización de la UE[18] quedarán limitados en cuanto a la aplicación del RIA y se someterán a la regulación específica que aplica para su producción o fabricación.

16 RIA. Artículo 3.65

17 Remisión al artículo 50 del RIA en el que se determinan las obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA

18 RIA. Anexo I. Sección B

4.1 Sistemas de IA prohibidos

Se enumeran las prácticas de IA prohibidas[19] y que no interfiere con las prácticas prohibidas derivadas de la aplicación de otras normativas del Derecho de la Unión Europea como, por ejemplo, la aplicable a protección de datos personales, de protección de los consumidores, ciberseguridad, etcétera. Las prácticas prohibidas reguladas por el RIA son concebidas como *númerus clausus*.

(i) El uso de técnicas subliminales, manipuladoras o engañosas para distorsionar el comportamiento de una persona o un grupo de personas. El RIA desarrolla[20] el concepto de técnicas de manipulación del comportamiento humano con el riesgo de poder producir efectos en la salud física o mental o en los intereses financieros. Por lo tanto, se focaliza en los efectos que el uso de estas prácticas que pueden tener en la decisión de los individuos.
(ii) Además de lo anterior, el RIA presta especial atención a la utilización de estos sistemas de IA y los perjuicios que pueda ocasionar cuando se trata de personas vulnerables o de un determinado colectivo de personas, o de una situación social o económica, o por razón de su edad o discapacidad.
(iii) Puntuación ciudadana (*social scoring*).
La norma específica un uso prohibido de la IA[21] cuando se utiliza para evaluar o clasificar a personas físicas o colectivos basándose en su comportamiento social o características personales durante un periodo de tiempo determinado, siempre que dicha evaluación resulte perjudicial o desfavorable. En consecuencia, no todos los usos de la IA en evaluaciones de las personas o clasificaciones están prohibidos, solo aquellos que generen consecuencias negativas, como un trato desfavorable en contextos diferentes a donde se recopilaron los datos, o un trato desproporcionado o injustificado según el comportamiento social de la persona o colectivo. El objetivo es evitar sistemas de "crédito social", como los que existen en terceros países. Sin embargo, la normativa no afecta a las evaluaciones de las

19 RIA. Artículo 5

20 RIA. Considerando 29

21 RIA. Artículo 5.1 c)

personas físicas que se realicen con un fin específico que cumpla con la normativa de la Unión Europea o nacional.

(iv) Sistemas de IA para evaluar o predecir el riesgo de que una persona física cometa un delito, basándose únicamente en su perfil o sus rasgos de personalidad.

En el desarrollo de esta prohibición[22] podemos deducir que aplica (i) a los sistemas de IA diseñados específicamente para estas finalidades, y a aquellos que permiten alcanzar los mismos resultados; (ii) a los que se refiera a delitos, excluyendo infracciones administrativas; (iii) a los sistemas de IA que elaboran perfiles o rasgos de personalidad cuando este es el único fundamento del sistema, por lo tanto, si la elaboración de perfiles no es el único propósito, el sistema de IA no estaría prohibido.

El RIA regula una excepción refiriéndose a que estos sistemas de IA puedan utilizarse en el caso de personas sospechosas, siempre que la sospecha se base en hechos objetivos y verificables, y estén directamente relacionados con una actividad delictiva. Es decir, la prohibición no se aplica cuando la IA apoya una evaluación humana en el contexto de delitos basados en pruebas objetivas.

(v) Sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión.

Es necesario acudir a la definición de datos biométricos[23] que se conecta con (v.i) el RGPD y de la LOPDGDD[24]; (v.ii) Directiva (UE) 2016/680 del Parlamento Europeo y Del Consejo de 27 de abril de 2016, y de la trasposición en la Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales[25].

Para tratar datos biométricos deberá contarse con una base que legitime su tratamiento. Para poder interpretar el concepto de "datos biométricos" el RIA los clasifica[26] (i) autenticación; (ii) identificación o

22 RIA. Considerando 42

23 RIA. Artículo 3.24

24 RGPD. Artículo 9. LOPDGDD. Artículo 9

25 Ley Orgánica 7/2021. Artículo 13

26 RIA. Considerando 14

categorización; y, (iii) el reconocimiento de emociones de personas físicas.

El RIA no prohíbe todo tipo de reconocimiento facial, sino únicamente aquel que ocurra bajo ciertas condiciones específicas descritas en su normativa. Se busca proteger[27] a las personas contra la vigilancia masiva, que implica la recopilación no selectiva de imágenes. La prohibición se centra en la creación o ampliación de bases de datos con imágenes obtenidas de manera no selectiva, como a través de internet o circuitos cerrados de televisión.

(vi) Sistemas de IA de categorización biométrica que clasifican individualmente a las personas físicas sobre la base de sus datos biométricos con el objetivo de deducir o inferir datos sensibles.

Este supuesto prohíbe los sistemas de IA[28] que reconocen emociones por su potencial para generar resultados discriminatorios, y vulnerar los derechos y libertades de las personas, especialmente en ciertos contextos. Estos sistemas se definen[29] como *"(…) aquellos que infieren emociones o intenciones a partir de sus datos biométricos (…)"*.

El RIA prohíbe[30] específicamente la introducción en el mercado o el uso de sistemas de IA para inferir emociones en lugares de trabajo y centros educativos, debido a su carácter intrusivo y los efectos inadmisibles que puede tener en estos entornos. Sin embargo, existe una excepción a esta prohibición cuando el sistema de IA se utiliza por razones médicas o de seguridad.

(vii) Sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos, excepto cuando su uso esté destinado a fines médicos o de seguridad. Para comprender la extensión de esta prohibición[31] es necesario acudir a la definición realizada sobre la categorización biométrica[32].

El RIA define la categorización biométrica como un sistema que clasifica a las personas en categorías específicas basadas en sus datos biométricos, salvo que sea un uso accesorio y estrictamente necesa-

27 RIA. Considerando 43

28 RIA. Considerando 44

29 RIA. Artículo 3.39

30 RIA. Artículo 5.1 f)

31 RIA. Artículo 3.40

32 RIA. Artículo 5.1. Considerandos 16 y 30

rio por razones técnicas. Estas categorías pueden incluir características como el género, la edad, el color del pelo, los rasgos conductuales o las creencias políticas o religiosas.
La normativa prohíbe el uso de sistemas de IA que tengan como fin clasificar a personas basándose en datos biométricos para inferir características sensibles, como la raza, orientación sexual o afiliación política, consideradas de especial protección por otras normativas como la regulación aplicable a la protección de datos personales. No obstante, se permite el uso de categorización biométrica en ciertos contextos legales, como el cumplimiento del derecho, y para etiquetar o filtrar datos biométricos adquiridos de manera legal, como imágenes, que no infringen la normativa. Por ejemplo, la clasificación de imágenes basada en el color del pelo o los ojos podría permitirse en determinados supuestos.

(viii) Sistemas de identificación biométrica remota en tiempo real en espacios de acceso público con fines de garantía del cumplimiento de la ley, con ciertas excepciones sujetas a condiciones y garantías.
El RIA explica el concepto de "identificación biométrica"[33] y ahonda en el concepto y requisitos para la identificación biométrica[34], como el modo de caminar, la postura, la frecuencia cardíaca, la presión arterial, pulsaciones de tecla, movimiento ocular, características conductuales, etcétera. El RIA exceptúa de este supuesto los sistemas de IA con finalidades de verificación biométrica, que comprende la autenticación al indicar que su finalidad es la de confirmar que una persona es quién dice ser, con la finalidad de acceso a un servicio, acceso de seguridad o desbloqueo de un dispositivo.
En cuanto al uso de estos sistemas de IA para la identificación biométrica remota en tiempo real en espacios públicos con fines de cumplimiento normativo, puede generar desconfianza, sensación de control con el consecuente impacto en los derechos y libertades de las personas físicas como, por ejemplo, producir resultados sesgados que lleven a conductas discriminatorias, en función de su edad, etnia, raza o género.
Se permite su uso en situaciones excepciones en las que su uso sea estrictamente necesario, como la búsqueda selectiva de víctimas

33 RIA. Artículo 3.35

34 RIA. Considerandos 15 y 32

de delitos graves, la prevención de amenaza real actual o previsible de la seguridad física, o la identificación de sospechosos de delitos graves. En estos casos se articularán las garantías necesarias como llevarlo a cabo a través de una autorización judicial previa además de realizar la evaluación de impacto sobre protección de datos, y la notificación a autoridades competentes.

4.2 Sistemas de IA de riesgo alto

La clasificación realizada por el RIA atiende al criterio del impacto que puede tener en los derechos fundamentales, así como en la salud y en la seguridad. Estos sistemas están permitidos siempre que cumplan con una serie de requisitos y siempre que el riesgo pase la evaluación de conformidad o, la evaluación de impacto.

Se clasificarán de alto riesgo si cumplen dos condiciones de forma cumulativa:

i) Que el sistema de IA sean componentes de un producto o un producto incluido en la legislación de la Unión Europea[35], o sea un componente de seguridad de estos productos. Como, por ejemplo, un dispositivo médico, equipos radioeléctricos, productos sanitarios para diagnóstico, etcétera.
ii) Que se trate de uno de los sistemas de IA que deba someterse a una evaluación de conformidad[36].

En las evaluaciones de conformidad[37] se analiza si el sistema de IA de alto riesgo cumple con la normativa, evalúa el sistema de gestión de calidad y documentación técnica e indica dos procedimientos alternativos de evaluación dependiendo de la normativa que utilizara. Deberá tenerse en cuenta que, si un sistema de IA se modifica de forma sustancial, deberá someterse a una nueva evaluación. Hay que añadir que el cumplimiento con los requisitos de evaluación de conformidad, recibirán el marcado CE físico o digital, dependiendo del tipo de producto y solo podrán comercializarse en la UE.

