ÉTICA Y COMPLIANCE EN EL USO DE LA INTELIGENCIA ARTIFICIAL

Una iniciativa de

isms forum

Ética y Compliance en el uso de la Inteligencia Artificial

Copyright: Todos los derechos reservados. Puede descargar, almacenar, utilizar o imprimir la presente Guía sobre Ética y Compliance en el uso de la Inteligencia Artificial de ISMS Forum, atendiendo a las siguientes condiciones: (a) la guía no puede ser utilizada con fines comerciales; (b) en ningún caso la guía puede ser modificada o alterada en ninguna de sus partes; (c) la guía no puede ser publicada sin consentimiento; y (d) el copyright no puede ser eliminado del mismo.

AUTORES

DIRECTORES	Ángel Pérez
	Francisco Lázaro
COORDINADORES	Rubén Cabezas
PARTICIPANTES	Ana de la Higuera
	Berta Balanzategui
	Diego Fernandez
	Elena Mora
	Esther Garcia
	Ignacio Hornes
	María Cristina Köhler
	María Cumbreras
	María Pilar Alapont
	Miguel Angel Cabezas
	Pablo Diaz
	Silvia Tortosa
	Xavier Alberdi
GESTOR DE PROYECTOS	Beatriz García
DISEÑO/MAQUETACIÓN	Rim Souri

CONTENIDOS

CONTENIDOS

1 MARCO LEGAL Y NORMATIVO EN EL ÁMBITO DE LA IA

1.1. PROPUESTA EUROPEA DE MARCO JURÍDICO SOBRE IA

En el presente apartado se analizarán los elementos que la UE ha considerado más relevantes para abordar los riesgos generados por los usos específicos de la IA. Aunque se considera que la mayoría de los sistemas de IA no generan riesgos para los ciudadanos, ciertos sistemas de IA sí pueden dar lugar a riesgos o consecuencias negativas para personas concretas o la sociedad en su conjunto.

En este sentido, la UE considera que legislación vigente no proporciona una protección suficiente para hacer frente a los desafíos específicos que pueden plantear los sistemas de IA y, por este motivo, ha propuesto una serie de iniciativas legislativas que tienen como objetivo abordar los riesgos vinculados a determinados usos de esta nueva tecnología y, al mismo tiempo, proporcionar a los desarrolladores, implementadores y usuarios de IA la claridad que necesitan y facilitar, de este modo, la creación de una IA confiable.

En concreto, estas iniciativas legislativas europeas se agrupan en tres grupos fundamentales :

I La creación de un marco jurídico basado en cuatro niveles de riesgo diferentes: riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo .

II Un marco de responsabilidad civil adaptado a la era digital y a la IA ;

III Revisión de la legislación sectorial en materia de seguridad (por ejemplo, el Reglamento sobre máquinas o el Reglamento sobre la seguridad general de los productos).

[1] *Un enfoque europeo de la inteligencia artificial.*

[2] *Propuesta de marco normativo sobre inteligencia artificial.*

[3] *Directiva sobre responsabilidad por los daños causados por productos defectuosos. Adaptación de las normas de responsabilidad a la era digital, la economía circular y las cadenas de valor mundiales.*

[4] *Proposal for a Regulation of the European Parliament and of the Council on machinery products.*

[5] *Reglamento relativo a la seguridad general de los productos.*

Antes de abordar la propuesta de regulación de la inteligencia artificial -que, en el momento de cierre de este documento, está ya debatiéndose en el ámbito de los trílogos (Los trílogos, una forma de agilizar la tramitación de leyes en la Unión Europea (europa.eu) al haberse pronunciado los tres actores relevantes (Comisión, Consejo y Parlamento)- de cara a un mejor entendimiento de la misma, nos parece relevante abordar los antecedentes de esta norma que marcará un antes y un después en la creación y uso de la inteligencia artificial.

En este sentido, si nos centramos en el proceso en el ámbito de la Unión Europea los orígenes de esta regulación se remontan a marzo de 2018 cuando la Comisión Europea publica junto con la declaración del Grupo Europeo de Ética de la Ciencia y Tecnología, la creación de un grupo de expertos en inteligencia artificial a través del cual recabar la opinión de los expertos y forjar alianzas entre las diversas partes implicadas con la misión de elaborar un propuesta de Directrices sobre ética de la IA.

Desde ese primer hito y hasta el momento, en el ámbito europeo se han producido multitud de eventos que se detallan en el anexo I, con el objetivo de facilitar el entendimiento del proceso legislativo y el diseño de los procesos de implementación de la norma, que están llevando a cabo o deberán llevar a cabo las entidades que utilicen esta tecnología:

Entrando someramente en el detalle de los textos señalados, cabe destacar, y por ello profundizaremos en otros apartados de este documento, las Directrices éticas para una IA fiable articulan un marco para lograr una IA fiable (licita, ética y robusta), directrices que sin detenerse en el primer aspecto (licitud), -como veremos- identifican los principios éticos y valores conexos que deberían respetarse en el desarrollo, despliegue y utilización de los sistemas IA fiables. Estas directrices ofrecen una orientación muy práctica sobre cómo alcanzar esa IA fiable a través del cumplimiento de 7 requisitos (Acción y supervisión humana, solidez técnica y seguridad, gestión de la privacidad de los datos, transparencia, diversidad, no discriminación y equidad, bienestar social y ambiental y rendición de cuentas) basados en 4 grandes fundamentos (respeto a de la autonomía humana, prevención del daño, equidad y explicabilidad).

El segundo gran hito que merece la pena de los señalados anteriormente es el Libro Blanco sobre Inteligencia Artificial (febrero de 2020) en el cual se abordaba la necesidad de establecer:

- El marco político de la IA: A través de la creación de un ecosistema de excelencia en la UE para el desarrollo y la adopción de la IA).
- El marco regulador: A través de creación de un ecosistema de confianza que evite la fragmentación y que aborde la IA desde un enfoque a riesgos para los derechos y libertades de los ciudadanos).
- Una definición de IA suficientemente flexible para adaptarse al progreso y que a la vez permita la seguridad jurídica.
- Unos requisitos obligatorios para los sistemas de IA.
- Dirigidos a aquellos que estén en el mejor posición de cumplir en los cadena de desarrollo, implementación y uso de los sistemas de IA.
- Una gobernanza europea sobre la IA que, al igual que el marco regulador, huya de la fragmentación de responsabilidades.

A estos hitos cabría añadir, a nivel español, las dos guías que hasta el momento ha publicado la Agencia Española de Protección de Datos ("AEPD") y dos pronunciamientos sobre la norma que han realizado tres actores muy relevantes: el Banco Central Europeo ("BCX") , el Comité Europeo de Protección de Datos ("CEPD") y el Supervisor Europeo de Protección de Datos ("SEPD") .

Necesariamente cuando la utilización de una IA implique un tratamiento de datos personales deberemos tener en cuenta las dos guías de la AEPD. La primera guía "Adecuación al RGPD de tratamientos que incorporan IA" recurre a la definición de IA realizada por el grupo de Expertos en las Directrices para una IA fiable que mencionábamos con anterioridad y, haciendo un repasado de las principales obligaciones derivadas del RGPD, establece cómo, ante la existencia de algoritmos en los tratamientos de datos personales, deben cumplirse no solo los principios generales del artículo 5 del RGPD sino, además, como deben gestionarse los derechos reconocidos a los titulares de los datos, cómo debe procederse en la realización de las correspondientes Evaluaciones de impacto en la privacidad o cómo se deberá analizar el cumplimiento de las previsiones relativas a la elaboración de perfiles y la toma de decisiones basadas únicamente en un tratamiento automatizado, debiendo evitar el diseño de sistemas con la orientación "Dead man switch" y dar siempre la opción de que un operador humano pueda ignorar el algoritmo en un momento dado.

En 2021 se publica la segunda guía de la AEPD, "Requisitos para auditorias de tratamientos que incluyen IA" en la que se establecen 144 controles agrupados en 5 grandes materias (identificación y transparencia del componente, propósito del componente, fundamentos del componente, gestión de los datos y verificación y validación) que se perfilan como una guía muy útil a disposición de responsables y encargados del tratamiento para, a través de la supervisión de los tratamientos, se puedan determinar los planes de acción necesarios para que, por defecto y desde el diseño, los tratamientos de datos que incluyan inteligencia artificial cumplan con el RGPD.

Acabamos esta sección, como avanzábamos, con referencias a dos pronunciamientos: el Dictamen "BCE" y el Dictamen conjunto CEPD-SEPD sobre la propuesta de Reglamento de IA.

[1] *A European approach to artificial intelligence*

[2] *Directrices éticas para una IA fiable*

[3] *LIBRO BLANCO sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza*

[4] *Adecuación al RGPD de tratamientos que incorporan IA. Una introducción (febrero de 2020) y Requisitos para auditorias de tratamientos que incluyen IA (enero de 2021)*

[5] *OPINION OF THE EUROPEAN CENTRAL BANK of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence*

[6] *Dictamen conjunto 5/2021 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial*

En relación con el primero, señalamos su importancia debido a que la propuesta de Reglamento de IA contiene disposiciones que afectan a las competencias del BCE, en particular a sus funciones relacionados con la supervisión prudencial de las Entidades de crédito que, como sabemos, se verán impactadas al encontrarse los sistemas destinados a evaluar la solvencia de las personas físicas o a establecer su calificación crediticia entre los sistemas de IA de alto riesgo. En este sentido, el BCE ha solicitado clarificación de sus competencias en el marco de la propuesta de Reglamento de IA en 3 aspectos:

I. La autoridad de vigilancia de los mercados.

II. La evaluación de la conformidad requerida a los sistemas de IA.

III. Las competencias de supervisión prudencial en general.

Hay que destacar finalmente de este Dictamen que el BCE admite la necesaria supervisión del SEPD en relación con el uso del supervisor bancarios de sistemas de IA y hace una llamada a no considerar IA los sistemas que "aprovechan el uso independiente de la regresión lineal o logística o los diagramas de decisiones bajo supervisión humana siempre que el efecto de dichos enfoques aplicados a la evaluación de la solvencia o la calificación crediticia de las personas físicas no sea significativo".

En relación con el segundo pronunciamiento, el Dictamen conjunto del CEPD y SEPD destacan sus llamadas a garantizar la claridad de:

- La relación de la propuesta de Reglamento de IA con la legislación vigente en materia de protección destacando, entre otros, si el enfoque basado en los factores de riegos de la propuesta y el concepto de "riesgo para los ¬derechos fundamentales" se ajusta al previsto en el RGPD, o bien si la interpretación relativa a si un tipo de tratamiento puede dar lugar a un riesgo elevado de conformidad con el RGPD, debe hacerse con independencia de la propuesta de Reglamento de IA, mientras que la clasificación de un sistema de IA como «de alto riesgo» debido a su impacto en los derechos fundamentales debido a la mencionado propuesta daría lugar a una presunción de «alto riesgo» en el marco del RGPD, o la configuración como requisito previo al marcado como producto de la CE del cumplimiento de las obligaciones legales derivadas de su legislación.
- Necesidad de constante actualización de las listas donde se concreten los sistemas de alto riesgo.
- Necesidad de mejora de la lista de supuestos de prácticas prohibidas para incluir la clasificación social de las personas en general por cualquier actor (público o privado) o la inferencia de emociones de las personas.
- La dicotomía entre las obligaciones para los proveedores de sistemas de IA con relación a la evaluación de riesgos cuando en muchos casos los responsables de tratamiento serán los usuarios de los mismos, así como la indisponibilidad del proveedor para evaluar todos los usos de su sistema de IA (lo que parece que propone solucionar a través de la realización de EIDPs por parte de estos).
- Competencias del SEPD en relación con ser autoridad de vigilancia del mercado y las autoridades de protección de datos nacionales, como autoridades de supervisión para garantizar un enfoque regulador más armonizado.
- Ámbito de aplicación y los objetivos de los espacios de pruebas.

1.2. PROPUESTA DE REGLAMENTO SOBRE IA (UE)

1.2.1. INTRODUCCIÓN

Tal y como se recoge en la Exposición de Motivos de la propuesta de Reglamento en materia de IA, el objetivo de este Reglamento consiste en abordar los riesgos vinculados a determinados usos de esta nueva tecnología para inspirar confianza en los ciudadanos, con el objetivo de que puedan adoptar soluciones basadas en la IA al tiempo que trata de animar a las empresas a que desarrollen este tipo de soluciones. Por este motivo, el Reglamento se centra en las personas, a fin de que los ciudadanos tengan la seguridad de que esta tecnología se usará de un modo seguro y con respecto a sus derechos fundamentales.

De la misma manera, debido a la rápida evolución de la IA, la propuesta de Reglamento tiene un enfoque a prueba de futuro que permitirá que las normas se adapten a los rápidos cambios tecnológicos.

