

Guía de *inteligencia artificial responsable* para áreas de negocio



¿Te has preguntado alguna vez qué impacto tienen en las personas los productos y servicios en los que estás trabajando?



La Inteligencia Artificial (IA) es una tecnología clave para la innovación, y por tanto para nuestro negocio. Actualmente vivimos una época de auge y democratización de la Inteligencia Artificial. Las oportunidades de aplicación de la IA aumentan cada día y, en consecuencia, el espacio e impacto que alcanza la IA en nuestras vidas también es mayor. En Telefónica creemos que esta situación ofrece la posibilidad de aprovechar todos los beneficios de la IA, pero siempre desde un uso comprometido y responsable, que afronte éticamente los desafíos que la IA supone.

La IA puede tener posibles resultados no deseados con impacto negativo en las personas, en la sociedad o el medio ambiente, como, por ejemplo:

- **posibles discriminaciones o sesgos injustos,**
- **falta de transparencia y de interpretabilidad de sus conclusiones,**
- **un nivel insuficiente de privacidad y/o calidad en los datos utilizados,**
- **fallos de seguridad por no considerar aspectos como la robustez de los modelos,**
- **una supervisión humana inadecuada,**
- **posibles efectos nocivos para colectivos vulnerables como los menores de edad,**
- **destrucción de empleo,**
- **difusión de noticias falsas, polarización y fanatismo político,**
- **unas elevadas emisiones de carbono...**



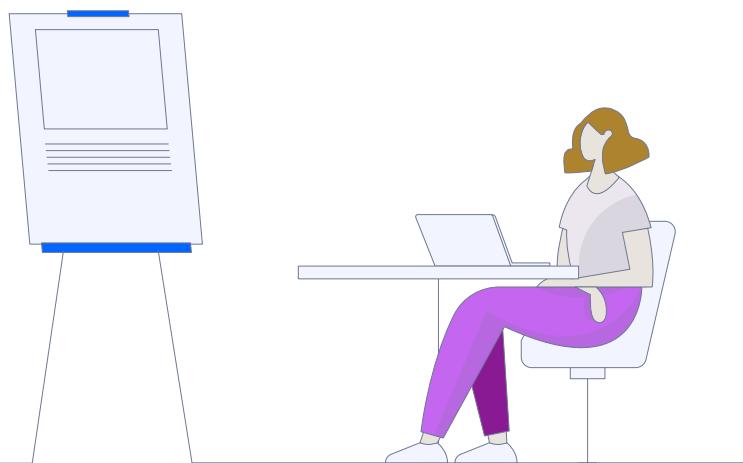
Para evitar estos posibles impactos negativos, debes asegurarte de que los sistemas de IA en los que estás trabajando se diseñan, desarrollan, o se usan de manera responsable, desde un enfoque "by design". Y, por tanto, es necesario que te plantees una serie de reflexiones éticas, como, por ejemplo:

- ¿Es ético el objetivo del producto?**
- ¿Cuál es su grado de impacto social o medioambiental?**
- ¿Es ético el uso de las variables que considera el algoritmo?**
- ¿El caso de uso del sistema está alineado con los valores de Telefónica?**
- ¿Estás documentando el sistema a lo largo de su proceso de desarrollo?, etc.**



En Telefónica tenemos nuestros principios de IA desde 2018. Estos principios son nuestro código de conducta de IA, y tiene como objetivo promover la adopción de una inteligencia artificial centrada en el ser humano y confiable, así como garantizar un alto nivel de protección de la salud, la seguridad, los derechos fundamentales y el medio ambiente frente a potenciales efectos nocivos de la aplicación de la inteligencia artificial.

Todo ello se materializa en un modelo de gobierno que nos permite aprovechar todo el potencial de la IA al tiempo que protegemos a las personas y minimizamos los posibles riesgos. Este modelo de gobierno y código de conducta aplica a todos nuestros productos y servicios desde el diseño y se extiende a nuestros proveedores y socios comerciales. Además, nos va a permitir cumplir con la Ley de Inteligencia Artificial de la UE: el AI Act.