35 RIA. Anexo I

36 RIA. Anexo I

37 RIA. Artículo 43

Además de lo anterior, define a los sistemas de IA de riesgo alto[38] como aquellos que no estén vinculados a productos que no sean componentes de seguridad de productos, o que son productos en sí mismos, pero pueden causar perjuicios para la salud y la seguridad, o a los derechos fundamentales de las personas físicas sobre la base de la probabilidad y el impacto si llegara a producirse el perjuicio[39].

Para que un sistema de IA no se considere de alto riesgo a pesar de estar incluido en el anexo III del RIA, deberá determinarse su influencia en la toma de decisiones, y si cumple alguna o varias de las siguientes condiciones[40]:

(i) El sistema lleva a cabo una tarea de procedimiento limitada[41] consiste en "(…) *realizar una tarea de procedimiento delimitada, como un sistema de IA que transforme datos no estructurados en datos estructurados, un sistema de IA que clasifique en categorías los documentos recibidos o un sistema de IA que se utilice para detectar duplicados entre un gran número de aplicaciones* (…)".

(ii) La tarea esté destinada a la mejora actividades humanas previas llevadas a cabo por un ser humano y el sistema de IA complementa esta actividad humana y se ejemplifica[42] "(…) *mejorar el lenguaje utilizado en documentos ya redactados* (…)".

(iii) Esté destinado a detectar patrones de decisión o desviación sin reemplazar la intervención humana y especifica[43] "(…) *por ejemplo* (…) *utilizarse para comprobar a posteriori si un profesor puede haberse desviado de su patrón de calificación determinado, a fin de llamar la atención sobre posibles incoherencias o anomalías* (…)" ; o, (iv) lleva a cabo tareas preparatorias y el se refiere al ejemplo[44] "(…) *soluciones inteligentes para la gestión de archivos, lo que incluye funciones diversas tales como la indexación, la búsqueda, el tratamiento de texto y del habla o la vinculación de datos a otras fuentes de datos, o bien los sistemas de IA utilizados para la traducción de los documentos iniciales* (…)".

(iv) Para asegurar la trazabilidad y la transparencia en el uso de sistemas de IA, los proveedores que determinen que un sistema de IA no es de alto riesgo, basándose en las condiciones anteriores, deben preparar una evaluación documentada antes de su comercialización o puesta en servicio. Esta documentación debe estar disponible para las autoridades nacionales competentes si es solicitada. Además, los proveedores están obligados a registrar el sistema en una base de datos de la UE creada al amparo del RIA, lo que refuerza el control y seguimiento de estos sistemas.

38 RIA. Anexo III

39 RIA. Anexo I

40 Revista de Privacidad y Derecho Digital. AÑO IX • MAYO-AGOSTO 2024 • NÚMERO 34

41 RIA. Considerando 53

42 RIA. Considerando 53

43 RIA. Considerando 53

44 RIA. Considerando 53

Los sistemas de IA clasificados como alto riesgo tienen como característica su avanzada capacidad técnica y el impacto que genera en la esfera personal y social. Por este motivo se establecen siete (7) requisitos de obligado cumplimiento[45]:

(i) Establecer, implantar, documentar y mantener un sistema de gestión de riesgos en relación con los sistemas de IA de alto riesgo.

(ii) El entrenamiento de modelos de IA se desarrollará a partir de conjuntos de datos de entrenamiento, validación y prueba que cumplan los criterios de calidad establecidos en el RIA[46] "(…) *Los conjuntos de datos de entrenamiento, validación y prueba se someterán a prácticas de gobernanza y gestión de datos adecuadas para la finalidad prevista del sistema de IA de alto riesgo* (…)".

(iii) La documentación técnica del sistema de IA deberá de elaborarse antes de introducirlo en el mercado o su puesta en servicio, además de la obligación de mantenerla actualizada.

(iv) Permitirán el registro automático de archivos de registro durante el ciclo de vida del sistema.

(v) Se diseñarán y desarrollarán de forma que se garantice el nivel de transparencia necesario que permita a los responsables del despliegue interpretar y utilizar los resultados de salida.

(vi) Se diseñarán y desarrollarán para que puedan ser supervisados con intervención humana de forma proporcional al riesgo, al nivel de au-

45 RIA. Sección 2

46 RIA. Artículo 10.2

tonomía y al caso de uso, a lo largo del período de utilización con el fin de prevenir o reducir los riesgos para la salud, la seguridad o los derechos fundamentales.

(vii) Se diseñarán y desarrollarán con el objetivo de tener un nivel adecuado de precisión, solidez y ciberseguridad de forma que se garantice durante el ciclo de vida.

4.2.1 Obligaciones de los proveedores

Además de los requisitos que el RIA establece para los sistemas de IA de alto riesgo, establece una serie de obligaciones para los proveedores de los sistemas de IA de alto riesgo.

Los proveedores de estos sistemas deben asegurarse del cumplimiento de todos los requisitos normativos[47] aquí referenciados, e incluirán el establecimiento de:

(i) Sistema de gestión de registro, se impone la obligación específica de conservar los archivos de registro que se generen de forma automática a lo largo de todo el ciclo de vida del sistema y que siempre estén bajo su control.
(ii) De un sistema de gestión de la calidad.
(iii) El cumplimiento de los procedimientos de evaluación de la conformidad, antes de su introducción en el mercado.
(iv) Una evaluación de impacto[48] relativa a los derechos fundamentales para los sistemas de IA[49].

47 RIA. Sección 2

48 RIA. Artículo 27

49 Las organizaciones deben documentar las evaluaciones de riesgos de IA para demostrar su responsabilidad. A un alto nivel, la documentación debe reflejar los riesgos identificados durante la evaluación, las medidas adoptadas para mitigar los riesgos y si, en conjunto, las medidas de mitigación son adecuadas y suficientes para hacer frente a los riesgos para que la organización continúe con la actividad de tratamiento de la IA. Tales evaluaciones de riesgos y evaluaciones de impacto, ya existen y se exigen en normativas basadas en aproximación al riesgo como el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). Comisión Europea-Preguntas y respuestas. Bruselas, 1 de agosto de 2024: "(…) *Los operadores que suministren sistemas de inteligencia artificial de alto riesgo que realicen evaluaciones de la solvencia crediticia o evaluaciones de precios y de riesgos en los seguros de salud y de vida, deberán llevar a cabo una evaluación de las repercusiones en los derechos fundamentales y notificar los resultados a la autoridad nacional. En la práctica, muchos usuarios también tendrán que llevar a cabo una evaluación de impacto relativa a la protección de datos. Para evitar solapamientos sustanciales en tales casos, la evaluación de impacto sobre los derechos fundamentales se llevará a cabo junto con esa evaluación de impacto relativa a la protección de datos* (…)".

(v) Procedimientos de modificación del sistema de IA de alto riesgo.
(vi) Procedimientos y técnicas de diseño, desarrollo, control de calidad y garantía de calidad del sistema de IA de alto riesgo.
(vii) Procedimientos de prueba y validación a lo largo del ciclo de vida de desarrollo del sistema de IA de alto riesgo.
(viii) Sistemas y procedimientos de gestión de datos. Estos sistemas que impliquen el entrenamiento de modelos de IA se desarrollarán a partir de un conjunto de datos de entrenamiento, validación y prueba deberán cumplir con los criterios de calidad del artículo 10 apartados del 2 al 5 del RIA.
Como cuestión práctica nos preguntaremos[50] por la exactitud, la disponibilidad de sus registros, la exhaustividad, si se ajustan a la normativa, si son coherentes y se ajustan a los patrones, si son redundantes sus patrones o atributos, si son exactas las relaciones entre los datos, y si éstos están actualizados y disponibles cuando se necesitan.
(ix) Sistema de gestión de riesgos y un sistema de seguimiento de postcomercialización. El RIA impone la obligación específica al proveedor que considere que no cumple con el RIA, establecer medidas correctoras necesarias para cumplir, para retirarlo del mercado, desactivarlo o recuperarlo, además de informar a los distribuidores de estos sistemas y, si procediera a los responsables del despliegue, al representante autorizado y a los importadores.
(x) Procesos de comunicación y de notificación de incidentes graves. El RIA impone de forma específica estas obligaciones a los proveedores: (ix.i) previa solicitud, deberán de cooperar con las autoridades competentes para demostrar la conformidad del sistema de IA de alto riesgo, así como el acceso a los archivos de registro que se generen de forma automática; (ix.ii) obligaciones establecidas en

50 ISACA, Applied Data Management for Privacy, Security and Digital Trust, 2023

el artículo 13 del RIA; (ix.iii) notificación de incidentes graves a las autoridades competentes donde se hubiera producido el incidente después de que el proveedor hubiera verificado el nexo causal entre el sistema IA y el incidente o la probabilidad razonable de que este exista y, en todo caso, quince (15) días después de que el proveedor o el responsable del despliegue tuvieran conocimiento.

(xi) Conservación de la documentación[51]. Durante un periodo de diez (10) años desde la introducción en el mercado o la puesta en servicio, mantendrá a disposición de las autoridades nacionales la documentación técnica, del sistema de gestión de calidad, la relativa a los cambios aprobados por organismos notificados, las decisiones y otros documentos expedidos por organismos notificados y la declaración UE de conformidad.

Como se ha señalado, los proveedores desempeñan un papel clave y están obligados a cumplir con una serie de obligaciones. Sin embargo, en ciertas situaciones, los distribuidores, importadores, responsables o terceros de estos sistemas también podrán ser considerados proveedores, situación que implicará cumplir con las obligaciones establecidas para los proveedores del artículo 16 del RIA[52].

Siguiendo la regulación[53] se considera que los responsables deben adoptar las medidas técnicas y organizativas adecuadas para garantizar que el uso de estos sistemas, siguen las instrucciones, aseguran la supervisión humana cualificada, y verifican que los datos utilizados sean pertinentes y representativos para el propósito de estos sistemas. La supervisión humana es necesaria teniendo en cuenta el potencial impacto que podrían causar estos sistemas, y se exige con la finalidad de reducir el riesgo para la salud, la seguridad o los derechos fundamentales que podría derivarse de su uso como, por ejemplo, advertir sobre aquellos que faciliten información o impliquen recomendaciones para la toma de decisiones; ayudar con la interpretación que realiza el sistema; permitir ignorar; anular o revertir los resultados, etcétera[54]. Con carácter adicional se indica que antes de utilizar estos

51 RIA. Artículo 18

52 RIA. Artículo 25

53 RIA. Artículo 26

54 ISACA. Understanding the EU AI Act: Requirements and Next Steps

sistemas en el entorno laboral, los responsables del despliegue que sean empleadores deberán informar a los representantes de los trabajadores y a los trabajadores afectados de su exposición a la utilización de este sistema.