En este sentido, las razones y objetivos específicos que se han identificado en la propuesta de Reglamento son los siguientes:

- Garantizar que los sistemas de IA introducidos y usados en el mercado de la UE sean seguros y respeten la legislación vigente en materia de derechos fundamentales y valores de la Unión.
- Garantizar la seguridad jurídica para facilitar la inversión e innovación en IA.
- Mejorar la gobernanza y la aplicación efectiva de la legislación vigente en materia de derechos fundamentales y los requisitos de seguridad aplicables a los sistemas de IA.
- Facilitar el desarrollo de un mercado único para hacer un uso legal, seguro y fiable de las aplicaciones de IA y evitar la fragmentación del mercado.

Tal y como veremos a continuación, la propuesta de Reglamento establece normas armonizadas para el desarrollo, la introducción en el mercado y la utilización de sistemas de IA en la UE a partir de un enfoque basado en los riesgos, con la finalidad de abordar los riesgos creados específicamente por las aplicaciones de IA y definir aquellos sistemas de IA que plantean un "alto riesgo", que deberán cumplir una serie de requisitos obligatorios que garanticen su fiabilidad. También se propone una definición única de la IA que puede resistir el paso del tiempo.

Asimismo, prohíbe determinadas prácticas particularmente perjudiciales de IA por ir en contra de los valores de la UE y propone restricciones y salvaguardias específicas con relación a determinados usos de los sistemas de identificación biométrica remota con fines de aplicación de la ley.

Finalmente, la propuesta de Reglamento impondrá obligaciones específicas a los proveedores y los usuarios durante todo el ciclo de vida de los sistemas de IA, en especial para los sistemas de alto riesgo, proponiendo una evaluación de la conformidad antes de que el sistema de IA se ponga en servicio o se introduzca en el mercado. En este mismo sentido, se determina que las aplicaciones de IA deberán seguir siendo fiables, incluso después de haber sido comercializadas, circunstancia que requerirá una gestión continua de los riesgos por parte de los proveedores de esta tecnología.

1.2.2. DEFINICIÓN DE UN SISTEMA DE IA, PRÁCTICAS DE IA PROHIBIDAS Y CLASIFICACIÓN DE SISTEMAS DE IA

Según hemos analizado, la intención de la UE consiste en articular un marco normativo que permita asegurar la fiabilidad de las IA de tal manera que garanticen los derechos humanos consagrados en Carta de los Derechos Fundamentales de la Unión Europea (la «Carta de la UE»), así como en la pertinente legislación internacional de derechos humanos. De esta manera la UE quiere crear el primer marco jurídico sobre IA y desempeñar un papel de liderazgo a nivel mundial, el denominado "*Efecto Bruselas*".

Para alcanzar dichos objetivos, la propuesta de Reglamento establece unos requisitos mínimos necesarios para subsanar los riesgos y problemas vinculados a la IA, sin obstaculizar ni impedir indebidamente el desarrollo tecnológico y sin aumentar de un modo desproporcionado el coste de introducir soluciones de IA en el mercado.

Asimismo, prohíbe determinadas prácticas particularmente perjudiciales de IA y propone restricciones y salvaguardias específicas en relación con determinados usos de los sistemas de identificación biométrica remota. La propuesta establece además una metodología de gestión de riesgos para definir aquellos sistemas de IA que plantean un "alto riesgo".

Este planteamiento inicial está recogido en el actual borrador del Reglamento sobre IA donde propone una definición única de la IA que pretende ser lo más tecnológicamente neutra posible y resistir al paso del tiempo lo mejor posible, habida cuenta de la rápida evolución tecnológica y del mercado en relación con la IA.

En concreto, con la definición en su artículo 31 de lo que se considera un "sistema de IA":

» Sistema basado en máquinas que puede, para un conjunto determinado de objetivos definidos por seres humanos, hacer predicciones, recomendaciones o decisiones que influyan en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con diferentes niveles de autonomía.

[1] DIRECTRICES ÉTICAS para una IA FIABLE 2018 Comisión Europea

Claramente se aprecia en el planteamiento de la UE los siguientes aspectos:

- Presentar la IA como un sistema en oposición a una percepción antropomorfista, la cual nos llevaría a pensar en la IA como un ser humano en su forma de" pensar y actuar".
- Asimilar su funcionamiento a mecanismos complejos formados por múltiples elementos, como podrían ser las piezas que forman un reloj
- Supervisar sus interactuaciones por seres humanos, dejando poco margen a un independencia o autonomía plena de estos sistemas
- Evitar que tomen el control sobre los entornos, permitiendo únicamente su influencia.

Unos planteamientos totalmente garantistas de derechos frente a unos posibles usos y prácticas intrusivas o limitadoras de derechos y libertades. Siguiendo por este camino, el regulador es consciente, antes que cualquier otra cosa, del mal uso que se puede llevar a cabo con los sistemas de IA, por ello lo primero que establece en el borrador de Reglamento de IA son las prácticas consideradas como prohibidas.

Si bien la última versión no recoge una lista exhaustiva de sistemas o inteligencias artificiales prohibidas en la UE, propone en su Título II un listado cuyo uso se considera no aceptable por ser contrarios a los valores de la UE, como por ejemplo los siguientes sistemas:

- Puntuación social, donde se podrían asignar valoraciones o clasificaciones a las personas en función de su comportamiento, intereses, conductas sociales, características personales, orientaciones sexuales o ideológicas, entre otras.
- El uso de sistemas de identificación biométrica remota y en tiempo real en espacios públicos como el reconocimiento facial, salvo excepciones limitadas.
- Manipulación o explotación del comportamiento humano mediante técnicas subliminales, de tal manera que tienen un elevado potencial para trascender su consciencia, o que aprovechan las vulnerabilidades de grupos concretos como los menores o las personas con discapacidad para alterar de manera sustancial su comportamiento, de un modo que es probable que les provoque perjuicios físicos o psicológicos a ellos o a otras personas.

En el resto de los casos, este mismo marco regulatorio aborda los riesgos asociados con el uso de la IA, y establece que algunas prácticas pueden ser restringidas o reguladas en el futuro. Para ello establece una clasificación basada en tres niveles de riesgo, de acuerdo con el siguiente esquema:

- Riesgo inaceptable, usos prohibidos tal y como vimos anteriormente.
- Alto riesgo.
- Riesgo bajo o mínimo.

De conformidad con el considerando 27, la calificación «de alto riesgo» debe limitarse a aquellos sistemas de IA que tengan consecuencias perjudiciales importantes para la salud, la seguridad o los derechos fundamentales de las personas físicas de la Unión, y dicha limitación reduce al mínimo cualquier posible restricción del comercio internacional, si la hubiera. Dichos sistemas de IA están permitidos en el mercado europeo siempre que cumplan determinados requisitos obligatorios y sean sometidos a una evaluación de la conformidad ex ante

Se establecen unas normas de clasificación del sistema de IA como de riesgo algo en atención a la función que lleve a cabo, la finalidad específica y las modalidades para las que se use dicho sistema. De igual modo, se definen las dos categorías principales de sistemas de alto riesgo:

I **Sistemas de IA diseñados para utilizarse como componentes de seguridad de productos sujetos a una evaluación de la conformidad ex ante realizada por terceros.**

II **Otros sistemas de IA independientes que se recogen explícitamente en el Anexo III, listado que podrá ser ampliado por la Comisión.**

Otros sistemas de IA:

- Infraestructuras críticas, debido a que su fallo o defecto podrían poner en peligro la vida y la salud de los ciudadanos.
- Educación o formación profesional, en especial aquellas que determinan el acceso o distribuyen a las personas entre distintas instituciones educativas y de formación profesional o aquellas que evalúan a las personas a partir de pruebas realizadas en el marco de su educación o como condición necesaria para acceder a ella.
- Componentes de seguridad de los productos, o que son productos en sí mismos, como pueden ser máquinas, juguetes, ascensores, equipos y sistemas de protección para uso en atmósferas potencialmente explosivas, equipos radioeléctricos, equipos a presión, equipo de embarcaciones de recreo, instalaciones de transporte por cable, aparatos que queman combustibles gaseosos, productos sanitarios y productos sanitarios para diagnóstico in vitro.
- Empleo, gestión de los trabajadores y el acceso al autoempleo, sobre todo para la contratación y la selección de personal; para la toma de decisiones relativas a la promoción y la rescisión de contratos; y para la asignación de tareas y el seguimiento o la evaluación de personas en relaciones contractuales de índole laboral, dado que pueden afectar de un modo considerable a las futuras perspectivas laborales y los medios de subsistencia de dichas personas.
- Servicios públicos y privados esenciales o necesarios para que las personas participen en la sociedad o cuenten con unas condiciones de vida mejores.
- Actuaciones de las autoridades encargadas de la aplicación de la ley y que puedan interferir con los derechos fundamentales de las personas.
- Gestión de la migración, el asilo y el control fronterizo.
- Administración de justicia y procesos democráticos, dado que pueden tener efectos potencialmente importantes para la democracia, el Estado de Derecho, las libertades individuales y el derecho a la tutela judicial efectiva y a un juez imparcial.
- Identificación biométrica remota, ya que pueden dar lugar a resultados sesgados y tener consecuencias discriminatorias

1.2.3. REQUISITOS DE LOS SISTEMAS DE IA DE ALTO RIESGO Y RESPONSABILIDADES DE DIVERSOS AGENTES DE LA CADENA DE VALOR DE LA IA

Todos los sistemas anteriormente citados y catalogados de "Alto Riesgo" estarán sujetos a un conjunto de obligaciones previas a su comercialización y antes de que puedan ser integrados en los usos anteriormente citados, siendo estas:

- **Gestión de riesgos:** Con sistemas adecuados de evaluación y mitigación de riesgos durante todo el ciclo de vida de un sistema de IA.
- **Gobernanza de datos:** Utilización de técnicas de entrenamiento de modelos a partir de conjuntos de datos de entrenamiento, validación y prueba que cumplan los criterios de alta calidad de los conjuntos de datos, que alimentan el sistema para minimizar los riesgos y los resultados discriminatorios.
- **Documentación técnica:** Incluirá como mínimo los elementos contemplados en el Anexo IV del Reglamento.
- **Trazabilidad:** Con un registro de la actividad para garantizar la evaluación de los resultados y la aparición de situaciones que presenten un riesgo.
- **Transparencia:** Documentando detalladamente toda la información necesaria sobre el sistema y su finalidad.
- **Información:** Clara y adecuada para el usuario.
- **upervisión humana:** Medidas adecuadas para minimizar el riesgo.
- **Fiabilidad:** Con un alto nivel de robustez, seguridad y precisión.

Fuera de los casos anteriores, la propuesta de Reglamento plantea la existencia de otros sistemas que serán catalogados como de alto riesgo y que estarán regulados bajo otras disposiciones normativas y, en algunos casos, sujetos a la autorización de un órgano judicial u otro organismo independiente y a límites apropiados en el tiempo, en el alcance geográfico y en las fuentes de datos consultadas.

Los requisitos anteriores quedaran asociados a los diversos agentes que forman parte la cadena de valor de la IA de alto riesgo. La finalidad es que estos asuman una serie de responsabilidades que garanticen la seguridad y cumplimiento normativo exigidos a los sistemas de IA, de acuerdo con esto se han identificado los siguientes actores y los requisitos que tienen que cumplir:

ACTORES	REQUISITOS
FABRICANTES	▪ Cumplir con los requisitos de seguridad establecidos en el reglamento. ▪ Llevar a cabo una evaluación exhaustiva de riesgos y realizar pruebas adecuadas antes de poner el sistema en el mercado. ▪ Proporcionar documentación técnica completa y precisa que describa el diseño, el funcionamiento y las características del sistema de IA. ▪ Etiquetar el sistema de IA claramente para indicar que es un sistema de alto riesgo y proporcionar información sobre el uso seguro y las limitaciones. ▪ Establecer mecanismos adecuados para la detección y gestión de eventos adversos y para la notificación de incidentes a las autoridades competentes.
IMPORTADORES Y DISTRIBUIDORES	▪ Asegurar que los sistemas de IA que ponen en el mercado cumplan con los requisitos del reglamento. ▪ Verificar que los fabricantes han llevado a cabo la evaluación de riesgos adecuada y proporcionen la documentación técnica requerida. ▪ Mantener registros de los sistemas de IA que importan o distribuyen y garantizar que estén disponibles para las autoridades competentes cuando sea necesario.
USUARIOS	▪ Utilizar los sistemas de IA de manera adecuada, siguiendo las instrucciones del fabricante y cumpliendo con las limitaciones y advertencias indicadas. ▪ Informar a los fabricantes sobre cualquier evento adverso o incidente que ocurra durante el uso del sistema de IA. ▪ Cooperar con los fabricantes y proveedores en la aplicación de las medidas de mitigación de riesgos y en la implementación de las acciones correctivas necesarias.