Además, los responsables del despliegue deberán monitorizar el funcionamiento de estos sistemas sobre la base de las instrucciones de uso, realizar evaluaciones de impacto de los derechos fundamentales e informar a los proveedores. En el caso de las organizaciones financieras que están sujetos a requisitos específicos de gobernanza, la obligación de monitorear se considera cumplida cuando se respeten las normas sobre sistemas, procesos y mecanismos de gobernanza interna acorde al Derecho que aplique a los servicios financieros.

En caso de detección de un incidente grave, deberán informar de inmediato al proveedor y al importador o distribuidor y a la autoridad de vigilancia competente con exclusión de los datos operativos sensible de los responsables del despliegue de sistemas de IA que sean autoridades competentes. La normativa define los incidentes graves[55] como un mal funcionamiento o incidente de un sistema de IA que provoque la muerte o daños a la salud, una perturbación grave e irreparable de infraestructuras críticas, el incumplimiento de obligaciones destinadas a proteger derechos fundamentales o daños graves a la propiedad o al medio ambiente.

Podemos decir que el enfoque basado en el riesgo del RIA es fundamental para garantizar la responsabilidad y seguridad, proporcionando un marco para mitigar los riesgos asociados en el uso de los sistemas de IA de alto riesgo, protegiendo a los ciudadanos, personas físicas y orientando a las organizaciones desde una perspectiva ética.

Anticipación y flexibilidad: claves para enfrentar la rápida evolución de la IA

El RIA es el reflejo de una evolución que busca adaptarse a la rápida transformación tecnológica. Por ejemplo, otorga a la Comisión Europea la facultad de revisar y actualizar el anexo III[56], y puede añadir nuevos sistemas que cumplan con los criterios de riesgo preestablecidos y eliminar aquellos que ya no representen un riesgo significativo. La inclusión de nuevos sistemas

55 RIA. Artículo 3.49

56 RIA. Artículo 7

dependerá de si estos operan en áreas ya cubiertas por el Anexo III y si presentan un nivel de riesgo equivalente o superior en cuanto a la salud, seguridad o derechos fundamentales.

Desde el 2 de agosto de 2026, con carácter anual la Comisión evaluará la necesidad de modificar el anexo III, así como la lista de prácticas prohibidas previstas en el artículo 5[57]. Además, se establece que, desde el 2 de agosto de 2028, y con posterioridad en periodos de cuatro (4) años la Comisión evaluará determinados puntos relacionados con el anexo III, las medidas de transparencia adicionales a los proveedores y responsables del despliegue de determinados sistemas de IA, y la mejora en la eficacia de supervisión y gobernanza.

La inclusión de nuevos sistemas de IA en la lista de alto riesgo tendrá en cuenta si suponen un riesgo para la salud y seguridad o un impacto negativo en los derechos fundamentales.

4.2.2 Requisitos específicos para los sistemas de IA de alto riesgo

Para los sistemas de IA de alto riesgo se deberá diseñar, implementar, documentar, supervisar el sistema de gestión de riesgos relacionado con este sistema a lo largo del ciclo de vida del sistema. Siguiendo la regulación del RIA, deberán incluirse en el sistema de gestión de calidad, el sistema de gestión de riesgos. Todo ello con la finalidad de controlar y mitigar los riesgos desde el inicio hasta el final de vida del sistema de IA.

Se establece una clasificación de los riesgos[58] en función si éstos (i) son conocidos y previsibles; (ii) los que pudieran surgir por una utilización correcta o los derivados de un uso indebido; y, (iii) aquellos que pudieran surgir a partir de los datos postcomercialización.

Los requisitos de postcomercialización[59] se derivan de la necesidad de garantizar un cumplimiento continuo y de esta forma poder identificar necesidades de aplicar medidas preventivas o correctivas. Todo ello teniendo en

57 RIA. Artículo 112

58 RIA. Artículo 9

59 RIA. Artículo 72 y Considerando 155

cuenta cada caso en el que el sistema de IA de alto riesgo interactúa con otros sistemas de IA. Además de las obligaciones de informar sobre la producción de un incidente grave a las autoridades competentes.

Las medidas de gestión de riesgos considerarán aceptables los riesgos residuales que se asocien a cada peligro.

En cuanto a la gestión del riesgo, el RIA[60] facilita las medidas de gestión más adecuadas:

(i) Eliminar o reducir. Se tendrán en cuenta los conocimientos técnicos, la experiencia, la educación y formación que se espera del responsable del despliegue, como el contexto en el que esté prevista la utilización del sistema de alto riesgo.
(ii) Implementar medidas de mitigación y control apropiadas que no puedan eliminarse.
(iii) Facilitar información requerida y, cuando proceda, impartir formación a los responsables del despliegue.

Las pruebas de los sistemas de IA de alto riesgo a la que se someterán las pruebas que puedan determinar las medidas de gestión más adecuadas, se realizarán antes de introducirlas en el mercado o la puesta en servicio y en cualquier momento del proceso de desarrollo.

4.2.3 La gestión del riesgo de sistemas IA de alto riesgo

Además de evaluar los riesgos de cumplimiento normativo, deberán de analizarse y evaluar los riesgos asociados derivados de la utilización de los sistemas IA de alto riego.

Este proceso deberá de ser iterativo teniendo en cuenta la evolución de las tecnologías basadas en IA, además de los casos de uso. Una vez puestos en servicio, los sistemas de IA continúan en continuo aprendizaje y con el fin de reducir o eliminar el riesgo de resultados sesgados deberá diseñarse un proceso de reevaluación de riesgos[61].

60 RIA. Artículo 9.5

61 ISACA understanding the EU AI Act: Requirements and Next Steps

(i) Identificar la tecnología, procesos, colectivos y derechos fundamentales afectados a lo largo de todo su ciclo de vida como la planificación, el desarrollo, la implementación, operaciones y terminación de su vida útil.

(ii) Identificar y definir el alcance del caso de uso. Por ejemplo, si la tecnología basada en IA se valora aplicar a procesos de gestión de siniestros, de reclamaciones, de pagos, para soporte al cliente, para prevenir el fraude, para el proceso de tarificación de los riesgos en el proceso de suscripción, para personalizar los presupuestos y las pólizas acorde a las necesidades del cliente, etcétera.

(iii) Analizar y evaluar los riesgos asociados al sistema IA de alto riesgo[62]. Pueden darse diferentes tipologías de riesgos asociados a los datos personales[63], a la dependencia y automatización, a las fuentes de los datos, al diseño de algoritmos, a la ética, a la seguridad, al cumplimiento normativo del RIA y otras normativas afectadas, al medioambiente, a la escalabilidad y al rendimiento, a la propiedad intelectual, a la esfera social y la laboral, la no discriminación, a la ciberseguridad[64] que puede afectar a los sistemas de IA y/o al uso malicioso[65].

No existe un método único para realizar una evaluación de riesgos y será cada organización la que deberá de definir e implementar la metodología que considere adecuada para la gestión de riesgos de la utilización de sistemas IA[66].

62 Companion Guide On Securing AI Systems. October 2024

63 ISO/IEC 27701 como marco de gestión para la protección de la privacidad y el tratamiento de los datos personales.

64 The NIST Cybersecurity Framework (CSF) 2.0. February 26, 2024.

65 https://owasp.org/www-project-developer-guide/draft/verification/guides/web_security_testing_guide

66 Tal y como se explica en el Sistema de Gestión de Inteligencia Artificial en Tecnologías de la Información de la Organización Internacional de Normalización (ISO/IEC 42001:2023 en 6.1.2(d)(2)), una vez que la organización ha clasificado los riesgos, evaluará la probabilidad y el impacto de la materialización de la amenaza. Utilizando una matriz de riesgo, una organización puede analizar el impacto -vinculado a diferentes niveles de impacto asociados a un valor cuantitativo-, y la probabilidad -vinculado a diferentes niveles de probabilidad asociados a un valor cuantitativo- de daño según su metodología que, por ejemplo, podrá definirse como crítica, moderada o baja. Dependiendo del nivel de riesgo, una organización puede necesitar implementar medidas de mitigación y salvaguardas adicionales.

En esta fase, además del análisis de riesgos[67], implicará, en su caso, una evaluación de impacto en derechos fundamentales y protección de datos personales y que en todo caso deberán de ser documentadas siguiendo la metodología implementada en cada organización. Tal y como se ha indicado en apartados anteriores, con el fin de evitar solapamientos sustanciales en estos casos, la evaluación de impacto sobre los derechos fundamentales se llevará a cabo junto con esa evaluación de impacto relativa a la protección de datos.

(iv) Identificar los riesgos derivados de la cadena de suministro. Como el análisis de proveedores IT, priorización según su nivel de criticidad, incluir a proveedores y/o terceros relevantes para poder dar respuesta y poder recuperarse ante incidentes.

(v) Identificar los controles o las medidas adecuadas a los riesgos identificados[68]. Estos controles o medidas podrán dividirse en bloques diferenciando los (i) controles operativos, como los relacionados con la ciberseguridad, transparencia, supervisión humana, auditorías, sistema de gobernanza adecuado, auditorías de algoritmos, políticas con el objetivo de verificar la calidad de los datos con los que el sistema se ha entrenado, etcétera; (ii) controles documentales, como los relacionados con la documentación técnica, la conservación de registros, etcétera; (iii) controles propios establecidos por cada organización; y, (iv) controles asociados a la formación y concienciación.

(vi) Implementación en la organización y sus procesos. Incluye formación específica a las áreas afectadas.

(vii) Operación y monitorización. Deberá atenderse a la criticidad derivada del análisis de riesgo y a las obligaciones derivadas del tipo de riesgo según el RIA.

Esta fase incluye la supervisión y evaluación de los riesgos emergentes una vez que el sistema de IA se implementa.

(viii) Terminación de su vida útil.

67 Para el caso que no se trate de un sistema de IA de alto riesgo, deberá tenerse en cuenta los diferentes niveles de riesgo establecidos del RIA aquí especificados. Es decir, se tendrá en cuenta la clasificación con las respectivas obligaciones de (i) los casos de uso prohibidos; (ii) los de riesgo limitado; (iii) los modelos de IA de uso general. Además, podrán utilizarse bases de datos que especifiquen los riesgos de la utilización de la IA.

68 Sobre la base de publicación en LinkdIn de Xabier Ribas, acudir para información adicional

4.3 Sistemas de IA de riesgo limitado

En el RIA se especifica que algunos sistemas de IA diseñados para interactuar con personas o crear contenidos, presentan riesgos potenciales de suplantación o engaño[69]. Estos peligros pueden surgir con independencia de si dichos sistemas son clasificados como de alto riesgo.

Por lo tanto, se subraya la necesidad de vigilancia adicional para evitar posibles manipulaciones o confusiones de los riesgos derivados de la capacidad de la IA para imitar comportamientos humanos o crear información que podría influir y engañar a los usuarios.

Para determinados sistemas de IA con independencia de su clasificación de alto riesgo[70], se imponen requisitos específicos de transparencia; por ejemplo, cuando existe un riesgo claro de manipulación (v. gr., mediante el uso de *chatbots*), con el fin que los usuarios sean conscientes de que están interactuando con un sistema de IA.

En este sentido, el RIA[71] establece obligaciones de transparencia para (i) sistemas de alto riesgo; y, (ii) para algunos sistemas de IA aunque no sean de alto riesgo y que serán los denominados de riesgo limitado. Se establecen cuatro (4) supuestos:

(i) Sistemas destinados a interactuar con personas físicas. Esta obligación se impone a los proveedores del sistema IA.
(ii) Sistemas de IA que generen contenido sintético de audio, imagen, video o texto. Los proveedores velarán por que la información de salida del sistema de IA esté en un formato legible y que ha sido generada o manipulada de manera artificial.
(iii) Sistemas de reconocimiento de emociones o de categorización biométrica. Los responsables del sistema deberán informar del funcionamiento del sistema a las personas físicas expuestas a él.

69 RIA. Considerando 132

70 Reglamento (UE) de IA. Un marco jurídico pionero sobre inteligencia artificial. Guía práctica. Julio 2024. Cuatrecasas

71 RIA. Artículo 50

(iv) Sistemas que generen o manipulen imágenes o contenidos de audio o vídeo que constituyan una ultrasuplantación[72], y contenidos que informen sobre asuntos de interés público. En este supuesto, los responsables del despliegue deberán divulgar que el contenido se ha generado o manipulado de manera artificial.

4.4 Sistemas de IA de riesgo bajo o nulo

Nos referimos a todos los demás sistemas de IA que pueden desarrollarse y utilizarse conforme a la legislación vigente sin obligaciones adicionales. El RIA no los regula, pero sí que hace una mención[73], y se decide por la opción que los proveedores de estos sistemas puedan optar de forma voluntaria por aplicar las directrices para una IA ética, y a diseñar e implementar o adherirse a códigos de conducta de forma voluntaria.

4.5. Los modelos de IA de uso general que se integran en los sistemas de IA

El RIA[74] aporta las definiciones que deberán ser tenidas en cuenta para su correcta interpretación. Entre todas las aportadas no se encuentra la definición de IA generativa, pero si nos ofrece un marco de referencia[75], y nos indica que los grandes modelos de IA generativa son un claro ejemplo de sistemas de uso general, ya que tienen la capacidad de generar contenido en diversos formatos como texto, audio, imágenes o video. Esta flexibilidad les permite adaptarse a una amplia variedad de tareas y aplicaciones.

Los modelos de IA generativa aprenden patrones y estructuras a partir de los datos de entrenamiento, lo que les permite generar nuevos datos con características similares, y abarcan una gran variedad de usos potenciales, contemplados o no originalmente por los creadores del sistema.

72 RIA. Artículo 3.60

73 RIA. Considerandos 165 y 166

74 RIA. Artículo 3

75 RIA. Considerando 99

Podemos decir que son aquellos que (i) tienen un grado considerable de generalidad; (ii) son capaces de realizar una gran variedad de tareas; y, (iii) pueden integrarse en diversos sistemas o aplicaciones de IA.

Entre otros ejemplos destaca el ChatGPT de OpenAI que se califica como un sistema de IA de uso general al tratarse de un sistema de IA basado en un modelo de IA de uso general.

4.5.1. Obligaciones:

Deberán cumplirse con una serie de obligaciones:

(i) Documentar el proceso de entrenamiento y sus resultados de evaluación.
(ii) Informar a los proveedores de sistemas de IA que tengan previsto integrar en sus sistemas el modelo de IA de uso general sobre sus características y requisitos legales.
(iii) Establecer una política de cumplimiento de la normativa de la UE sobre derechos de autor y derechos afines, especialmente en lo que respecta a la minería de textos y datos.
(iv) Divulgar públicamente un resumen detallado del contenido utilizado para entrenar el modelo de IA de uso general.

Debido a sus capacidades de gran impacto, se considera que determinados modelos de IA de uso general plantean un riesgo sistémico. Para mitigar estos riesgos, los proveedores deben cumplir requisitos adicionales que se detallan en el siguiente apartado.

4.5.2. Riesgos

Siguiendo la clasificación que venimos analizando podemos encontrarnos con sistemas de IA generativa de con consecuencias muy diversas dependiendo los riesgos que puedan llegar a generar.

En ciertos casos pueden llegar a calificarse como usos prohibidos en la utilización de la IA generativa[76], como la prohibición de uso de sistemas de IA:

76 RIA. Artículo 5

- Que impliquen el uso de técnicas subliminales o manipuladoras en sistemas de inteligencia artificial que puedan influir en el comportamiento de una persona o grupo, alterando de forma significativa su capacidad para tomar decisiones informadas. Estas técnicas pueden llevar a que se tomen decisiones que no se hubieran adoptado de manera consciente, generando potencialmente daños considerables a individuos o colectivos.
- Que aprovechen vulnerabilidades de personas o colectivos debido a factores como la edad, discapacidad o situación socioeconómica. Estos sistemas no deben alterar de manera significativa el comportamiento de los individuos, especialmente si su utilización puede causar daños considerables a las personas físicas o a un colectivo.

Por otra parte, también podemos encontrarnos con sistemas de IA de uso general de riesgo alto remitiéndonos a lo especificado en el apartado de esta guía y que recordamos de forma resumida. Los sistemas de IA se consideran de alto riesgo si están en los ámbitos listados en el Anexo III del RIA y presentan un riesgo significativo para la salud, la seguridad o los derechos fundamentales de las personas, influyendo de manera sustancial en la toma de decisiones[77].

Y, por exclusión, no se clasificarán como de alto riesgo cuando no representen un peligro considerable en estos aspectos, ni afecten de forma relevante las decisiones, excepto cuando el sistema de IA esté diseñado para elaborar perfiles de personas, en cuyo caso siempre será considerado de alto riesgo. Sobre los derechos que pueden ser susceptibles de incurrir en un riesgo relevante, el RIA especifica[78] una serie de derechos fundamentales, incluyendo la dignidad humana, la privacidad, la libertad de expresión y reunión, el derecho a la tutela judicial efectiva, protección a los consumidores, así como derechos laborales y educativos. Se enfatiza la importancia de la no discriminación y la igualdad de género. También se mencionan derechos específicos para menores, según la Convención sobre los Derechos del Niño, que requieren una atención especial en el entorno digital. Por último, se subraya la necesidad de considerar el impacto de la inteligencia artificial en la salud y la seguridad, así como el derecho a un medio ambiente protegido.

77 Revista de Privacidad y Derecho Digital. AÑO IX • MAYO-AGOSTO 2024 • NÚMERO 34
78 RIA. Considerando 48

En consecuencia, deberán de identificarse los riesgos asociados al uso de la IA generativa y si éstos pueden suponer un riego alto clasificado por el RIA, deberán diseñarse los controles y las obligaciones que le apliquen.

4.5.3. Riesgos sistémicos de la IA de uso general

El RIA hace una distinción entre aquellos modelos de IA de uso general[79] que suponen un riesgo sistémico y aquellos que no, disponiendo obligaciones adicionales a los que presentan un riesgo sistémico tal y como hemos indicado.

El riesgo sistémico se define[80] como "(…) *un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general, que tienen unas repercusiones considerables* (…) *debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor* (…)". Por su parte, el RIA define el modelo de IA de uso general[81] como "(…) *uno entrenado con un gran volumen de datos utilizando autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas, independientemente de la manera en que el modelo se introduzca en el mercado, y que puede integrarse en diversos sistemas o aplicaciones posteriores, excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su introducción en el mercado* (…)".

En el concepto de riesgo sistémico[82] podrían entenderse:

(i) Efectos negativos reales o razonablemente previsibles en relación con accidentes graves, perturbaciones de sectores críticos y consecuencias graves para la salud y la seguridad públicas.
(ii) Efectos negativos reales o razonablemente previsibles sobre los procesos democráticos, la seguridad pública y seguridad económica.

79 Revista de Privacidad y Derecho Digital. AÑO IX • MAYO-AGOSTO 2024 • NÚMERO 34
80 RIA. Artículo 3.65
81 RIA. Artículo 3.63
82 RIA. Considerando 110

(iii) La difusión de contenidos ilegales, falsos o discriminatorios.