1.2.4. ACLARACIÓN DEL ÁMBITO DE APLICACIÓN DE LA PROPUESTA DE REGLAMENTO DE INTELIGENCIA ARTIFICIAL Y DISPOSICIONES RELATIVAS A LAS AUTORIDADES ENCARGADAS DE LA APLICACIÓN DE LA LEY

El ámbito de aplicación del futuro Reglamento de Inteligencia Artificial es muy amplio y abarca todos los niveles de riesgo que pueden presentarse. No obstante, las disposiciones más estrictas están enfocadas a los sistemas de IA catalogados de Alto Riesgo, cuya finalidad es garantizar unos niveles de seguridad adecuados y un respeto a los derechos fundamentales que rigen la normativa de la UE.

En cuanto a las autoridades encargadas de la aplicación de esta futura norma, el borrador establece la creación de una autoridad europea de inteligencia artificial como un órgano independiente. Esta autoridad tendría la responsabilidad de supervisar y hacer cumplir el Reglamento, así como de emitir directrices y opiniones sobre su aplicación.

Al igual que en otras normativas, el Reglamento prevé que cada estado miembro de la UE designe una o varias autoridades nacionales competentes en materia de inteligencia artificial. Estas autoridades nacionales serían responsables de la supervisión y aplicación del reglamento a nivel nacional, incluyendo la realización de inspecciones y la imposición de sanciones en caso de incumplimiento.

Además, el borrador del Reglamento recoge la cooperación entre las autoridades nacionales y la autoridad europea a fin de garantizar una aplicación coherente y eficaz del reglamento en toda la UE.

A este respecto, en España se creó en 2020 la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), cuyos Estatutos han sido publicados a finales de agosto de 2023. De este modo, España se ha convertido en el primer país de la Unión Europea con un organismo destinado a la supervisión de la IA, adelantándose a la entrada en vigor del futuro Reglamento Europeo sobre IA, que establece la obligación de que los Estados miembros dispongan de una autoridad supervisora en esta materia.

1.2.5. EVALUACIONES DE CONFORMIDAD, MARCO DE GOBERNANZA, VIGILANCIA DEL MERCADO, APLICACIÓN Y SANCIONES

En el borrador del Reglamento se establecen disposiciones relacionadas con las evaluaciones de conformidad, el marco de gobernanza, la vigilancia del mercado, la aplicación y las sanciones, brevemente reseñamos cada uno de estos aspectos:

- **Evaluaciones de conformidad:** Se establecerá que los sistemas de IA de alto riesgo deberán someterse a una evaluación de conformidad antes de ser puestos en el mercado o utilizados. Estas permitirán verificar que los sistemas cumplen con los requisitos y obligaciones establecidos en el reglamento.

- **Marco de gobernanza:** Se propondrá establecer un marco de gobernanza de la inteligencia artificial que incluya la creación de una autoridad europea de IA, como se indicó anteriormente, sería la responsable de supervisar su cumplimiento y de emitir directrices y opiniones sobre la aplicación del reglamento

- **Vigilancia del mercado:** La responsabilidad de llevar a cabo las actividades de vigilancia del mercado y del cumplimiento normativo recaerá sobre las autoridades nacionales. Para ello, podrán llevar a cabo inspecciones, recopilar información sobre los sistemas de IA en el mercado y tomar medidas en caso de incumplimiento

- **Aplicación y sanciones:** Se prevé la imposición de sanciones, por parte de las autoridades nacionales competentes, cuando se detecte un incumplimiento de lo establecido en el reglamento. Siguiendo las directrices de otros reglamentos, como el RGPD, se establecerán multas proporcionales a los ingresos anuales o al volumen de negocios del infractor. Además, se establece que los Estados miembros deben establecer medidas efectivas, proporcionadas y disuasorias para garantizar la aplicación del reglamento y el cumplimiento de las sanciones.

1.2.6. TRANSPARENCIA Y OTRAS DISPOSICIONES A FAVOR DE LAS PERSONAS AFECTADAS

El futuro Reglamento prevé incluir disposiciones a favor de las personas afectadas por los sistemas de IA. Podemos agrupar estas medidas en dos grandes tipos: las aplicadas exclusivamente a los sistemas de IA de Alto Riesgo y aquellas que recaerán en el resto de los sistemas de IA.

Comenzaremos por el análisis de los requerimientos específicos de los sistemas de Alto riesgo:

- **Transparencia:** Los proveedores de IA deberán proporcionar información clara y comprensible sobre el funcionamiento del sistema, incluyendo sus capacidades y limitaciones. Esto incluye la divulgación de información sobre los datos utilizados, los algoritmos empleados y los posibles riesgos asociados.
- **Aplicabilidad:** Cuando los sistemas interactúen con personas, se exige que los proveedores puedan proporcionar explicaciones sobre las decisiones tomadas por el sistema, cuando sea requerido por la persona afectada. Esto permite una mayor comprensión y rendición de cuentas en situaciones críticas.
- **Evaluación de impacto en los derechos fundamentales:** Los proveedores deberán realizar una evaluación de impacto en los derechos fundamentales antes de su despliegue. Esta evaluación debe identificar y abordar los posibles riesgos para los derechos fundamentales y adoptar medidas adecuadas para mitigarlos.

El resto de las disposiciones se aplicarán a todos los sistemas de IA y están enfocadas hacia los siguientes aspectos:

» **Derechos fundamentales y no discriminación:** Los sistemas de IA deberán respetar los derechos fundamentales de las personas y no pueden ser utilizados para discriminar de forma injusta o injustificada. Esto incluye la prohibición de la creación o despliegue de sistemas de IA que vallan en contra de ellos.

» **Protección de datos personales:** el futuro reglamento está totalmente alineado con el RGPD y establece que los sistemas de IA deben cumplir con los todos los principios de protección de datos personales.

1.2.7. MEDIDAS DE APOYO A LA INNOVACIÓN

El futuro reglamento de incluirá medidas de apoyo a la innovación en el ámbito de la inteligencia artificial. Con el objetivo de fomentar la investigación, el desarrollo y la adopción de estas tecnologías en la UE. Algunas de estas medidas consisten en las siguientes:

- **Sandbox de IA:** Estableciendo unos entornos de experimentación y prueba controlados en la fase de desarrollo y previa a la comercialización, con vistas a garantizar que los sistemas de IA innovadores cumplan lo dispuesto en el futuro Reglamento y en otra legislación pertinente de la Unión y los Estados miembros
- **Programas de ayudas:** Destinados a apoyar la investigación y el desarrollo de tecnologías de IA.
- **Apoyo a las pymes:** Mediante el fomento de la cooperación entre empresas, centros de investigación y universidades en el ámbito de la IA. Además, facilitar el acceso a recursos, financiación o conocimientos para desarrollar soluciones de IA innovadoras
- **Estándares y certificaciones:** Promoviendo estándares técnicos o certificaciones para los sistemas de IA. Con la idea de impulsar la confianza, la interoperabilidad, la adopción de tecnologías de IA y promoviendo una competencia justa.

1.3. PRINCIPALES IMPACTOS REGULATORIOS

1.3.1. IMPACTO EN LA PRIVACIDAD

Aunque se han identificado de manera indirecta los riesgos potenciales en materia de privacidad a lo largo del presente documento, hemos destacado en este apartado el impacto específico del nuevo marco legal que regulará la IA con la normativa que actualmente existe con relación al tratamiento de datos personales (el RGPD y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de Derechos Digitales ("LOPD").

En el caso del RGPD, al tratarse de un reglamento horizontal con un enfoque intersectorial establece una serie de principios generales que se pueden aplicar en el contexto de la inteligencia artificial siempre que exista un tratamiento de datos personales. Aunque en el RGPD no se mencionan expresamente las tecnologías de Inteligencia Artificial esto se debe a que su enfoque es tecnológicamente neutro, con el objetivo de garantizar su adaptación a las tecnologías emergentes y a los nuevos usos que puedan venir (Considerando 15).

Por lo tanto, la protección de los datos personales no debe depender de las tecnologías utilizadas para su tratamiento debido a que esto podría conllevar un riesgo de elusión de esta protección , sin imponer ni discriminar el uso de la IA, o cualquier otra tecnología, para conseguir la protección de cualquier tratamiento de datos personales.

Es necesario recordar la especial relevancia que supondrá el cumplimiento de la normativa europea de protección de datos en las distintas fases de los sistemas de IA, debido a que el recurso fundamental de esta tecnología consiste en el tratamiento de datos, entre los que necesariamente se encontrarán datos personales, tanto en el desarrollo de la IA, como a la hora de utilizarla para analizar o tomar decisiones. A la vista del gran impacto en la protección de datos en el desarrollo y despliegue de la IA, resultará imprescindible que las empresas definan la base de legitimación para el tratamiento de la información personal y limitar las finalidades.

[1] Sentencia del Tribunal de Justicia en el asunto C–25/17

En una de las guías publicadas por la AEPD que se han mencionada anteriormente , se describe de manera detallada los posibles tratamientos de datos personales en las diferentes etapas del ciclo de vida de una solución de IA:

- **Entrenamiento:** Este tratamiento podría incorporar, entre otros, las actividades de minería de datos para la obtención del conjunto de datos de interés, el preprocesamiento de la información, la partición del conjunto de datos para verificación (splitting), así como la trazabilidad y auditoría.

- **Validación**: Aunque esta fase también puede realizarse a través de modelos analíticos, en el caso de que se utilicen datos que se correspondan con la situación real de los afectados y resulten distintos de aquellos utilizados en la etapa de entrenamiento, resultará necesario realizar una nueva identificación de la legitimación y limitar las finalidades del tratamiento de manera diferenciada del análisis realizado en la fase anterior.

- **Despliegue:** Durante esta etapa podrían producirse un tratamiento de datos personales en aquellas situaciones en las que la solución de IA permita identificar a una persona física viva o hacerla identificable como, por ejemplo, cuando la lógica del modelo de IA contenga ejemplos que puedan llevar a la identificación de determinadas personas cuyos datos hubieran sido utilizados en la fase de entrenamiento.

- **Explotación:** La AEPD ha identificado los tratamientos de información personal que podrían producirse en las distintas actividades de explotación de la solución IA, en concreto en los siguientes:

    - **INFERENCIA**: Cuando se traten datos personales del interesado o de un tercero o cuando datos e inferencias del interesado se almacenen.

    - **DECISIÓN:** Cualquier decisión que adopte el sistema de IA relativo a una persona física, identificada o identificable, constituye un tratamiento de datos personales sujeto al cumplimiento de las obligaciones establecidas en el RGPD.

    - **EVOLUCIÓN:** Resultaría igualmente aplicable el RGPD en esta etapa en los supuestos en los que se envíen a terceros los datos y resultados de personas físicas con el objetivo de modificar el modelo de IA.

- **Retirada:** Resultará igualmente aplicable el principio de conservación establecido en el RGPD, de manera que los datos personales solo podrán conservarse durante el menor tiempo posible y, por ello, resultará necesario que las entidades establezcan internamente plazos máximos de supresión o revisión de los datos en los sistemas de IA, con base al principio de responsabilidad proactiva, que deberán tener en cuenta los motivos por los que resultaría necesario mantener el tratamiento de los datos, así como las obligaciones legales que pudieran resultar de aplicación para la conservación de los datos en el sistema de IA.

[1] *Guía Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*

Por otro lado, resulta destacable la similitud existente entre la metodología de gestión de los riesgos de la propuesta de Reglamento de la IA, con el enfoque basado en los factores de riesgo en el tratamiento de la información personal establecido en el RGPD. A su vez, la propuesta de Reglamento presenta una propuesta similar a la realización de los Evaluaciones de Impacto en la Privacidad previstas en el RGPD, que consiste en la incorporación de evaluaciones del cumplimiento de las obligaciones establecidas en el Reglamento de IA ex ante en los sistemas de IA de alto riesgo.

Conviene explicar que la propuesta de Reglamento de la IA contiene una mención específica a la obligación del cumplimiento de otros requisitos derivados de la legislación de la UE en materia de protección de datos, con el objetivo de evitar el riesgo identificado por el CEPD y SEPD consistente en que un sistema de IA, a pesar de estar certificado (marcado CE) en virtud de la propuesta de Reglamento, para su introducción en el mercado o puesta en servicio, no cumpla con las normas y principios de protección de datos y, en particular, de protección de datos desde el diseño y por defecto.

Por otro lado, en las versiones iniciales de la propuesta de Reglamento de IA no se prohibía de manera expresa el posible uso de sistemas de IA para la categorización de las personas a partir de la biometría (como el reconocimiento facial) en función del origen étnico, el género, así como la orientación política o sexual, u otros motivos prohibidos de discriminación, o sistemas de IA cuya validez científica no se encontraba demostrada o, incluso, que estaban en conflicto directo con los valores esenciales de la UE. No obstante lo anterior, debido precisamente al impacto de estas medidas en la privacidad de los ciudadanos, tanto el CEPD como el SEPD solicitaron la incorporación de una prohibición general del uso de la IA para el reconocimiento automatizado de rasgos humanos en espacios de acceso púbico, la clasificación de las personas a partir de sus datos biométricos en grupos por razón de su origen étnico, sexo, orientación política o sexual u otros motivos de discriminación, o la utilización de esta tecnología para la inferencia de emociones, como así se ha recogido finalmente en las últimas propuestas del Reglamento de IA.