Se considera que un modelo de IA de uso general presenta riesgo sistémico[83] cuando se cumplen alguna de las siguientes condiciones:

(i) El modelo tiene capacidades de gran impacto, entendiéndose como tal aquellas capacidades de gran impacto evaluadas a partir de herramientas que igualan o exceden las capacidades registradas en los modelos de IA que ofrecen indicadores y parámetros de referencia. En este sentido, el RIA presume que existen capacidades con riesgo sistémico cuando la cantidad de cálculo utilizada para su entrenamiento sea mayor que 10^{25}.
En estos casos, el proveedor del modelo deberá notificarlo a la Comisión Europea sin dilación y, en todo caso, en el plazo de dos (2) semanas desde el conocimiento del requisito o desde que se tenga conocimiento de su cumplimiento, aunque durante este periodo podrá presentar argumentos que podrán ser estimados o desestimados para demostrar que el modelo no presenta un riesgo sistémico y, no clasificarlo como modelo de IA de uso general con riesgo sistémico.
(ii) Cuando la Comisión Europea lo determine de oficio o tras recibir una alerta cualificada por expertos de que un modelo de IA de uso general, que se basarán en la capacidad o en las consecuencias similares al establecido en modelos de capacidades de gran impacto, teniendo en cuenta los criterios del anexo XIII del RIA, como el número de parámetros del modelo, la calidad o el tamaño del conjunto de datos, el número de usuarios finales registrados, los parámetros de referencia y las evaluaciones de las capacidades del modelo, si sus repercusiones para el mercado interior son importantes debido a su alcance, que se dará por supuesto cuando se pusiera a disposición de diez mil (10.000) usuarios profesionales registrados en la UE, etcétera.
Hay que tener en cuenta que el RIA[84], establece que, tras la designación de un modelo de IA de uso general como de riesgo sistémico, los proveedores deben esperar al menos seis (6) meses antes de solicitar una reevaluación. Si la Comisión, tras la reevaluación, decide

83 RIA. Artículo 51.1
84 RIA. Artículo 52.5

mantener esa designación, los proveedores deberán esperar otros seis (6) meses antes de poder solicitar una nueva revisión.

4.5.4. Obligaciones de los proveedores de modelos de IA de uso general

La normativa[85] indica que las empresas que desarrollan modelos de IA de uso general deben mantener registros detallados sobre su desarrollo y pruebas, compartiendo esta información con otras empresas interesadas, sin comprometer su propiedad intelectual. Lo anterior no aplica a los modelos de código abierto, salvo que representen riesgos sistémicos. Además, las empresas deben colaborar con la Comisión Europea y autoridades nacionales, y pueden utilizar códigos de prácticas aprobados para demostrar cumplimiento normativo hasta que se establezcan normas definitivas, las cuales podrán actualizarse conforme avance la tecnología.

Todos aquellos proveedores de modelos de IA de uso general -tengan o no riesgo sistémico- deberán cumplir las siguientes obligaciones:

4.5.4.1. Obligaciones de transparencia de proveedores y responsables

Tal y como hemos indicado en apartados anteriores, el RIA[86] determina que los proveedores y responsables del despliegue se les impondrán obligaciones adicionales y específicas en cuanto a la información que deben proporcionar[87].

Entre los diferentes casos de uso, se refiere a los sistemas IA destinados a interactuar directamente con personas físicas; (ii) los proveedores de sistemas IA -con inclusión de los sistemas IA de uso general- que generen contenido sintético de audio, ,imagen, vídeo o texto; (iii) Los responsables del despliegue de un sistema de reconocimiento de emociones o de un sistema de categorización biométrica; (iv) los responsables del despliegue de un sistema de IA que genere o manipule imágenes o contenidos de audio o vídeo que

85 RIA. Artículo 53

86 RIA. Artículo 50

87 La obligación de transparencia impuesta en los cuatro primeros puntos se entenderá sin perjuicio de otras obligaciones de transparencia reguladas en el capítulo III y de las impuestas por el Derecho nacional o de la Unión para los responsables del despliegue de sistemas de IA.

constituyan una ultrasuplantación[88]; y, (v) los responsables del despliegue de un sistema de IA que genere o manipule texto que se publique con el fin de informar al público sobre asuntos de interés público.

En los tres primeros supuestos el RIA establece la excepción de informar cuando "(…) *Esta obligación no se aplicará a los sistemas de IA autorizados por ley para detectar, prevenir, investigar o enjuiciar delitos* (…)".

4.5.4.2. Obligaciones adicionales

Elaborar y mantener actualizada:

(i) Documentación técnica del modelo -con al menos la información detallada en el anexo XI del RIA- para poder facilitarla a la Oficina de IA y/o a las autoridades competentes.
(ii) Información para aquellos proveedores de sistemas de IA que quieran integrar el modelo en sus sistemas, que contenga al menos el contenido del anexo XII del RIA.
(iii) Establecer directrices para cumplir la legislación aplicable en materia de derechos de autor. En particular, deberán respetar las reservas de derechos.
(iv) Poner a disposición del público un resumen del contenido utilizado para el entrenamiento del modelo IA.
(v) Cooperar con la Comisión y las autoridades competentes.

Además de las obligaciones mencionadas para todos los proveedores de modelos de IA de uso general, aquellos proveedores de modelos que presenten un riesgo sistémico deberán cumplir:

(i) Presentar y elaborar la documentación técnica del modelo incluyendo el proceso de formación, ensayo y los resultados de su evaluación, con el mínimo de información detallada en el anexo XI, que se proporcione a la Oficina de IA y a las autoridades nacionales competentes si lo solicitan la siguiente información adicional.
(ii) Los proveedores de modelos de IA de propósito general deberán proporcionar información y documentación actualizada a aquellos

88 RIA. Artículo 3.60

que deseen integrar estos modelos en sus sistemas de IA. Esta información debe adherirse a la normativa de protección de propiedad intelectual y secretos comerciales, al tiempo que permita a los proveedores entender de manera clara las capacidades y limitaciones del modelo, además de cumplir con las obligaciones establecidas por el RIA. Además, la documentación debe incluir, al menos, los elementos especificados en el Anexo XII Establecer una política para garantizar el cumplimiento de la legislación de la Unión Europea en materia de derechos de autor y derechos afines. Esta política debe incluir mecanismos para identificar y respetar las reservas de derechos expresadas según el artículo 4, apartado 3, de la Directiva (UE) 2019/790.

(iii) Crear y poner a disposición un resumen detallado del contenido utilizado para entrenar el modelo de IA de propósito general, siguiendo una plantilla proporcionada por la Oficina de IA, con el fin de garantizar la transparencia en el uso de datos para el entrenamiento del modelo.

(iv) Deber de cooperación con la Comisión y las autoridades nacionales competentes.

(v) Los proveedores de modelos de IA de propósito general pueden usar códigos de prácticas[89], para demostrar que cumplen con sus obligaciones, hasta que se establezcan normas armonizadas. Cumplir con estas normas ofrece una presunción de conformidad. Si los proveedores no siguen un código de buenas prácticas o no cumplen con las normas armonizadas europeas, deberán presentar alternativas adecuadas que serán evaluadas por la Comisión. Códigos de buenas prácticas para proveedores de modelos de IA de uso general.

(vi) Los proveedores de modelos de IA de propósito general[90] pueden utilizar códigos de prácticas, con el objetivo[91] de demostrar el cumplimiento de sus obligaciones hasta que se publique una norma armonizada que otorgará presunción de conformidad. Sin embargo, si los proveedores no se adhieren a un código de buenas prácticas aprobado o no cumplen con una norma armonizada, deberán presentar métodos alternativos adecuados que serán evaluados por la Comisión.

89 RIA. Artículo 56

90 RIA. Artículo 53.4

91 RIA. Artículo 56

La adhesión a códigos de buenas prácticas puede facilitar a los proveedores de estos modelos el cumplimiento con las obligaciones que establece el RIA. Además de lo anterior, deberán establecer un nivel adecuado de protección de ciberseguridad[92] para el modelo de uso general con riesgo sistémico.

RIA establece que la Oficina de IA de la Comisión debe apoyar la creación, revisión y modificación de códigos de conducta integrando diversos puntos de vista.

5. LA GOBERNANZA EN EL DESARROLLO Y UTILIZACIÓN DE LA INTELIGENCIA ARTIFICIAL

Desde el grupo de trabajo Legal Risk de AGERS, tras diversas reuniones con expertos en la materia y los datos que ha aportado la encuesta realizada, se ha concluido que en la actualidad todavía es muy limitado el número de organizaciones (empresas, instituciones) que utilizan sistemas de IA y de IA Generativa; aunque va aumentando de forma constante, aquellas que se plantean usar estas nuevas tecnologías para mejorar sus procesos.

En ambos casos, la gobernanza es clave para la utilización provechosa de la IA en la actividad de las organizaciones, lo que puede abordarse partiendo de dos cuestiones. En primer lugar, si en este estado inicial en el que se encuentra la IA, se está utilizando en las organizaciones y, en segundo lugar, cual es el proceso de implementación de la IA y los órganos de los que dependería el desarrollo y uso de la misma.

Respecto a la primera cuestión, el resultado de la encuesta realizada y de las reuniones mantenidas con profesionales y expertos, muestra el limitado uso actual de la IA en las organizaciones, si bien se contempla su introducción progresiva en algunos procesos de negocio. Se constata que las grandes corporaciones desarrollan y utilizan IA en varios aspectos de su actividad

92 Se recomienda apoyarse en "El Marco de Seguridad Cibernética (CSF) 2.0 del NIST, 26 febrero, 2024" que tiene como objetivo colaborar con las organizaciones con independencia del sector para mejorar las capacidades preventivas y defensivas ante los incidentes de seguridad, además de facilitar herramientas para gestionarlo con éxito

organizativa y predictiva[93]. En las PYMEs, en cambio, es muy escaso el uso de esta nueva tecnología, si bien se produce la siguiente paradoja: aunque la empresa como tal no emplea ni forma a sus trabajadores en IA, algunos de ellos la utilizan personalmente en el trabajo para experimentar y encontrar herramientas que faciliten y mejoren los resultados de su trabajo. A veces la dirección de la empresa no es consciente de ello hasta que se produce algún incidente de protección de datos, confidencialidad, inexactitud de resultados, etc., que pueden generar un incremento del riesgo en la gestión de la empresa y una responsabilidad para la organización por actuación negligente.

Del análisis de los resultados de la encuesta se desprende que las organizaciones no están aplicando la IA de forma coordinada y organizada. Es probable que exista un grado de desconexión entre la IA utilizada y la pirámide de gobierno, en el sentido de que cuanto más altos son los estamentos en la organización, más difícil es una utilización real.

En segundo lugar, en cuanto al proceso y órganos que vigilen y supervisen la IA, en los encuentros mantenidos con directivos de empresas e instituciones, se debate la conveniencia de crear un "Comité de IA". Algunas empresas señalan que ya lo tienen constituido y otras han arrancado el desarrollo y utilización de la IA sin ese Comité. La conclusión es que parece conveniente disponer de un equipo de especialistas que controle el desarrollo y uso de la IA.

Como consecuencia de lo anterior, surgen las siguientes preguntas a las que trataremos de dar respuesta:

¿De qué órganos debe depender la aprobación y desarrollo de la IA en una organización?
¿Es oportuno la creación de un Comité de IA, y que composición debería tener?
¿Cuáles son las funciones y objetivos del Comité de IA?
¿De qué órgano o personas debería depender el Comité de IA?

93 Los expertos señalaban los siguientes ejemplos: reclamaciones, segmentación, diseño de ofertas personalizadas, tramitación de correos, chatbox, etc.

5.1 Procedimiento de implementación de la Gobernanza en el uso de la IA.

La gestión de una empresa lleva aparejada la evaluación de todos aquellos riesgos que puedan llegar a afectar a la misma, y los procesos de uso de IA no son una excepción. La transformación digital y el uso de tecnología en nuestro entorno pone cada vez con mayor frecuencia probabilidades de uso de herramientas con componentes de IA, bien por iniciativa de la organización, o en ocasiones de forma espontánea por el uso no consciente de algún proceso o empleado que la incluya. En todos los supuestos la empresa debe poseer una gobernanza adecuada que comprenda la gestión de riesgos, el establecimiento eficaz de controles internos y de medidas para su supervisión.

En los resultados de la encuesta realizada, se pone de manifiesto que se está incluyendo el análisis de otro tipo de riesgos relativos a la información y a la seguridad, así como asuntos digitales, pero no se incluye expresamente la IA. Es posible que exista una intención de control a nivel formal, pero hay una necesidad de bajar la metodología a la realidad de las empresas, dado que se trata de un sistema en plena evolución que debe monitorizarse de forma continua.

Desde el punto de vista del gobierno corporativo pueden enumerarse algunos aspectos fundamentales a tener en cuenta en el proceso de implantación en el uso de IA de la siguiente forma:

5.2 Definición del riesgo de la empresa por el órgano de administración

En primer paso en la gestión de este proceso es que la empresa cuente con un adecuado análisis del riesgo. Para ello la organización debe realizar un análisis previo que defina su nivel de exposición al riesgo y su intención de uso de las tecnologías y particularmente la IA. El perfil del riesgo y los límites de tolerancia en el uso de IA debe formar parte del análisis general del sistema de gestión de riesgos de la empresa y de su planificación estratégica.

La evaluación y decisión sobre ello, constituye un objetivo estratégico, que requiere un proceso de toma de decisiones final por parte del órgano de Administración.

El máximo órgano responsable de la gestión de la empresa debe determinar, en primer lugar, su interés o no en hacer uso de estas herramientas en el desarrollo del interés social de la entidad. En segundo lugar, si es aceptable, o no, el uso de estos sistemas en procesos de negocio (la empresa se encuentra en condiciones y madurez suficiente para incluirla en su gestión) o por el contrario se limita a algunos procesos corporativos.

En algunas ocasiones, y atendiendo al apetito al riesgo de la entidad en esta materia y el tipo de negocio, los responsables pueden determinar el uso de un sistema de IA de riesgo alto[94]. Finalmente puede considerarse que su utilización es inaceptable.

Relacionado con lo anterior, y la implicación de los directivos en los procesos de implementación de la IA la encuesta puso de manifiesto, en el 75% de las respuestas, que el grado de implicación del encuestado era ningún o muy bajo. En este aspecto hay que tener en cuenta que la encuesta fue respondida en un 41 % por miembros de consejo de administración y personas pertenecientes a la Dirección ejecutiva[95].

El proceso de decisión anterior debe estar documentado, deben participar, con su apoyo y asesoramiento, aquellas áreas de la empresa experta en estos riesgos y aquellas funciones fundamentales del sistema de gobierno de una organización como la dirección de riesgos y la dirección de cumplimiento. El propósito es identificar y evaluar entre todas las vulnerabilidades de la empresa y los beneficios del uso de esas herramientas, conteniendo la información necesaria para la toma de decisión.

Finalmente, el uso de los sistemas de IA deberá incluirse en la política de tecnologías de la información de la empresa aprobado por el órgano de Administración. Y, como cualquier política y proceso organizativo debe estar sujetos a actualizaciones, cambios, y supervisión.

94 La consideración de un sistema de IA de alto riesgo viene definido en el artículo 6 del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial.

95 La encuesta fue respondida por personas pertenecientes en un 12,2% al Consejo de administración y en un 28,6 % a la Dirección ejecutiva.

5.3 Diseño y control interno en el uso del sistema de IA en los procesos de la empresa.

La evaluación del riesgo previo y su aceptación determinara posteriormente la selección de aquel o aquellos procesos (o procedimientos) de negocio o corporativos en los que la empresa decida la aplicación de IA.

El proceso de implementación sea a partir de desarrollo interno o mediante la adquisición de un sistema a un proveedor, debe ser nuevamente objeto de análisis sobre la conveniencia y oportunidad, sus riesgos y vulnerabilidades, impacto, testeos iniciales, y revisiones periódicas.

En los supuestos de externalización, se deberá seguir la política implantada por la entidad en este tipo de procesos realizados por terceros, siendo especialmente rigurosos en la selección del proveedor.

Finalmente, una buena gobernanza corporativa en la implementación por la empresa de herramientas de IA, debe tener en cuenta en su evaluación de riesgos y controles la posibilidad de ocurrencia de diversos escenarios relacionados con la aplicación de IA que se exponen a continuación.

- El primer escenario sería el uso del sistema de IA de acuerdo con la finalidad prevista. Esto es el uso definido por el proveedor que incluye el contexto y las condiciones de uso concretos, según la información facilitada por el proveedor en las instrucciones de uso, los materiales y las declaraciones de promoción y venta, y la documentación técnica[96].
- El segundo escenario se refiere a un posible uso indebido razonablemente previsible, qué, si bien no se ajusta estrictamente a la finalidad prevista y definiciones de uso del proveedor, su uso puede derivarse de un comportamiento humano o una interacción con otros sistemas, incluidos otros sistemas de IA, razonablemente previsible.

96 El concepto de “finalidad prevista” viene definido en el artículo 3 apartado 12 del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial.

- Y, el tercer escenario a considerar es la posibilidad de que en su uso se devenguen prácticas prohibidas.

En relación con la implementación, los resultados de la encuesta muestran que hoy en día, la gestión de la IA sigue teniendo un componente eminentemente teórico para las empresas, y que de forma generalizada no se estaría incluyendo en las políticas de riesgos tecnológicos aquellos relativos a IA. A pesar de la rápida evolución y de la influencia directa e indirecta que puede tener en los negocios, la realidad es que de forma generalizada la IA no está considerada en el marco práctico de los riesgos de las mepresas. Esto que aparentemente podría indicar que el impacto del riesgo es mínimo, no refleja la realidad, dado que puede dar un vuelco debido a la velocidad de evolución de la IA y su capacidad disruptiva.

5.4 Actualización, revisión y ética en el uso de los sistemas de IA.

La gobernanza en el uso de los sistemas de IA en una organización debe estar sujeta a actualizaciones, revisiones periódicas y a un proceso de supervisión.

El sistema de control interno de la empresa, al margen de sus controles *in situ* (propios del proceso), debe prever con la periodicidad adecuada verificaciones sobre la adecuación del sistema de IA utilizado por la empresa.

El contenido de la verificación debe asegurar que la organización está utilizando un sistema de IA alineado con su ética empresarial, cultura y valores. Se debe comprobar que los controles implantados son eficaces de forma que el sistema de IA corresponde al definido.

En el ámbito del control de la ética y un uso adecuado de la IA, al margen de las áreas de control interno y cumplimiento, supone una buena práctica constituir un Comité de IA.

El uso de IA en la empresa puede hacer emerger importantes riesgos operacionales, de ética y reputacionales no deseados que puedan producir un grave perjuicio en el cumplimiento del interés social de la empresa. Un ejemplo de ello sería el uso de una IA no ajustadas al perfil del riesgo de la empresa o el uso de prácticas prohibidas de IA.

La creación de un Comité de IA en la organización implica la búsqueda de un equipo de apoyo y asesoramiento especializado. Su composición debería ser heterogénea estando representadas distintas áreas y perfiles de la empresa como dirección general, responsables de cumplimiento, gestión de riesgos, personas del área financiero, comercial. Los encuentros con varios profesionales han manifestado la importancia de que estén presentes asesores externos que aporten asesoramiento objetivo y al margen del negocio de la empresa.

Las funciones y objetivos del Comité de IA se extenderían a velar por el cumplimiento de la legalidad, además del cumplimiento ético y la transparencia que cada organización quiera incluir entre las funciones y objetivos de este órgano.

Una cuestión relevante con relación a este comité es determinar de quienes debería depender el comité de IA. Las respuestas obtenidas en los encuentros mantenidos con profesionales han sido variadas:

- Destaca en primer lugar la dependencia de la alta dirección, seguida por el consejo de administración y, por último, otros órganos.
- Y, desde un punto de vista de optimización de gobernanza se considera que en el actual momento en el que nos encontramos debería ser el órgano de administración el que supervisara y controlara, directa o indirectamente, el Comité de IA.