Relacionado con lo anterior, en el marco de la consulta pública realizada en la Propuesta de Directiva para adaptar las normas de responsabilidad extracontractual en la IA, cuyas implicaciones se analizarán más adelante, el CEPD publicó en febrero de 2022 una carta subrayando algunos elementos que desde una óptica de privacidad deberían considerarse destacando, entre otras cuestiones, que este nueva normativa debería garantizar la seguridad en el tratamiento datos personales y el uso de sistemas de IA. Es conveniente destacar que, de conformidad con el RGPD, solo los responsables o encargados del tratamiento pueden ser responsables, por ejemplo, en el supuesto de una violación de la seguridad de los datos personales, de manera que resultaría esencial reforzar el régimen de responsabilidad de los proveedores de IA.

En el RGPD podemos identificar un conjunto de artículos que tienen un impacto directo sobre el uso de la inteligencia artificial cuando existe un tratamiento de datos personales, entre los que se destacan a continuación los aspectos más relevantes que las empresas deberán tener en cuenta en aquellos tratamientos de datos personales que incluyan IA, pero sin que se pretenda realizar una descripción exhaustiva de lo establecido en el RGPD.

Ente ellos podemos identificar los principios de protección de datos desde el diseño y por defecto, de licitud, lealtad y transparencia, el principio de limitación de la finalidad, de minimización de datos, de exactitud, de limitación del plazo de conservación y de responsabilidad proactiva :

- **ARTÍCULO 5 – Principios relacionados con el tratamiento:** Cuando usamos la IA y este uso implique el tratamiento de datos personales debemos cumplir estos principios y debe realizarse la evaluación de riesgos correspondiente. El principio de transparencia, minimización y exactitud toman especial relevancia en este tipo de tratamientos.

- **ARTÍCULOS 6 AL 11- Legitimación y licitud del tratamiento:** Los sistemas de IA pueden tratar grandes cantidades de datos personales en las distintas etapas del ciclo de vida, de manera que resultará necesario disponer de una base legitimadora sobre las cuales se podría basar el tratamiento de datos personales. El RGPD establece varias bases jurídicas distintas que pueden utilizarse en cada etapa del ciclo de vida de un sistema de IA, en función de la finalidad perseguida. Las bases jurídicas más habituales que legitimarán el tratamiento de datos personales en una solución de IA son la ejecución de un contrato, la aplicación de medidas precontractuales, el cumplimiento de obligaciones legales, el interés legítimo, en especial para tratamientos que requieren acceso a datos de entrenamiento pero que requerirá, en cualquier caso, que la empresa realice internamente y con carácter previo una prueba de sopesamiento para poder aplicar esta base jurídica, así como el consentimiento de los interesados, por ejemplo, en ciertos casos específicos que requieren el uso de categorías especiales de datos, sin perjuicio de que resultará necesario levantar la prohibición previa establecida en el citado artículo 9 de RGPD con las limitaciones adicionales establecidas en la LOPD, también en su artículo 9.

  Relacionado con lo anterior, el principio de limitación de la finalidad del RGPD determina que los datos no podrán ser tratados ulteriormente para finalidades distintas para las que se recogieron inicialmente pero, paradójicamente, los sistemas de IA requieren a menudo el uso de mucha información que las empresas podrían haber obtenido inicialmente para otros fines, circunstancia que requerirá un análisis interno previo respecto de si la finalidad del tratamiento de los datos personales que serán reutilizados para esta tecnología resultan compatibles o no con la finalidad inicial para la cual se obtuvieron dichos datos.

  A este respecto, las empresas pueden tener en cuenta que el artículo 89 del RGPD determina que las operaciones de tratamiento ulterior con fines de archivo en interés público, fines de investigación científica e histórica o fines estadísticos deben considerarse operaciones de tratamiento lícitas compatibles, siempre que se cumpla con el resto de las garantías y excepciones establecidas en el RGPD.

[1] *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*

- **ARTÍCULOS 13 Y 14 – Información**: El contenido concreto de la información que se ha de proporcionar a los interesados se deberá adaptar a la etapa del ciclo de vida de la IA en la que se realice el tratamiento. Se establece la posibilidad de ofrecer esta información mediante capas o niveles.

  De conformidad con los criterios que ha determinado la AEPD, en la primera capa de información se deberá ofrecer a los usuarios la siguiente información:

  - La identidad del responsable del tratamiento o de su representante.
  - La finalidad del tratamiento.
  - La posibilidad de ejercer los derechos 15 al 22 RGPD.
  - Si el tratamiento incluye la elaboración de perfiles o decisiones automatizadas.
    - Hay que informar claramente que se produce esta circunstancia.
    - Informando de su derecho a oponerse a la adopción de decisiones individuales automatizadas de acuerdo con el art. 22 RGPD.
    - Información significativa sobre la lógica aplicada,
    - Importancia y las consecuencias previstas de dicho tratamiento para el interesado.
  - Si los datos personales objeto del tratamiento no han sido obtenidos directamente del afectado, la información básica incluirá también:
    - Las categorías de datos objeto de tratamiento.
    - Las fuentes de las que procedieran los datos.
  - En la segunda capa, el resto de la información establecida en los artículos 13 y 14 del RGPD.

Este principio pretende evitar que el sistema de IA resulte hermético para las personas cuyos datos resulten objetivo de tratamiento, de manera que solo se obtenga información básica respecto de las características de entrada concreta de datos y su salida, pero sin obtener una explicación relevante respecto del funcionamiento interno del sistema y, de manera especial, la toma de decisiones en los procesos de entrenamiento y creación del modelo de IA generativa, lo que permitirá a los individuos comprender y anticipar cómo se comportará el sistema en una situación particular.

Por este motivo, la información que se facilite al interesado debe resultar fácilmente accesible y comprensible, por lo que deberá ser explicada en un lenguaje sencillo y claro con diferentes grados de detalle en la explicación del modelo, dependiendo del individuo y el contexto de las decisiones basadas en IA, con el objetivo de evitar los denominados sistemas de "caja negra" en los que no se explica de manera adecuada su funcionamiento.

- La AEPD ha recordado que no resultará admisible ofrecer simplemente una referencia técnica a la implementación del algoritmo, e introduce una serie de ejemplos que detallamos a continuación que podrán ser implantados por las empresas con el objetivo de evitar facilitar información a los interesados que pueda resultar opaca, confusa e, incluso, conducir a la fatiga informativa.

  - » El detalle de los datos empleados para la toma de decisión, más allá de la categoría, y en particular información sobre los plazos de uso de los datos (su antigüedad).
  - » La importancia relativa que cada uno de ellos tiene en la toma de decisión.
  - » La calidad de los datos de entrenamiento y el tipo de patrones utilizados.
  - » Los perfilados realizados y sus implicaciones.
  - » Valores de precisión o error según la métrica adecuada para medir la bondad de la inferencia.
  - » La existencia o no de supervisión humana cualificada.
  - » La referencia a auditorías, especialmente sobre las posibles desviaciones de los resultados de las inferencias, así como la certificación o certificaciones realizadas sobre el sistema de IA. En el caso de sistemas adaptativos o evolutivos, la última auditoría realizada.
  - » En el caso de que el sistema IA contenga información de terceros identificables, la prohibición de tratar esa información sin legitimación y de las consecuencias de realizarlo .

- **ARTÍCULOS 15 A 21 - Ejercicios de derechos.** Es necesario garantizar la atención de las solicitudes de derechos que se reciban adaptando los procedimientos internos y los canales de atención de este tipo de solicitudes a las soluciones de IA utilizadas en la entidad. Pueden resultar necesaria una aproximación caso a caso en los derechos supresión de los datos recogidos durante la etapa de entrenamiento de los modelos, incluir las medidas de bloqueo de los datos relativos al proceso de inferencia o la rectificación de los datos generados por los perfiles elaborados por la IA.

- **ARTÍCULO 22 – Decisiones individuales automatizadas y perfilados.** Un sistema de IA puede facilitar la creación de perfiles y la automatización de las decisiones, y tienen el potencial de afectar de forma significativa a los derechos y libertades de las personas de manera que requieren unas garantías adecuadas para abordar los riesgos específicos derivados de este tipo de tratamientos. En concreto, se deberá facilitar información significativa sobre la lógica aplicada de manera que el afectado pueda entender el comportamiento del tratamiento.

- **ARTÍCULO 25 - Protección de datos desde el diseño y por defecto.** La complejidad que rodea la Inteligencia Artificial hace necesario utilizar un enfoque orientado a la gestión del riesgo y de responsabilidad proactiva para establecer estrategias que incorporen la protección de la privacidad a lo largo de todo el ciclo de vida de una solución de IA, pasando por las fases comunes a todas las soluciones de IA: concepción y análisis, desarrollo, explotación y Retirada final. El objetivo último consiste en que la protección de datos esté presente desde las primeras fases de desarrollo y no sea una capa añadida a un producto o sistema.

- **ARTÍCULO 35 - Evaluación de impacto relativa a la protección de datos ("EIPD").** Tal y como se prevé en el nuevo marco regulatorio los sistemas de Inteligencia Artificial de alto riesgo van a requerir la realización de EIPD, sin perjuicio de su posible ampliación cuando se realice la elaboración de perfiles, basados en tratamientos automatizados (pero no necesariamente exclusivamente automatizados), sobre los que se tomen decisiones que produzcan efectos jurídicos para las personas físicas o que les afecten significativamente en las soluciones de IA.

    - Por otro lado, existen una serie de principios en el RGPD que pueden resultar claves en el desarrollo de la IA. Como hemos comentado, para el desarrollo de una IA en el que exista un tratamiento de datos personales se deben aplicar necesariamente un conjunto mínimo de principios con el objetivo de garantizar la conformidad del tratamiento realizado a la normativa de protección de datos. Dentro de estos principios básicos podemos destacar los siguientes debido a su criticidad en este tipo de tratamientos :

PRINCIPIO DE TRANSPARENCIA: La propuesta de Reglamento de IA establece un concepto de transparencia consistente en la información que proveedores de sistemas de IA destinan a los usuarios, entendidos como entidades que despliegan estos sistemas y afecta a diseñadores, desarrolladores, proveedores y usuarios/entidades que despliegan sistemas de IA. En cambio, esta definición difiere del mismo término establecido en el RGPD, que se refiere a categorías de información distinta, tanto en contenido como en redacción, y se dirigen a destinatarios diferentes y debe ser implementado, exclusivamente, por aquellas entidades que resulten responsables del tratamiento.

- En el caso de tratamientos basados en IA este principio cobra especial relevancia ya que en la mayoría de estos sistemas no será evidente que datos están siendo tratados ni cómo funcionan los sistemas que los tratan. El RGPD especifica que el responsable del tratamiento adoptará las medidas apropiadas para facilitar al interesado toda la información relativa al tratamiento de forma concisa, transparente, inteligible y fácilmente accesible, utilizando un lenguaje claro y sencillo, de manera que permita a los interesados ser conscientes del impacto que el empleo de dichas soluciones de IA puede llevar asociado.

- Existen unos riesgos específicos para la transparencia como, por ejemplo, facilitar información sobre la posibilidad de reidentificar al interesado a partir de los datos del modelo durante la etapa de entrenamiento, o los problemas de preservación de la propiedad intelectual a la hora de facilitar transparencia con relación a los algoritmos de IA.

- El propio RGPD determina que este principio es un aspecto crítico para garantizar la aplicación del principio de privacidad por defecto, que permite la supervisión del tratamiento por los propios afectados y conocer el impacto, capacidades y limitaciones que los sistemas de IA pueden llevar aparejados y, por este motivo, la transparencia afecta a todos y cada uno de los elementos y participantes que intervienen en la solución de IA.

- El DPO, figura clave dentro de las empresas para garantizar el cumplimiento de la privacidad, se constituye como una figura de gran relevancia para aquellas entidades que traten datos personales en soluciones basadas en IA, en particular en la fase de entrenamiento de los modelos, a través de la gestión del riesgo en materia de protección de datos personales y el cumplimiento del principio de responsabilidad proactiva.

PRINCIPIO DE EQUIDAD: Un aspecto crítico de los sistemas de IA consiste en la posible existencia de sesgos, que revisten especial relevancia cuando derivan en discriminaciones de un grupo en favor de otro. Algunos sesgos pueden afectar a la exactitud de los datos, como por ejemplo sesgos de evaluación y agregación cuando se construye un modelo de análisis, o sesgos de realimentación cuando la solución de IA es utilizada mayoritariamente por un grupo de sujetos con características particulares. En este sentido, se podrán utilizar técnicas orientadas a examinar y determinar la posible existencia de sesgos en los algoritmos utilizados en las soluciones de IA y a garantizar la equidad en la implementación del modelo. En el contexto de la Inteligencia Artificial, la equipad tiende a tener una aplicación más concreta ya que la posibilidad de existir un sesgo o discriminación dentro del modelo de IA (racial, de género, entre otros) pueden llevar a que el tratamiento sea considerado como ilícito.

PRINCIPIO DE EXPLICABILIDAD: Este principio significa la obligación de informar de manera significativa sobre la lógica aplicada que permita al interesado comprender el tipo de tratamiento que se está llevando a cabo en la solución de IA con sus datos personales, y le proporcione certeza y confianza sobre sus resultados. Si pensamos en términos RGPD esto se indica en el artículo 14 cuando se habla del contexto de los datos personales, en el que se indica que se deberá poner a disposición de los interesados información relevante para el interesado, como por ejemplo los datos empleados para la toma de la decisión, los patrones utilizados en la toma de decisión, los perfilados realizados y sus implicaciones, la existencia o no de supervisión humana, entre otras.

PRINCIPIO DE MINIMIZACIÓN: La aplicación de este principio requerirá que las empresas realicen, con carácter previo, un análisis de la tipología de datos que se deberán utilizar en cada fase del desarrollo de la IA. Esta situación puede suponer un reto para las empresas debido a que, en muchas ocasiones, resulta difícil establecer en un primer momento cuál será la información necesaria que deberá utilizar un determinado sistema de IA.

» Por otro lado, las organizaciones consideran, de manera errónea, que cuantos más datos y más variados el sistema de IA será mejor cuando, al contrario, más datos no necesariamente mejorarán el rendimiento de los modelos de IA, debido a que puede implicar la generación de mayores sesgos.

Del mismo modo, el RGPD exige un tratamiento proporcional de los datos a su finalidad, de manera que un tratamiento masivo y no justificado de datos vulneraría este principio que obliga a que los datos que se utilicen en la IA resulten adecuados, pertinentes y limitados a lo necesario con relación a las finalidades para los que son tratados dentro de un sistema de IA. Con el objetivo de ofrecer criterios prácticos a las empresas en la aplicación de este principio, se reproducen a continuación las indicaciones que la AEPD ha establecido:

- Limitar la extensión de las categorías de datos que se utilizan en cada fase del tratamiento a aquellas que son estrictamente necesarias y relevantes.
- Limitar el grado de detalle o precisión de la información, la granularidad de la recogida en tiempo y frecuencia y la antigüedad de la información utilizada.
- Limitar la extensión en el número de interesados de los que se tratan los datos.
- Limitar la accesibilidad de las distintas categorías de datos al personal del responsable/encargado o incluso al usuario final (si hay datos de terceros en los modelos de IA) en todas las fases del tratamiento.

PRINCIPIO DE EXACTITUD Y LIMITACIÓN DEL PLAZO DE CONSERVACIÓN: Este principio requiere implantar en el sistema de gestión de la IA una evaluación continua de las necesidades reales de cada sistema, incorporando las medidas necesarias para evitar el riesgo de inexactitud de los datos, debiendo para ello establecer mecanismos internos tendentes a la supresión o rectifican de los datos personales inexactos y garanticen su fidelidad e integridad.

- Relacionado con lo anterior, las empresas deberán incorporar dentro de su modelo de gobierno de la IA un análisis que determine el periodo de vida útil de los datos, de manera que se garantice un periodo de conservación y se eliminen aquellos datos que permitan la identificación de los interesados en el sistema por más tiempo del necesario.

- A este respecto, la AEPD ha recordado en diversos informes que el borrado material de los datos personales utilizados en los sistemas de IA quedaría excluido solo en las siguientes circunstancias: (i) para la puesta a disposición de los datos a los jueces y tribunales, el Ministerio Fiscal o las Administraciones Públicas competentes, en particular de las autoridades de protección de datos, y (ii) para la exigencia de posibles responsabilidades derivadas del tratamiento del sistema de IA y solo por el plazo de prescripción de estas. Este último punto debe interpretarse como el "plazo de prescripción de las acciones" encaminadas a la exigencia de tales responsabilidades derivadas de la distinta normativa que resulte aplicable a cada sistema de IA.

- A la vista de lo anterior, en aquellos supuestos en los que exista una normativa legal aplicable que exige la conservación de los datos por un plazo determinado, las empresas no deberían suprimir los datos. Una vez finalice este plazo, sí deberán proceder a la destrucción (borrado) de los datos, no pudiendo ser tratados para ninguna finalidad en el sistema de IA.

Finalmente, las empresas deberán valorar la utilización de la denominadas "Privacy Enhancing Technologies (PETs)", que permiten el desarrollo de algunos sistemas de IA sin necesidad de comunicar los datos personales entre los intervinientes, explotando la información de forma sostenible y protegiendo, de este modo, los derechos fundamentales. La AEPD ha destacado en numerosos artículos las bondades de este tipo de estrategias, en particular las arquitecturas de Federated Learning o aprendizaje federado.

1.3.2. IMPACTO EN LOS DERECHOS DE PROPIEDAD INTELECTUAL

Otra de los importantes impactos legales que se deberán abordar en el ámbito de la IA y que plantea un enorme reto, tanto para los autores como las empresas que comercializan su creación, y al que se deberá prestar especial atención, consiste en la protección de los derechos de propiedad intelectual. En concreto, el Parlamento Europeo alertó de la dificultad de trazar los derechos de propiedad intelectual y su aplicación a los resultados generados con apoyo y/o mediante esta tecnología, circunstancia que podría impedir que reciban una remuneración justa los creadores humanos, cuyo trabajo original se utilice para alimentar dichas tecnologías.

El riesgo fundamental consiste en la utilización de contenidos generados por IA que se encuentre protegidos por derechos de autor de terceros que puede, incluso, constituir un plagio o afectar a derechos de imagen, por ejemplo, en aquellos supuestos en los que la solución de IA pueda producir imágenes digitales a partir de una descripción en texto hecha por el usuario ("text prompt"), si la obtención y utilización de las imágenes de las bases de datos empleadas para el entrenamiento del sistema no cumple con las necesarias garantías legales.

Esta última cuestión presenta implicaciones importantes a la hora de determinar cuál debe ser la protección del resultado mediante el uso de IA, la autoría de una obra creada por un sistema de IA o los mecanismos de protección de la IA en sí misma. En la actualidad, la evolución de los sistemas de IA permite la creación de nuevas obras de forma autónoma, que incluyen la toma de decisiones independientes a lo largo de todo el proceso de manera que la obra es generada por el propio programa informático (red neuronal) a partir de unos parámetros definidos por los programadores.

La protección de las creaciones generadas por la IA dependerá, por lo tanto, de si nos encontramos ante una creación generada con asistencia de la IA o de forma autónoma por la IA. Si la creación la genera un ser humano con la asistencia de una IA, le corresponde al autor persona física la atribución de los derechos de propiedad intelectual sobre la obra resultante. En el segundo de los supuestos, cuando la persona humana únicamente solicita a la IA la generación de un contenido específico, se considera que la creación ha sido generada de forma autónoma por la IA y, debido a la ausencia de un autor persona física, la creación resultante no podría ser protegida de conformidad con el actual marco jurídico en materia de propiedad intelectual.

De manera complementaria, constituirá una obligación para las entidades la búsqueda de un necesario equilibrio entre las obligaciones de transparencia de la IA y la protección de los derechos de autor, circunstancia que permitirá que las empresas pueden beneficiarse de una protección eficiente y eficaz de sus derechos de propiedad intelectual. A este respecto, las obligaciones de transparencia no implican la divulgación de información técnica detallada sino información general del funcionamiento del sistema de IA como, por ejemplo, los datos utilizados por el sistema, las posibles comunicaciones a terceros o los posibles riesgos que pueden suponer para los derechos y libertades de los afectados, de manera que resultaría posible para las empresas facilitar información significativa a los usuarios de la IA sin que esto produzca una vulneración de los derechos de propiedad intelectual o industrial de las empresas.

1.3.3. RESPONSABILIDAD CIVIL

Las organizaciones también deberán analizar el impacto de la propuesta de Directiva sobre responsabilidad derivada del uso de la IA que establece nuevas reglas en los procedimientos de reclamación de daños y perjuicios y determina, entre otras cuestiones, normas específicas sobre quién debe ser el responsable de los daños ocasionados por un sistema de IA, tanto si es de alto riesgo como si no. De manera complementaria, esta Directiva tiene como objetivo que cualquier víctima, particular o empresa, pueda tener una oportunidad justa de indemnización en caso de perjuicio por la culpa u omisión de un proveedor, desarrollador o usuario de IA. Resulta también destacable que la propuesta de Directa abre la puerta a la posibilidad de establecer un régimen de seguro obligatorio para el uso de ciertos sistemas de IA por los posibles daños y perjuicios causados por éste.

Este nuevo marco normativo europeo resultará de especial relevancia para las empresas debido a que supone una reforma en los regímenes nacionales de responsabilidad subjetiva, así como su aplicación a las demandas civiles, de particulares o personas jurídicas, contra cualquier persona en las que se ejerciten, con base en responsabilidad por culpa o negligencia, reclamaciones de daños y perjuicios causados por un sistema de IA, contra cualquier persona que haya influido en dicho sistema.

1.3.4. OTRA NORMATIVA AFECTADA

Otras obligaciones que deberán asumir las entidades en el diseño y utilización de soluciones de IA consisten en garantizar el cumplimiento de la normativa, local y autonómica, aplicable a la protección de los consumidores en todo lo que se refiere a la interacción con los afectados, de manera especial con aquellos considerados vulnerables. De igual modo, las obligaciones de transparencia con los consumidores adquieren cuando se utilice esta tecnología una mayor relevancia.

La IA también puede suponer una capacidad desestabilizadora de la libre competencia por su capacidad de adoptar decisiones que impactan directamente en el mercado, pudiendo provocar prácticas colusorias o el abuso de posición dominante. Por este motivo, las entidades deberán verificar que el uso de esta tecnología no impide ni dificulta el desarrollo de una competencia efectiva, leal y no falseada.

1.4. DERECHO COMPARADO

1.4.1. ESTADOS UNIDOS

Estados Unidos se encuentra involucrado en varias propuestas para la regulación de la inteligencia artificial pero hoy en día no dispone de una normativa que dé respuesta a las necesidades regulatorias, aunque con motivo de la carta abierta para pausar el desarrollo de sistemas de IA , el actual presidente señaló la importancia de abordar los riesgos de la inteligencia artificial para la seguridad nacional y la economía.

En la carrera por adoptar una legislación garantista, en marzo de 2023 el Departamento de Ciencia, Innovación y Tecnología del Gobierno de Reino Unido, publicó el enfoque "A pro-innovation approach to AI regulation " donde identifican los matices de la futura propuesta legislativa para regular la inteligencia artificial señalando riesgos, oportunidades, herramientas para su confiabilidad y, uno de los puntos más importantes, la interoperabilidad mundial, señalando especialmente la importancia de participar en compromisos multilaterales para aportar el máximo valor a los debates sobre la gobernanza mundial de IA.

En este sentido, el enfoque de la regulación de la IA en Estados Unidos puede contrastarse con la propuesta de Reglamento de la IA de la Unión Europea, debido a que Estados Unidos pretende adoptar un enfoque normativo más flexible y menos prescriptivo, orientado en muchas ocasiones a una autoregulación a nivel sectorial, articulado a través de la definición de una serie de objetivos a alto nivel, con el objetivo de que se adapten de una manera sencilla e iterativa a medida que la tecnología se desarrolla y los riesgos asociados a la misma evolucionan.

Al igual que sucede con otras normas, el reto para muchas entidades de ámbito multinacional consistirá en desarrollar y utilizar un enfoque de cumplimiento en múltiples jurisdicciones que satisfaga estos estándares regulatorios divergentes. A un nivel práctico, es probable que el enfoque de Estados Unidos y de otros países anglosajones, como el Reino Unido, se considere en las entidades como un nivel de referencia de las obligaciones regulatorias de la IA, que puede ser lo suficientemente amplio como para tener una relevancia global, mientras que es probable que se considere que el enfoque de la UE requiere de unos estándares de cumplimiento dentro de las empresas significativamente más altos.

1.4.2. CHINA

China acaba de presentar su proyecto de ley para regular la investigación y desarrollo de la inteligencia artificial que, junto con la creación de un comité de expertos, permitiría establecer un sistema regulatorio que se aplicará a "la investigación, el desarrollo y la utilización de productos de inteligencia artificial generativa para prestar servicios al público dentro del territorio de la República Popular China." Dicha Ley será pieza clave en la culminación del "Plan de desarrollo de inteligencia artificial de nueva generación " presentado en julio de 2017 por el Consejo de Estado de China.

1.4.3. OTRAS REGIONES

La Agencia de Asuntos Culturales y la Oficina del Gabinete de Japón presentó en mayo de 2023 a través de un documento llamado "Regarding the relationship between AI and copyright ", su postura respecto al aprendizaje por IA generativa y la relación entre derechos de autor sobre productos y obras existentes, y se prevé que en breve presente una propuesta legislativa, aunque todo parece indicar que será más flexible que la futura ley europea.