En definitiva, este Comité es un órgano de soporte que puede colaborar con la empresa para el diseño de políticas, procedimientos, introducir reglas de actuación en caso de riesgos emergentes etc. Aunque la empresa decida alejarse del uso de sistemas de IA de alto riesgo no se está totalmente exento de que los sistemas generen escenarios prohibidos o de alto riesgo que no se ajusten a la cultura y valores de la organización. La empresa debe estar preparada para ello, teniendo capacidad de atender inmediatamente la contingencia detectada y minimizar al máximo ese riesgo emergente.

5.5 Formación

Finalmente, la concienciación, sensibilidad y formación de la empresa en su conjunto con relación al uso de los sistemas de IA es una medida solida de control interno.

La formación podría consistir sobre lo que es un sistema de IA, su utilidad, sus escenarios de uso, la comprensión sobre el contenido de un sistema de IA de riesgo alto y sobre las prácticas prohibidas de IA en el ámbito de la UE. Todo ello potenciará el conocimiento de los directivos y empleados sobre esta materia y mitigará riesgos futuros de transgresión de políticas internas o de un uso espontaneo que provoque a la empresa graves perjuicios.

6. EL IMPACTO Y RIESGOS DE LA UTILIZACIÓN DE HERRAMIENTAS DE IA EN LA GESTIÓN DE PERSONAS.

La gestión de riesgos tiene como objetivo último procurar el beneficio social y el desarrollo socioeconómico desde el análisis y actuación frente a circunstancias concreta; y, en particular, el RIA plantea como fin último fomentar los beneficios que la IA puedan generar en las personas, por lo que este ámbito adquiere una especial importancia dentro del marco empresarial.

Los procesos de Recursos Humanos estructuran la contribución de las personas a la generación de bienes y servicios de todo tipo a la par que pueden proporcionar otros intangibles a los propios trabajadores, medibles únicamente en términos de subjetivos de satisfacción y realización personal.

La optimización de los procesos de RRHH mediante IA es un aspecto de actualidad en cuanto puede suponer en una mejora sustancial en su eficiencia al automatizar parte de ellos y generar conocimiento adicional que a su vez mejora la productividad en lo procesos de negocio. Cuanto más intensivo en el factor trabajo sea un sector determinado, mayor será la utilidad marginal potencial del uso de la IA.

Algunos de los procesos en los que la IA puede impactar de forma más significativa son los siguientes:

(i) Selección de Personal. Filtrando de manera automática los CVs que mejor se adaptan a la Descripción del Puesto de Trabajo (DPT), realizando entrevistas virtuales o a través de chatbot o incluso entrevistas por video en las que se pueden utilizar patrones gestuales y de voz

(ii) Seguimiento de objetivos y evaluación del desempeño, analizando datos de performance, dando *feedback* instantáneo y adaptando los objetivos automáticamente a los cambios en las circunstancias.

(iii) Predicción de bajas voluntarias identificando y analizando combinaciones de comportamientos y actitudes y otorgando probabilidad personalizada de bajas voluntarias. Esta predicción puede ser particularmente útil para aplicar política de retención personalizadas combinadas con la gestión del talento de la empresa.

(iv) Mayor automatización de tareas recurrentes de tipo administrativo (nóminas y seguridad social) o legal (cumplimiento de obligaciones por parte de la empresa).

Y, esto son solo algunos de los ejemplos más frecuentemente citados por profesionales de RRHH y consultoras especializadas en la provisión de servicios a la función. No cabe ninguna duda que la lista es mucho más amplia y permanecerá constantemente abierta.

Desde el enfoque del análisis del riesgo, y atendiendo a dichos ejemplos podríamos destacar los siguientes:

(i) Sesgos. Posibilidad de que los sesgos en los algoritmos nos lleven a resultados contra la normativa o simplemente indeseados en términos de equidad. Sentencia del Tribunal de Bolonia 2949/2020, de 31 de diciembre que analiza un caso en que el algoritmo contaba con dos parámetros a la hora de valorar a los trabajadores, siendo estos la fiabilidad y disponibilidad. Sin embargo, a la hora de cumplir con dichos parámetros no se tienen en cuenta circunstancias del trabajador como el derecho a huelga o la posibilidad de enfermedad o accidente de trabajo, de tal modo que se produciría una discriminación hacia determinados trabajadores al no valorar los derechos que tienen por el mero hecho de serlo.

(ii) Riesgos de distorsión por parte de externos, entre los que destaca la automatización de los filtros de los CV o las entrevistas a través de *chatbots*. Por otra parte, el candidato, puede acudir a herramientas de IA que fabriquen un número indeterminado de versiones de su CV para adaptarlo a las distintas propuestas de trabajo, y así también podría utilizar la IA para dar las mejores respuestas posibles a las en-

trevistas automatizadas distorsionando la competencia entre candidatos y la elección del más idóneo.

(iii) Riesgo de inadaptación de las plantillas. Existen sectores económicos enteros con plantillas relativamente envejecidas que no se van a adaptar fácilmente a la utilización de herramientas de IA. La formación adaptativa puede muy bien no ser suficiente para que estas personas cambien su *mindset* en función de las nuevas necesidades. Puede ser por tanto necesario proceder a renovar buena parte de las plantillas con los costes económicos y sociales inherentes, así como las dificultades para captar nuevo talento digital.

(iv) Privacidad. La materia prima de los procesos de recursos humanos son personas, sus capacidades, su rendimiento, sus actitudes e incluso sus emociones. La ingente cantidad de datos que la IA puede manejar para obtener resultados útiles y sobre todo el salto cualitativo que se produce en el carácter de esos datos (ya nos son solamente datos objetivos sino comportamientos, reacciones y otros factores personales) hace temer por la importancia de su salvaguardia no solo en la cuestión de su seguridad sino de los límites del acceso a ellos por parte de los distintos estamentos de la empresa. La sentencia C-634/21 del Tribunal de Justicia de la Unión Europea, tiene como trasfondo el algoritmo de la empresa SCHUFA, agencia de información sobre solvencia de terceros. El algoritmo tenía como fin la generación automatizada de valores de puntuación crediticia (o *credit scoring*). A través de una serie de procedimientos matemáticos y estadísticos, y un análisis del historial de solvencia de diferentes consumidores, el algoritmo permitía predecir la posibilidad de que un determinado cliente falte al pago de sus deudas. El demandante, había visto su crédito denegado en base a la ejecución de dicho algoritmo.

(v) La elaboración de datos complejos a partir de observaciones del comportamiento de las personas puede resultar en riesgos de inequidad de difícil prevención.

(vi) Riesgos de dilución de la responsabilidad de tomar decisiones. Tradicionalmente las decisiones sobre valoración del rendimiento de las personas, gestión de su talento, formación, promoción, reconocimiento, contratación o incluso desvinculación de la empresa, se toman por personas (directivos o comités) valorando una serie de informaciones objetivos y subjetivas. La elaboración de esa información previa utilizando herramientas de IA puede llevar a que la responsabilidad de las decisiones no sea enteramente suya, sino que resulte de alguna manera compartida con los algoritmos que han generado y en ultimo termino condicionado la decisión. Pueden producirse por tanto ineficiencias o decisiones no éticas que se repitan en el tiempo al basarse en procesos de elaboración de información distorsionados.

7. CONCLUSIONES.

El desarrollo de la IA es uno de los elementos más disruptivos del momento actual, cuyo impacto puede notarse en todos los ámbitos del entorno social y económico. Y, al mismo tiempo implica la proliferación y amplificación de ciertos riesgos globales entre los que destaca: i) la pérdida de empleos y reemplazo de la mano de obra; ii) el uso con fines criminales, desarrollo de ciberataques y de la información falsa; iii) los sesgos y la discriminación; y, iv) su integración en armamento y utilización fines bélicos.

El marco jurídico de la inteligencia artificial (IA) expuesto en el texto revela varios aspectos críticos que deben ser considerados para una adecuada gestión de tecnologías basadas en IA. Se destaca la importancia de un enfoque basado en el riesgo sobre cada caso de uso, que permite identificar y mitigar las potenciales amenazas que la IA puede representar para los derechos fundamentales, la salud y la seguridad pública. La clasificación de los sistemas de IA en categorías de riesgo -prohibidos, altos, limitados y bajos o nulos- a los que hay que añadir los sistemas de IA de uso general con o sin riesgo sistémico, es esencial para establecer requisitos específicos y obligaciones que garanticen su uso responsable.

Además, la regulación enfatiza la necesidad de transparencia en el desarrollo e implementación de sistemas de IA, y regula la obligación específica de transparencia con el objetivo de facilitar información clara y específica sobre sus capacidades y limitaciones. La inclusión de los diferentes roles que intervendrán dependiendo si es un sistema o un modelo de IA, junto con las responsabilidades que se asocian a cada rol, asegura un marco de colaboración y de cumplimiento normativo.

El RIA fomenta un compromiso con el desarrollo ético y responsable que proteja a la sociedad en su conjunto sin limitarse a regular sobre el uso de tecnologías basadas en IA. El marco normativo además de responder a las necesidades actuales también dispone de mecanismos de flexibilidad con el objetivo de adaptarse a la evolución tecnológica y garantizar que la IA contribuya al bienestar humano y al progreso social.

Tras las encuestas realizadas podemos concluir que en la actualidad existen muy pocas organizaciones que dispongan de una política de gobernanza en el desarrollo y utilización de IA. Es frecuente la previsión de riesgos tecnológicos, pero es escasa la regulación de los temas específicos de IA.

El control de la implantación de IA es clave para la buena gobernanza. Por ello se considera muy conveniente constituir un Comité de IA (con representación transversal de varios departamentos, incluso con la participación de expertos externos), que permitirá facilitar:

- Que se evite un uso incorrecto, no ético y perjudicial de la IA.
- Que el órgano de administración pueda acreditar una diligencia en la prevención de los riesgos derivados de la IA.
- La mejora en la gestión de oportunidades a nivel corporativo

En todo caso, tras el análisis realizado podemos constatar un escaso uso de sistemas de IA en las organizaciones, si bien simultáneamente se comprueba el posible uso individual por empleados en áreas para agilizar alguna tarea, sin conocimiento de la organización y con el riesgo que eso genera.