Recientemente, hace un mes, Panamá ha presentado su primera iniciativa legislativa para regular la IA (Anteproyecto de ley 014) donde, entre otros, se incluye la creación de una agencia de supervisión, así como la prohibición de un mal uso de esta tecnología. Perú por su lado, ha promulgado hace una semana la ley N° 31814, que promueve el uso de la inteligencia artificial en favor del desarrollo económico y social del país.

2 LA DIMENSIÓN ÉTICA DE LA INTELIGENCIA ARTIFICIAL

2.1. ALCANCE Y OBJETIVOS

La inteligencia artificial fue definida en el Libro Blanco sobre IA - un enfoque europeo orientado a la excelencia y la confianza- como: "una tecnología estratégica que ofrece numerosas ventajas a los ciudadanos, las empresas y la sociedad en su conjunto, siempre y cuando sea antropocéntrica, ética y sostenible y respete los derechos y valores fundamentales". La IA aporta importantes mejoras de la eficiencia y la productividad y puede contribuir, además, a encontrar soluciones a la lucha contra el cambio climático y la degradación medioambiental, los retos relacionados con la sostenibilidad y los cambios demográficos, entre otros. No obstante, es importante desarrollar sistemas de IA merecedores de confianza, y por este motivo se debe garantizar que los riesgos y otros efectos adversos asociados a la aplicación de la IA se gestionen de manera adecuada y proporcionada.

Con este objetivo de crear una IA fiable, se articulan a continuación los derechos fundamentales y un conjunto de principios éticos asociados que resultan cruciales en el contexto de la IA. En este sentido, en noviembre de 2021 los 193 Estados miembros de la Conferencia General de la UNESCO adoptaron la Recomendación sobre la Ética de la Inteligencia Artificial , este fue el primer instrumento normativo mundial sobre el tema. Estas recomendaciones no sólo protegen, sino que también promueven los derechos y dignidad humana y se constituyen como una brújula o guía ética y una base normativa global que permitirá construir un sólido respeto por el estado de derecho en el mundo digital.

Las recomendaciones de la UNESCO van en encaminadas a considerar la ética como una base dinámica para la evaluación de otras normativas y una orientación en el uso de tecnologías basadas en Inteligencia Artificial debido a que la dignidad humana, el bienestar y evitar los posibles daños a las personas son la orientación y la raíz de la ciencia y la tecnología.

Los sistemas de IA tienen la capacidad de procesar datos e información de una manera que se asemeja a un comportamiento inteligente y, por lo general, incluye aspectos de razonamiento, aprendizaje, percepción, predicción, planificación o control.

La UNESCO ha determinado cuales deben ser los tres elementos centrales de este enfoque:

I Los sistemas de IA son tecnologías de procesamiento de información que integran modelos y algoritmos que producen la capacidad de aprender y realizar tareas cognitivas que conducen a resultados como la predicción y la toma de decisiones en entornos materiales y virtuales. Actúan con distintos grados de autonomía a través del modelado, la representación del conocimiento, la explotación de datos y el cálculo de correlaciones. Los métodos utilizados en la IA son muy variados: machine learning, machine reasoning, Internet de las cosas, sistemas robóticos, robótica social e interfaces hombre-computadora.

II Las cuestiones éticas se refieren a todas las etapas del ciclo de vida de una solución de IA, desde su concepción hasta su retirada final. En muchas ocasiones, un componente de IA no estará aislado y se integrará en un sistema específico junto a otros componentes. Las personas que intervienen en el proceso de la IA pueden intervenir en cualquier momento del ciclo de vida de una solución de IA, y pueden ser personas tanto físicas como jurídicas.

III Los sistemas de IA plantean nuevos problemas éticos en los siguientes campos:

- Impacto en la toma de decisiones.
- El empleo y el trabajo.
- La interacción social.
- La atención médica .
- La educación.
- Los medios y el acceso a la información.
- La brecha digital.
- Los datos personales y protección del consumidor.
- Medio ambiente.
- Democracia.
- Derechos humanos y libertades fundamentales.

Los algoritmos utilizados en las soluciones de IA generan nuevos desafíos éticos que producen y refuerzan los errores y sesgos, algunos de ellos no intencionales, existentes en la sociedad que pueden afectar a la equidad en la implementación del modelo de IA, de manera especial cuando afecta negativamente a determinados grupos.

La mayor parte de estos problemas tienen su origen en la capacidad de los sistemas de IA para realizar tareas que antes solo podían hacer los seres humanos, atribuyendo un papel relevante a la IA en la creación de un nuevo contexto para que seres humanos comprendan y acepten esta tecnología con un espíritu crítico, de manera que no asuman un "principio de autoridad" derivado de las expectativas creadas por los sistemas de IA e interpreten siempre sus resultados como ciertos e inamovibles .

En este mismo sentido, las recomendaciones planteadas por la UNESCO prestan una especial atención a las implicaciones éticas de los sistemas de IA con relación a los siguientes dominios:

- **Educación** – Se requiere de nuevas prácticas educativas, reflexión ética, pensamiento crítico, prácticas de diseño responsable y nuevas habilidades

- **Ciencias** - Incluye todos los campos académicos de las ciencias naturales, ciencias médicas, ciencias sociales y humanidades, ya que las tecnologías de IA aportan nuevas capacidades y enfoques de investigación, tienen implicaciones para nuestros conceptos de comprensión y explicación científica, y crean una nueva base para la toma de decisiones.

- **Identidad y diversidad cultural** - Las tecnologías de IA pueden enriquecer las industrias culturales y creativas, pero también pueden conducir a una mayor concentración de la oferta de contenido cultural, datos, mercados e ingresos en manos de unos pocos.

- **Comunicación e información** - Las tecnologías de IA desempeñan un papel cada vez más importante en el procesamiento estructuración y provisión de información. Esto puede acarrear problemas como la desinformación, la información errónea, el discurso de odio, el surgimiento de nuevas formas de narrativas, discriminación, libertad de expresión, privacidad y alfabetización mediática e informacional.

Las recomendaciones realizadas por la UNESCO se dirigen a todos los Estados miembros, actores de la IA, autoridades responsables de desarrollar marcos legales y regulatorios y de promover la responsabilidad empresarial y ofrece, por lo tanto, una orientación ética a todos los actores de la IA, incluidos los sectores público y privado, al proporcionar una base para una evaluación del impacto ético de los sistemas de IA a lo largo de todo su ciclo de vida.

En idéntico sentido, la Comisión Europea ha considerado que la fiabilidad de la IA se apoya en tres componentes que deben satisfacerse de manera simultánea a lo largo de todo el ciclo de vida del sistema: (i)

~ La IA debe ser lícita, de modo que se garantice el respeto de todas las leyes y reglamentos aplicables.

~ La IA ha de ser ética, es decir, asegurar el cumplimiento de los principios y valores éticos.

~ La IA debe ser robusta, tanto desde el punto de vista técnico como social, puesto que los sistemas de IA, incluso si las intenciones son buenas, pueden provocar daños accidentales.

La construcción de una IA fiable a lo largo de todo el ciclo de vida del sistema

Las recomendaciones éticas para la IA de la UNESCO tienen como objetivo principal proporcionar un instrumento normativo que sea aceptado a nivel mundial que no sólo articule valores y principios, sino que además pueda ser llevada a la práctica aplicando políticas concretas. Estas políticas deberán hacer un énfasis especial en la igualdad de género y la protección del medio ambiente.

Los objetivos específicos que se desarrollan en esta visión ética de la UNESCO se articulan en las siguientes recomendaciones:

- Suministrar un marco universal de valores, principios y acciones para guiar a los Estados de la UNESCO a la hora de desarrollar su legislación y sus políticas relacionados con la IA, de conformidad con el derecho internacional.
- Orientar las acciones de los individuos, grupos, comunidades, instituciones y empresas del sector privado para garantizar la incorporación de la ética en todas las etapas del ciclo de vida del sistema de IA.
- Proteger, promover y respetar los derechos humanos y las libertades fundamentales, la dignidad humana y la igualdad; salvaguardar los intereses de las generaciones presentes y futuras; preservar el medio ambiente, la biodiversidad y los ecosistemas y respetar la diversidad cultural en todas las etapas del ciclo de vida del sistema de IA.
- Promover el diálogo y la creación de consenso entre múltiples partes interesadas, multidisciplinares y plurales sobre cuestiones éticas relacionadas con los sistemas de IA.
- Promover el acceso equitativo a los avances y conocimientos en el campo de la IA y la distribución de beneficios, con especial atención a las necesidades y contribuciones de los países de bajos ingresos.

De manera complementaria, la Estrategia Europea de Datos tiene por objeto ayudar a la Unión Europea en convertirse en la economía con agilidad en el manejo de los datos más atractiva, segura y dinámica del mundo, lo que fortalecerá a la UE con información para reforzar sus decisiones y mejorar las vidas de todos sus ciudadanos.

Las directrices como marco para una IA fiable

A modo de conclusión, la ética de la IA persigue proteger una serie de valores del individuo frente a un razonamiento exclusivamente mecánico que puede afectar a su libertad, igualdad, dignidad e, incluso, autonomía de manera que la perspectiva ética de la IA se constituye como una de las principales preocupaciones y, aunque los valores éticos de las personas dependen en gran medida de cuestiones culturales, es importante destacar la existencia de unos principios de IA, integrados como una parte de la "ética digital", que se han aceptado de forma global.

A continuación, se definen los principios fundamentales de la ética que deberán tener en cuentas las organizaciones en la utilización y desarrollo de los sistemas de IA. Estos principios suministran la base para abordar los principales desafíos éticos relacionados con esta tecnología y garantizan un enfoque responsable en su implantación.

Dentro de estos principios encontramos los siguientes:

El funcionamiento de los sistemas de IA debe ser abierto y claro. Los algoritmos y modelos de IA deben ser comprensibles y explicables, de manera que las decisiones tomadas por estos sistemas puedan ser comprendidas y evaluadas por los usuarios y las partes afectadas. Esta transparencia implica proporcionar información clara sobre cómo se utilizan los datos y cómo se toman las decisiones.

Se debe garantizar que los beneficios y los riesgos de la IA se distribuyan de manera equitativa en la sociedad. Debemos evitar la discriminación y asegurarnos de que los sistemas de IA fomenten sesgos o desigualdades existentes. En el desarrollo de un sistema de inteligencia artificial debemos tener en cuenta la diversidad de datos y perspectivas y abordar posibles sesgos en los conjuntos de datos utilizados.

Este concepto se refiere a la protección de los datos personales y la información confidencial en el contexto de la IA. Es imprescindible que los sistemas de IA respeten la privacidad de los individuos y cumplan con las regulaciones y políticas de protección de datos. Debemos obtener el consentimiento informado de los usuarios y garantizar la seguridad y confidencialidad de los datos que usamos.

Debemos considerar los valores éticos desde las etapas iniciales del diseño de los sistemas de IA. Esto implica que debemos tener en cuenta las implicaciones éticas y sociales de las decisiones de diseño, y asegurarse de que los sistemas sean desarrollados de acuerdo con principios éticos aceptables.

RESPONSABILIDAD

Este concepto está relacionado con la atribución de la responsabilidad por las decisiones y acciones de los sistemas de IA. Debemos fijar mecanismos para determinar quién es responsable en caso de daños o consecuencias negativas causadas por la IA. Se deben, por tanto, desarrollar marcos legales y éticos que definan las responsabilidades de los desarrolladores, proveedores y usuarios de sistemas de IA.

IMPACTO SOCIAL

La IA puede tener un impacto significativo en la sociedad, tanto positivo como negativo. Hay que considerar las implicaciones sociales, económicas y culturales de la IA y garantizar que se utilice de manera responsable y en beneficio de la sociedad en su conjunto. Se hace necesario realizar evaluaciones de impacto ético y social antes de implementar sistemas de IA y considerar las posibles consecuencias a largo plazo.

AUTONOMÍA HUMANA

Los sistemas de IA no deberían subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos de manera injustificada. Esto implica garantizar la supervisión y el control humanos sobre los procesos de trabajo de los sistemas de IA.

PREVENCIÓN DEL DAÑO

Se debe fomentar en las organizaciones el uso responsable de la IA, antes de implantar un sistema con IA se deben realizar las pruebas correspondientes para validar que funcione de forma segura, lo cual supone validar que cumple con los estándares de calidad y seguridad establecidos.

En similares términos se ha pronunciado la Unión Europea a través de diferentes documentos, entre otros, las Directrices para una IA fiable del grupo de expertos de la UE que acogía favorablemente los siete requisitos esenciales de la IA:

- Acción y supervisión humanas.
- Solidez técnica y seguridad.
- Gestión de la privacidad y de los datos.
- Transparencia.
- Diversidad, no discriminación y equidad.
- Bienestar social y medioambiental.
- Rendición de cuentas .

Interrelaciones existentes entre los siete requisitos: todos tienen idéntica importancia, se apoyan entre sí

Además de este conjunto de directrices no vinculantes, se destacó la importancia de acelerar la creación de un marco regulador claro para la UE que debe ser eficaz para alcanzar sus objetivos sin ser excesivamente prescriptivo, lo que podría generar una carga desproporcionada, en especial para las pymes. Para alcanzar este equilibrio, la Comisión europea consideró que debe seguir un enfoque basado en el riesgo que resulte de aplicación a los productos y servicios basados en la IA. El objetivo último consiste en la generación de confianza entre los consumidores y las empresas y contribuir a la creación de un mercado interior sin fricciones de cara al desarrollo y adopción futura de la IA.

2.2. INTEGRACIÓN DE LOS VALORES ÉTICOS EN LA IA DE LA ORGANIZACIÓN

2.2.1. INTRODUCCIÓN

Es aconsejable que las entidades consideren la integración de los valores éticos de esta tecnología en sentido amplio, de conformidad con los criterios que ha establecido la UNESCO en su documento Recomendación sobre la Ética de la Inteligencia Artificial, circunstancia que llevará aparejados una serie de beneficios:

- Para la Humanidad, para los individuos y la sociedad.
- Para el medioambiente y el ecosistema
- Para la prevención de posibles daños

Por este motivo, es necesario considerar la integración de los valores éticos de la IA en la organización, así como su materialización efectiva a través de políticas de obligado cumplimiento interno, en las que se enfatice la necesidad de reducir lo máximo posible potenciales diferencias de género o igualdad, entre otros, así como de maximizar la protección del medioambiente y los diferentes ecosistemas en los que puedan operar las organizaciones.

Este eje central podría definirse como la protección, promoción y respeto de los derechos humanos y las libertades fundamentales en un plano tanto individual como en el conjunto de la sociedad.

Una de las medidas que pueden llevar a cabo las organizaciones consisten en la revisión de los códigos de conducta internos a fin de incluir una política específica de IA o, en su caso, añadirla a alguna de las políticas ya existentes como un documento complementario, con el objetivo de dotarla de este modo del necesario apoyo organizativo.

2.2.2. ESTRATEGIA ÉTICA DE LA IA

A medida que las organizaciones recurren cada vez más a la IA para innovar, mejorar el rendimiento de sus procesos, reducir costes organizativos y/o mitigar riesgos, es importante que comprendan los requisitos éticos fundamentales para implantar con éxito una solución de IA. La IA es una tecnología que no funciona de manera independiente de otros sistemas o tecnologías empresariales, sino que depende de factores como la estrategia de la organización, las personas, los datos, la tecnología y la integración del ecosistema. Para garantizar que la organización aprovecha esta tecnología de forma ética, eficaz y responsable, debe tener en cuenta cada uno de estos componentes.

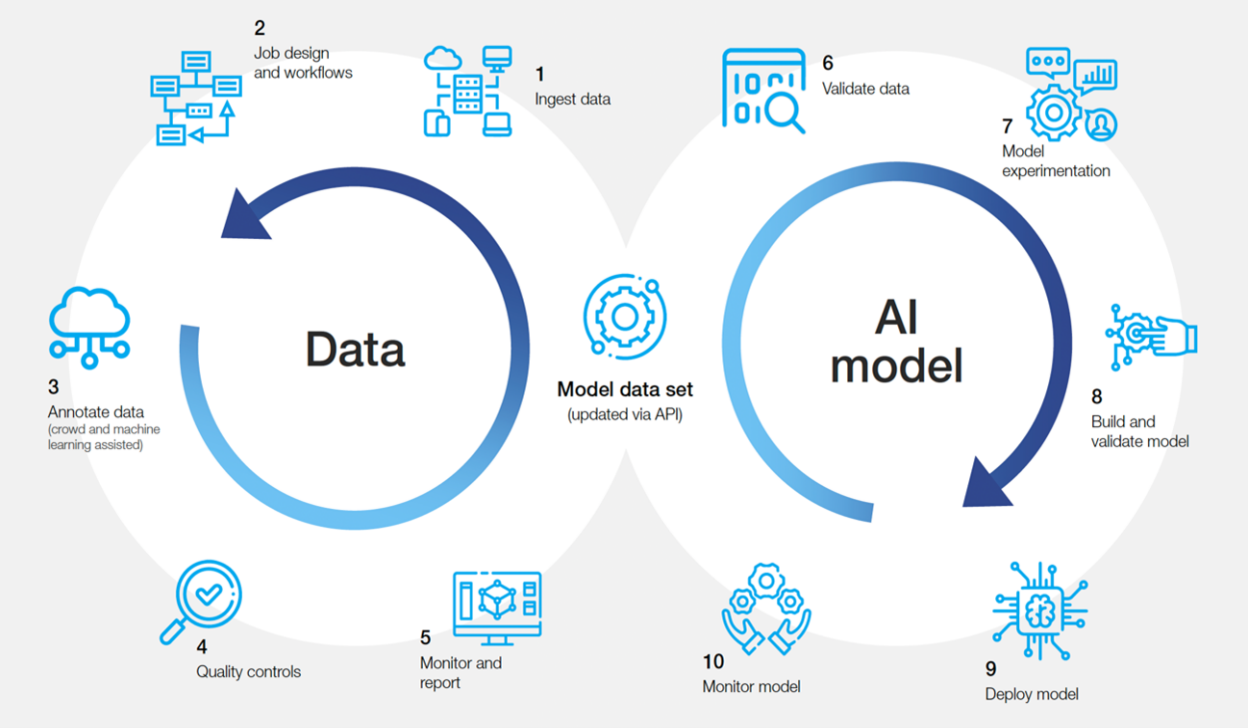
Por lo anteriores motivos, una estrategia ética de IA exitosa debe definir claramente una visión para la organización y cómo esa visión se despliega en objetivos específicos y medibles. En este sentido, será labor de la Alta Dirección definir claramente una visión ética para su organización e identificar cómo la IA puede apoyar la visión y la estrategia corporativa.

Desde una perspectiva ética, será de enorme relevancia **desarrollar una hoja de ruta ética desde el diseño** ("**Ethics by design**") que desemboque en políticas corporativas adecuadas, conectando así con los criterios ambientales, sociales y de gobernanza, así como **en la creación de "Políticas de IA Confiable y Responsable"** y "Hojas de Ruta de Cumplimiento Proactivo" como medidas clave para **mejorar la gestión de los riesgos legales** y éticos asociados a la inteligencia artificial.

A menudo, las entidades etiquetan erróneamente como IA iniciativas que, en puridad, no son realmente una solución de IA, sino una simple decisión automatizada, dejándose llevar por cuestiones de marketing.

En cualquier caso, centrándonos en las soluciones que sí reúnen los requisitos para ser consideradas tecnologías de IA de conformidad con la definición que establece la propuesta de Reglamento europeo, resultará necesaria la realización de un análisis previo relativo a si estas soluciones de IA tienen, en realidad, el potencial de ayudar a cumplir los principales objetivos de la organización.

Así, para la construcción de la estrategia de IA es clave diseñar previamente la estrategia de datos, circunstancia que comprende la adquisición, gestión y tratamiento de la información en una organización. Un buen punto de partida podría ser la evaluación de la disponibilidad de datos internos, así como, entender qué datos necesita la empresa y qué datos genera, lo que permitirá definir los diferentes casos de uso específicos de la IA. En definitiva, resultará necesario adoptar un enfoque estratégico de los datos, que permita a la entidad invertir en el tratamiento, depuración, actualización, conservación, protección y supervisión de los datos, en particular, teniendo en cuenta la alta actividad regulatoria en relación con la normativa de protección de datos y la inminente aprobación del Reglamento de IA en la UE.



Appen, "Technology, Single Data Collection and Annotation Pipeline", 2021

2.3. COMUNICACIÓN Y PARTICIPACIÓN CON TODAS LAS PARTES INTERESADAS

Una integración real de los valores éticos en el uso o desarrollo de IA en una organización ha de permitir la participación y comunicación por las partes interesadas, y para que ésta sea efectiva es necesario tener en cuenta los siguientes aspectos:

2.3.1. TRANSPARENCIA E INFORMACIÓN SIGNIFICATIVA SOBRE LA LÓGICA APLICADA

A fin de alcanzar los objetivos perseguidos, en particular la promoción del uso responsable de la IA, la creación de confianza y las garantías de reparación cuando proceda, resulta importante que se facilite información adecuada de manera proactiva en torno a cómo usar los sistemas de IA, de manera especial en aquellos que presentan un elevado riesgo.

Es importante también que la información facilitada sea objetiva, concisa y fácilmente comprensible. La manera en que ha de presentarse la información debe adaptarse al contexto específico, y aunque la legislación de protección de datos de la UE ya recoge algunas normas de este tipo (*artículo 13, apartado 2, letra f) del RGPD*), la propuesta de normativa introduce algunos requisitos adicionales para alcanzar los objetivos anteriormente mencionados

En este sentido, el cumplimiento del principio de transparencia del art.5 RGPD exige facilitar a los interesados la información relativa al tratamiento de manera concisa, transparente e inteligible, en lenguaje claro y sencillo. Dentro de esos deberes de información se incluye en este caso las propiedades, lógica y consecuencias derivadas de su uso . Esta información no sólo favorece la confianza de los usuarios en el uso de las tecnológicas, al facilitar un mayor control de sobre el uso de su información, sino que además forma parte de un tratamiento de datos ético, al garantizar que se adoptan las medidas adecuadas para proteger la privacidad y derechos de los usuarios.

A la vista de todo lo anterior, la información que cada entidad que utilice un sistema de IA deba proporcionar a los interesados y el contenido concreto se tendrá que adaptar a cada etapa del ciclo de vida de la IA y, en el supuesto de que se estén utilizando datos de carácter personal, se podrá ofrecer esta información mediante una aproximación por capas o niveles: una primera capa, de carácter general, con información básica del tratamiento y una segunda capa que completa la información de la primera con mayor nivel de detalle y que sea accesible desde ésta de forma fácil e inmediata, incluso por medios electrónicos. La información que se facilite deberá ser relevante a fin de permitir al interesado entender el comportamiento del tipo de componente IA utilizado y proporcionarle certeza y confianza sobre sus resultados. El ofrecimiento de información con referencias excesivamente técnicas sobre el algoritmo puede resultar opaco, confuso, e incluso conducir a la fatiga informativa.

2.3.2. EJERCICIO DE DERECHOS E INTERVENCIÓN HUMANA

Cuando las entidades hagan uso de soluciones de IA que impliquen un tratamiento de datos personales, resultará necesario que los interesados dispongan de medios y mecanismos accesibles para el ejercicio de sus derechos y, en su caso, que éstos les permitan solicitar intervención humana en la evaluación o quieran impugnar los resultados derivados del uso de esta tecnología, y por los cuales puedan verse afectados. Esta intervención humana garantiza una supervisión y control adecuado que permita evitar sesgos, discriminación y/o la adopción de resultados perjudiciales para los interesados, así como la detección de posibles defectos en el diseño general de los sistemas de IA, al mismo tiempo que como en el caso anterior, aumenta la confianza y aceptación por los usuarios, reduciendo los riesgos de una automatización completa, para lograr un equilibrio adecuado.

En este sentido, se debe garantizar que la acción y supervisión humana de la decisión sea realizada por una persona competente y autorizada para modificar la decisión, y para ello ha de realizar una acción significativa y no simbólica, desde la fase de diseño y a lo largo de todo el ciclo de vida de los productos y sistemas de IA.

Cada entidad deberá analizar el sistema de IA para, en función del análisis de los riesgos que pueda tener para las personas afectadas, determinar el nivel de supervisión humana que se deberá implantar en cada caso.

3.4. ELABORACIÓN DE UN CÓDIGO ÉTICO

La elaboración de un código ético interno en las organizaciones permite establecer principios y valores para regular el uso de la inteligencia artificial a través de claridad y orientación en las pautas a adoptar en el uso y desarrollo, entre otros, de esta tecnología, de manera que no transgreda los derechos y libertades fundamentales de los individuos, y permitiendo en todo caso, el cumplimiento regulatorio de la normativa aplicable.

Como se ha indicado previamente, el primer marco ético sobre inteligencia artificial ha sido adoptado por los 193 Estados miembros de la Unesco en noviembre de 2021, siendo de esta manera el primer instrumento normativo que servirá, tal y como se enuncia en el texto, como "brújula guía ética y una base normativa global que permitirá construir un sólido respeto por el estado de derecho en el mundo digital"

En diciembre de 2022, se publicó "TTC Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management " que presentaba la declaración realizada entre UE y Estados Unidos en donde se comprometían a la elaboración conjunta de una hoja de ruta que permitiera orientar el desarrollo de herramientas, metodologías y enfoques para la gestión de riesgos de la IA, manteniendo su compromiso con la Recomendación sobre la IA de la OCDE adoptada en mayo de 2019.