En definitiva, nuestro estudio trata de fomentar el análisis de riesgos y desarrollo de las políticas de gobernanza de la IA en las organizaciones, lo que permitirá seguir trabajando en la implementación de nuevos sistemas y procesos de forma segura. La visión de la gestión de riesgos resulta esencial para afrontar el reto de la IA de forma competitiva y segura, y permite a las organizaciones adelantarse a los problemas y avatares que pueden surgir en su implementación. Y, finalmente, la gestión de riesgos permite poner en el centro de estos desarrollos al ser humano y los principios éticos que deben considerarse los fines últimos y valores esenciales sobre el que pivotará la integración responsable de la IA en nuestras organizaciones.

8. ANEXO I

Análisis del Cuestionario sobre el Uso de la Inteligencia Artificial en las Organizaciones.

Pregunta 1: Sector de Actividad Principal de la Organización
Esta pregunta busca identificar el sector en el cual operan las organizaciones encuestadas.

Resultados:

- **Sector Asegurador:** La mayoría de las respuestas, aproximadamente el 60%, corresponden a este sector, indicando una alta representación de empresas de seguros.
- **Otros sectores:** El 40% restante se divide entre sectores como Transporte, Infraestructuras, Ingeniería, Logística, Salud, Energía Renovable y Jurídico.

Pregunta 2: Cargo dentro de la Organización
Esta pregunta recoge el nivel jerárquico o el rol específico de los encuestados.

Resultados:

- **Dirección Ejecutiva:** Aproximadamente un 50% de los encuestados se encuentran en la dirección ejecutiva, lo que sugiere que la mayoría de las opiniones provienen de niveles altos en la jerarquía organizacional.
- **Consejo de Administración y Gestión de Riesgos:** Entre el 30% restante, se encuentran cargos en el Consejo de Administración y roles en áreas de Riesgos y Cumplimiento.
- **Otros cargos:** Alrededor de un 20% representan roles como Ingeniería, Área Legal, Auditoría Interna y Analistas de Riesgos.

Pregunta 3: Grado de Implementación de IA en los Sistemas y Procesos (Escala del 1 al 5)
Esta pregunta mide el nivel de desarrollo de IA en los sistemas y procesos de la organización, siendo 1 “ninguno” y 5 “gran desarrollo”.

Resultados:

- **Nivel bajo de implementación (1 y 2):** Un 70% de las respuestas indican niveles bajos (1 o 2), lo cual sugiere que la mayoría de las organizaciones están en etapas iniciales o no han implementado IA en sus operaciones.
- **Nivel medio de implementación (3):** Un 20% indica un nivel medio de implementación.
- **Nivel alto de implementación (4 y 5):** Apenas el 10% de las organizaciones reporta tener un desarrollo significativo de IA.

Pregunta 4: Procesos en los que se Utiliza la IA

Esta pregunta examina en qué áreas o procesos específicos se está utilizando la IA.

Resultados:

- **No se utiliza IA:** Un 65% menciona que no se utiliza la IA en ningún proceso.
- **Procesos de negocio y atención al cliente:** Del 35% restante, se observa el uso de IA principalmente en procesos de negocio y en atención al cliente, con menor mención a prevención de incumplimientos.

Pregunta 5: Ejemplos de Uso de IA en la Organización

Esta pregunta permite conocer ejemplos específicos de implementación de IA en distintas organizaciones.

Resultados:

- **No utilizan IA:** El 60% reporta no utilizar IA.
- **Casos de uso específicos:** El 40% restante indica ejemplos como Machine Learning para eficiencia energética y riesgos, CRM para gestión de clientes, traducción y soporte en mails, venta cruzada y prevención de fraudes.

Pregunta 6: Grado de Conocimiento y Vinculación con Proyectos de IA (Escala del 1 al 5)

Esta pregunta evalúa el nivel de familiaridad e involucramiento de los encuestados con los proyectos de IA en sus organizaciones.

Resultados:

- **Vinculación baja (1 y 2):** Un 70% indica tener un vínculo bajo (1 o 2) con los proyectos de IA.
- **Vinculación media a alta (3, 4 y 5):** Un 30% muestra un mayor grado de participación o conocimiento.

Pregunta 7: Formación Relativa a la IA y sus Riesgos

Se explora si las organizaciones incluyen formación sobre IA en sus programas de capacitación anual.

Resultados:

- **No incluye formación:** Un 60% indica que no se realiza formación específica sobre IA.
- **Sí incluye formación:** El 40% reporta que sus programas de formación anual abordan la IA, aunque algunos mencionan que es limitada o relacionada solo con transformación digital.

Pregunta 8: Órganos de Toma de Decisiones para Proyectos de IA

Esta pregunta identifica los departamentos responsables de la toma de decisiones sobre proyectos de IA.

Resultados:

- **Dirección Ejecutiva:** Cerca del 50% señala que la dirección ejecutiva es la encargada de las decisiones.
- **Consejo de Administración:** Un 30% menciona el Consejo de Administración junto con la Dirección Ejecutiva.
- **Otros órganos:** El 20% restante indica otras áreas o combina diferentes órganos.

Pregunta 9: Órgano de Control para la Supervisión de IA
Esta pregunta indaga si existe un órgano específico para la supervisión de herramientas de IA.

Resultados:

- **No existe órgano de control:** Un 40% afirma que no tienen un órgano de control específico.
- **Área de TI:** Otro 35% menciona que el control se realiza principalmente en el Área de TI.
- **Consejo de Administración o Dirección Ejecutiva:** Un 25% señala que estos órganos participan en la supervisión.

Pregunta 10: Probabilidad de Error en Sistemas de IA (Escala del 1 al 5)
Esta pregunta evalúa la percepción del riesgo de error en los sistemas de IA.

Resultados:

- **Poca probabilidad (1):** Un 50% considera baja la probabilidad de error.
- **Media probabilidad (3):** Un 30% percibe un nivel de riesgo moderado.
- **Alta probabilidad (4 y 5):** Un 20% ve alta la probabilidad de error en los sistemas de IA.

Pregunta 11: Impacto Potencial de un Error en IA (Escala del 1 al 5)
Esta pregunta estima el posible impacto de un error en IA en la organización.

Resultados:

- **Impacto bajo (1 y 2):** El 60% considera que el impacto sería bajo.
- **Impacto medio (3):** Un 25% cree que el impacto sería moderado.
- **Impacto alto (4 y 5):** Un 15% percibe que el impacto sería considerablemente alto.

Pregunta 12: Políticas de Prevención en Riesgos Tecnológicos
Esta pregunta explora la existencia de políticas para prevenir riesgos tecnológicos.

Resultados:

- **Sí, pero no incluyen IA:** Un 60% menciona políticas de prevención sin referencias a IA.
- **Sí, incluyendo IA:** Un 30% indica que sus políticas de prevención tecnológica sí contemplan la IA.
- **No existen políticas:** El 10% restante afirma no contar con políticas específicas en este ámbito.

Pregunta 13: Políticas de Mitigación y Gestión de Brechas de Seguridad con Referencias a IA
Esta pregunta examina las políticas de mitigación de seguridad y su relación con la IA.

Resultados:

- **Sí, pero no incluyen IA:** Un 70% tiene políticas de mitigación que no abordan la IA.
- **Sí, incluyendo IA:** El 20% señala que sus políticas sí consideran aspectos de IA en la seguridad.
- **No existen políticas:** Un 10% no tiene políticas de mitigación de brechas de seguridad.

Pregunta 14: Otras Políticas Internas que Incluyan Referencias a IA
Esta pregunta indaga si existen otras políticas que mencionen el uso de IA.

Resultados:

- **Ninguna política relacionada:** Un 80% indica que no hay otras políticas con referencias a IA.
- **Normativas de Ciberseguridad y Privacidad:** Un 20% menciona que sus políticas de TI y privacidad de datos incluyen aspectos de IA.

Pregunta 15: Influencia de la IA en los Procesos de Decisión (Escala del 1 al 5)
Se mide el grado en que la IA afecta la toma de decisiones en la organización.

Resultados:

- **Grado bajo de influencia (1 y 2):** Un 70% indica una influencia baja.
- **Grado medio (3):** Un 20% percibe un impacto moderado.
- **Grado alto (4 y 5):** Solo el 10% siente que la IA influye considerablemente en la toma de decisiones.

Pregunta 16: Valor Añadido de la IA para el Negocio (Escala del 1 al 5)
Esta pregunta explora si la IA aporta valor añadido al negocio.

Resultados:

- **Valor bajo (1 y 2):** Un 65% percibe un valor añadido bajo.
- **Valor moderado (3):** Un 20% considera que aporta un valor medio.
- **Valor alto (4 y 5):** Solo el 15% observa que la IA ofrece un valor significativo.

Conclusiones Generales

Este análisis sugiere que la mayoría de las organizaciones encuestadas están en una fase inicial de implementación de la IA, con una baja integración en procesos críticos y políticas específicas. La percepción de riesgo y el bajo impacto en la toma de decisiones reflejan una oportunidad para explorar más usos y establecer marcos de gobernanza en IA.

La guía **"Gestión del Riesgo de la Inteligencia Artificial"** ofrece un marco completo para comprender y abordar los riesgos asociados al desarrollo y uso de la IA en el ámbito empresarial. A través de un enfoque basado en estándares internacionales y normativas como el Reglamento (UE) 2024/1689, el documento detalla las mejores prácticas para identificar, mitigar y gestionar los riesgos inherentes a estas tecnologías disruptivas.

Además, analiza en profundidad el impacto de la IA en los derechos fundamentales, la seguridad y la economía global, destacando la importancia de la gobernanza ética y responsable. La guía incluye estudios de caso, herramientas prácticas y recomendaciones para implementar sistemas de gestión de riesgos efectivos que no solo cumplan con las regulaciones actuales, sino que también anticipen desafíos futuros.

Dirigida a gerentes de riesgos, líderes empresariales y profesionales del sector, esta guía busca fomentar una integración segura y sostenible de la inteligencia artificial, promoviendo su adopción con un enfoque ético y estratégico para maximizar sus beneficios y minimizar sus riesgos.

agers