Estos códigos son solo un ejemplo del trabajo que se está llevando a cabo, pero actualmente hay varios grupos de instituciones solicitando a las empresas que desarrollan sistemas de Inteligencia Artificial la adopción de compromisos éticos como, por ejemplo, y tras la recomendación realizada por la UNESCO anteriormente mencionada, Microsoft y UNESC han firmado un Acuerdo de colaboración con el objetivo de promover la ética en el desarrollo de la Inteligencia Artificial. Otro ejemplo lo encontramos en un documento emitido por el Ministerio de Ciencia y Tecnología de China, con el objetivo de llegar a ser líder mundial de IA en el año 2030.

Desde un punto de vista práctico, las entidades pueden valorar la elaboración de un Código Ético con el objetivo de ir más allá del cumplimiento de la normativa aplicable en materia de IA. Un Código Ético no sustituye ni reemplaza el cumplimiento de cualquier obligación legal prevista en materia de IA, o del resto de la normativa que se verá impactada, sino que supone una muestra del compromiso con el cumplimiento de la legalidad, estableciendo un marco de cumplimiento ético interno que vaya más allá del mero cumplimiento regulatorio. De este modo, el tratamiento ético de las soluciones de IA permitirá la mejora de los procedimientos internos de las entidades, completando y mejorando el cumplimiento de la normativa de IA.

En la práctica, muchas organizaciones están elaborando códigos éticos de la IA con el objetivo de establecer un marco normativo sólido y coherente que promueva una tecnología de IA al servicio de los interesados, armonizando los derechos fundamentales de los afectados, con el objetivo de generar un entorno confiable para el desarrollo de esta tecnología, y asumiendo unas garantías adicionales en aquellos casos en los que la IA entrañe un alto riesgo para los derechos y libertades de los interesados.

ANEXO. ENFOQUE EUROPEO DE LA IA

PRINCIPALES HITOS

Según se ha indicado con anterioridad, el enfoque de la UE con respecto a la IA se centra en la excelencia y la confianza a través de una serie de normas y acciones concretas, que tienen como objetivo reforzar el potencial de Europa para competir a nivel mundial.

Para ello, se deberá garantizar que la IA esté centrada en el ser humano y sea fiable, lo que se traduce en la creación de una estrategia que impulse la investigación y la capacidad industrial, garantizando al mismo tiempo la seguridad y los derechos fundamentales de los ciudadanos.

A continuación, se incorporan todos los documentos publicados por la UE dentro del paquete de regulación de la IA hasta junio de 2023.

ABRIL 2019

Comunicación de la Comisión Europea: Crear confianza en la inteligencia artificial centrada en el ser humano.
Grupo de expertos de alto nivel sobre IA: Directrices éticas para una IA confiable.

DICIEMBRE 2018

Comisión Europea: Plan coordinado sobre IA.
Comisión Europea (Comunicado de prensa): IA fabricada en Europa.
Comunicación de la Comisión Europea: IA fabricada en Europa.
Consulta a las partes interesadas sobre el proyecto de directrices éticas para una IA fiable.

JUNIO 2018

Lanzamiento de la alianza europea de IA.
Creación del grupo de expertos de alto nivel sobre IA.

ABRIL 2018

Comunicado de prensa: Inteligencia artificial para Europa.
Comunicación: Inteligencia artificial para Europa.
Documento de trabajo de los servicios de la Comisión: Responsabilidad por las tecnologías digitales emergentes.
Declaración de cooperación en materia de inteligencia artificial.

MARZO 2018

Comunicado de prensa: Grupo de expertos en IA y alianza europea de IA.


OCTUBRE 2020
2.ª Asamblea Europea de la Alianza IA.
JULIO 2020
Evaluación de impacto inicial: Requisitos éticos y legales en materia de IA.
Grupo de expertos de alto nivel sobre IA: Lista final de evaluación de la IA fiable (ALTAI).
Grupo de expertos de alto nivel sobre IA: Recomendaciones sectoriales de una IA fiable.
FEBRERO 2020
Comisión Europea: Libro Blanco sobre IA: un enfoque europeo de la excelencia y la confianza.
Consulta pública sobre un enfoque europeo de la excelencia y la confianza en la IA.
DICIEMBRE 2019
Grupo de expertos de alto nivel sobre IA: Puesta a prueba de la lista de evaluación de la IA fiable.
JUNIO 2019
Primera Asamblea Europea de la Alianza IA.
Grupo de expertos de alto nivel sobre IA: Recomendaciones políticas y de inversión de la IA.
ABRIL 2019
Comunicación de la Comisión Europea: Crear confianza. en la inteligencia artificial centrada en el ser humano.
Grupo de expertos de alto nivel sobre IA: Directrices éticas para una IA confiable.

Entrando someramente en el detalle de los textos señalados, cabe destacar, y por ello profundizaremos en otros apartados de este documento, las Directrices éticas para una IA fiable articulan un marco para lograr una IA fiable (licita, ética y robusta), directrices que sin detenerse en el primer aspecto (licitud), -como veremos- identifican los principios éticos y valores conexos que deberían respetarse en el desarrollo, despliegue y utilización de los sistemas IA fiables. Estas directrices ofrecen una orientación muy práctica sobre cómo alcanzar esa IA fiable a través del cumplimiento de 7 requisitos (Acción y supervisión humana, solidez técnica y seguridad, gestión de la privacidad de los datos, transparencia, diversidad, no discriminación y equidad, bienestar social y ambiental y rendición de cuentas) basados en 4 grandes fundamentos (respeto a de la autonomía humana, prevención del daño, equidad y explicabilidad).

El segundo gran hito que merece la pena de los señalados anteriormente es el Libro Blanco sobre Inteligencia Artificial (febrero de 2020) en el cual se abordaba la necesidad de establecer i) el marco político de la IA (a través de la creación de un ecosistema de excelencia en la UE para el desarrollo y la adopción de la IA), ii) el marco regulador (a través de creación de un ecosistema de confianza que evite la fragmentación y que aborde la IA desde un enfoque a riesgos para los derechos y libertades de los ciudadanos), iii) una definición de IA suficientemente flexible para adaptarse al progreso y que a la vez permita la seguridad jurídica, iv) unos requisitos obligatorios para los sistemas de IA, v) dirigidos a aquellos que estén en el mejor posición de cumplir en los cadena de desarrollo, implementación y uso de los sistemas de IA y vi) una gobernanza europea sobre la IA que, al igual que el marco regulador, huya de la fragmentación de responsabilidades.

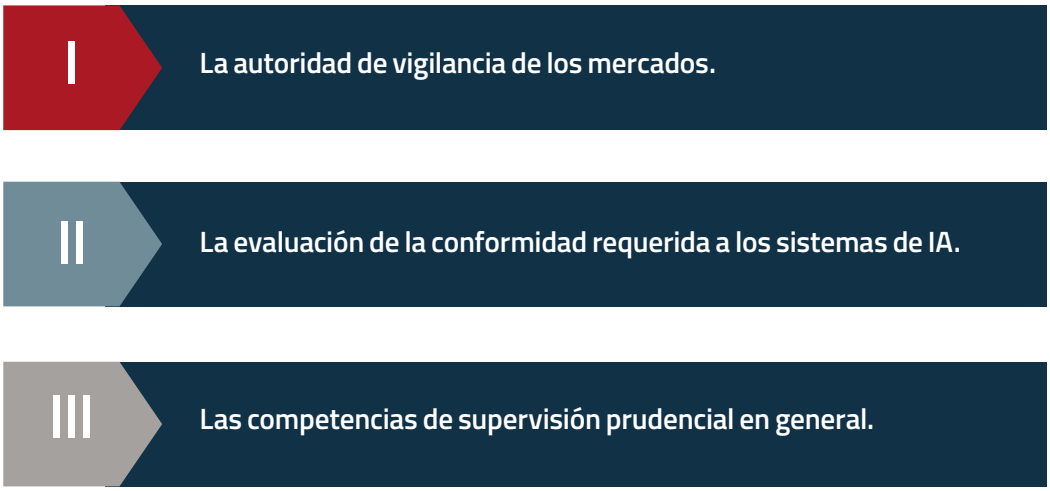
A estos hitos cabría añadir, a nivel español, las dos guías que hasta el momento ha publicado la Agencia Española de Protección de Datos ("AEPD") y dos pronunciamientos sobre la norma que han realizado tres actores muy relevantes: el Banco Central Europeo ("BCX"), el Comité Europeo de Protección de Datos ("CEPD") y el Supervisor Europeo de Protección de Datos ("SEPD").

Necesariamente cuando la utilización de una IA implique un tratamiento de datos personales se deberá tener en cuenta las dos guías de la AEPD. La primera guía "Adecuación al RGPD de tratamientos que incorporan IA" recurre a la definición de IA realizada por el grupo de Expertos en las Directrices para una IA fiable que mencionábamos con anterioridad y, haciendo un repasado de las principales obligaciones derivadas del RGPD, establece cómo, ante la existencia de algoritmos en los tratamientos de datos personales, deben cumplirse no solo los principios generales del artículo 5 del RGPD sino, además, como deben gestionarse los derechos reconocidos a los titulares de los datos, cómo debe procederse en la realización de las correspondientes Evaluaciones de impacto en la privacidad o cómo se deberá analizar el cumplimiento de las previsiones relativas a la elaboración de perfiles y la toma de decisiones basadas únicamente en un tratamiento automatizado, debiendo evitar el diseño de sistemas con la orientación "Dead man switch" y dar siempre la opción de que un operador humano pueda ignorar el algoritmo en un momento dado.

En 2021 se publica la segunda guía de la AEPD, "Requisitos para auditorias de tratamientos que incluyen IA" en la que se establecen 144 controles agrupados en 5 grandes materias (identificación y transparencia del componente, propósito del componente, fundamentos del componente, gestión de los datos y verificación y validación) que se perfilan como una guía muy útil a disposición de responsables y encargados del tratamiento para, a través de la supervisión de los tratamientos, se puedan determinar los planes de acción necesarios para que, por defecto y desde el diseño, los tratamientos de datos que incluyan inteligencia artificial cumplan con el RGPD.

Acabamos esta sección, como avanzábamos, con referencias a dos pronunciamientos: el Dictamen "BCE" y el Dictamen conjunto CEPD-SEPD sobre la propuesta de Reglamento de IA.

En relación con el primero, señalamos su importancia debido a que la propuesta de Reglamento de IA contiene disposiciones que afectan a las competencias del BCE, en particular a sus funciones relacionados con la supervisión prudencial de las Entidades de crédito que, como sabemos, se verán impactadas al encontrarse los sistemas destinados a evaluar la solvencia de las personas físicas o a establecer su calificación crediticia entre los sistemas de IA de alto riesgo. En este sentido, el BCE ha solicitado clarificación de sus competencias en el marco de la propuesta de Reglamento de IA en 3 aspectos:



Hay que destacar finalmente de este Dictamen que el BCE admite la necesaria supervisión del SEPD en relación con el uso del supervisor bancarios de sistemas de IA y hace una llamada a no considerar IA los sistemas que "aprovechan el uso independiente de la regresión lineal o logística o los diagramas de decisiones bajo supervisión humana siempre que el efecto de dichos enfoques aplicados a la evaluación de la solvencia o la calificación crediticia de las personas físicas no sea significativo".

En relación con el segundo pronunciamiento, el Dictamen conjunto del CEPD y SEPD destacan sus llamadas a garantizar la claridad de:

I. La relación de la propuesta de Reglamento de IA con la legislación vigente en materia de protección destacando, entre otros, si el enfoque basado en los factores de riegos de la propuesta y el concepto de "riesgo para los derechos fundamentales" se ajusta al previsto en el RGPD, o bien si la interpretación relativa a si un tipo de tratamiento puede dar lugar a un riesgo elevado de conformidad con el RGPD, debe hacerse con independencia de la propuesta de Reglamento de IA, mientras que la clasificación de un sistema de IA como «de alto riesgo» debido a su impacto en los derechos fundamentales debido a la mencionado propuesta daría lugar a una presunción de «alto riesgo» en el marco del RGPD, o la configuración como requisito previo al marcado como producto de la CE del cumplimiento de las obligaciones legales derivadas de su legislación.

II. **Necesidad de constante actualización de las listas** donde se concreten los sistemas de alto riesgo.

III. Necesidad de mejora de la lista de supuestos de prácticas prohibidas para incluir la clasificación social de las personas en general por cualquier actor (público o privado) o la inferencia de emociones de las personas.

IV. La dicotomía entre las obligaciones para los proveedores de sistemas de IA con relación a la evaluación de riesgos cuando en muchos casos los responsables de tratamiento serán los usuarios de los mismos, así como la indisponibilidad del proveedor para evaluar todos los usos de su sistema de IA (lo que parece que propone solucionar a través de la realización de EIDPs por parte de estos).

V. Competencias del SEPD en relación con ser autoridad de vigilancia del mercado y las autoridades de protección de datos nacionales, como autoridades de supervisión para garantizar un enfoque regulador más armonizado.

VI. Ámbito de aplicación y los objetivos de los espacios de pruebas

Ética y Compliance en el uso de la Inteligencia Artificial

Una iniciativa de

isms forum