AI Act: La ley de IA de la Unión Europea²

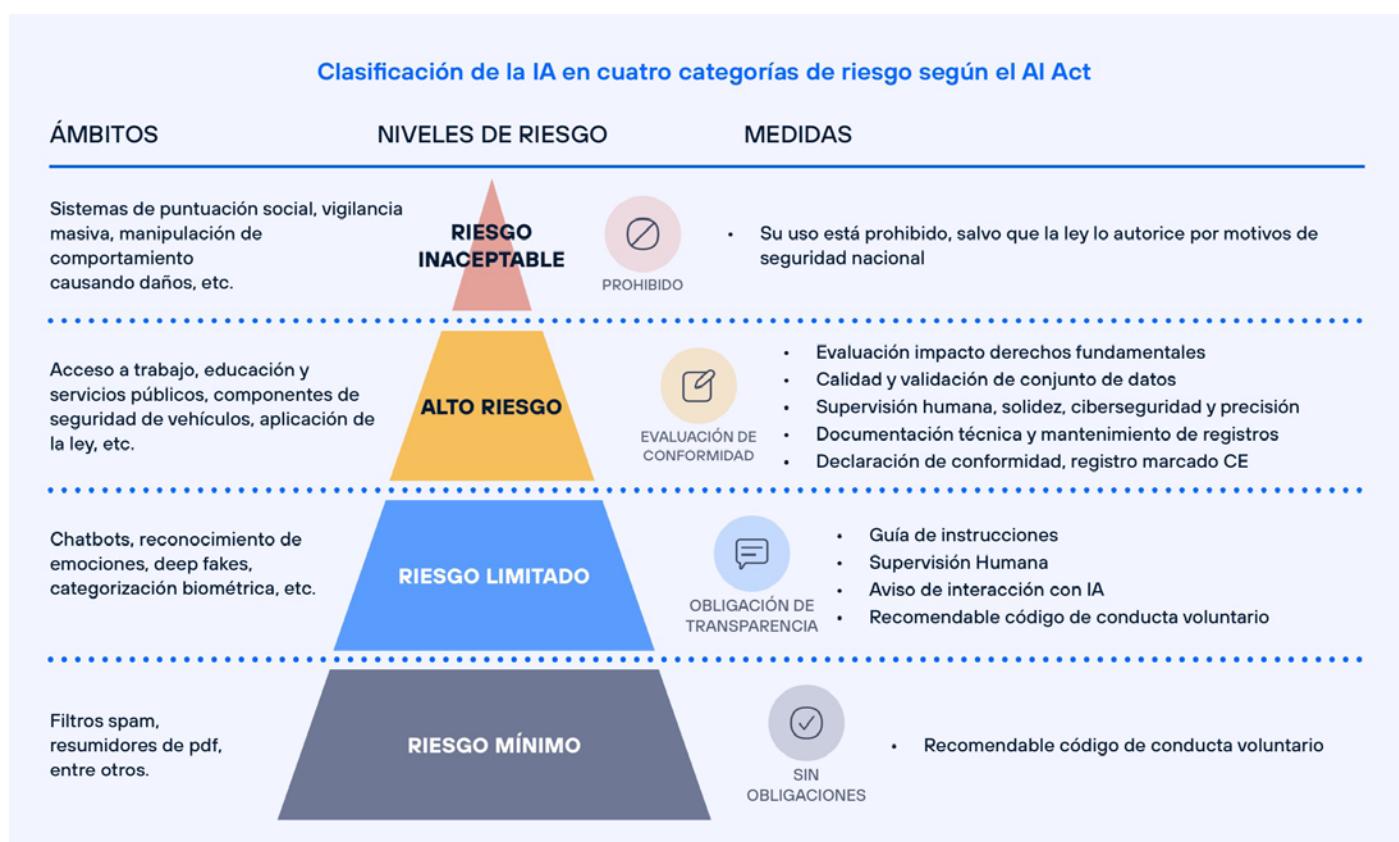
El Reglamento Europeo de Inteligencia artificial (AIA, por Artificial Intelligence Act) es el primer marco normativo integral sobre la materia. Esta norma clasifica el uso de los sistemas IA según el riesgo que supongan para las personas y la sociedad, teniendo en cuenta la función, finalidad y modalidades específicas para las que se utilice dicho sistema. Los distintos niveles de riesgo implicarán una mayor o menor aplicación de medidas regulatorias adecuadas y ajustadas para mitigar el riesgo de manera efectiva.

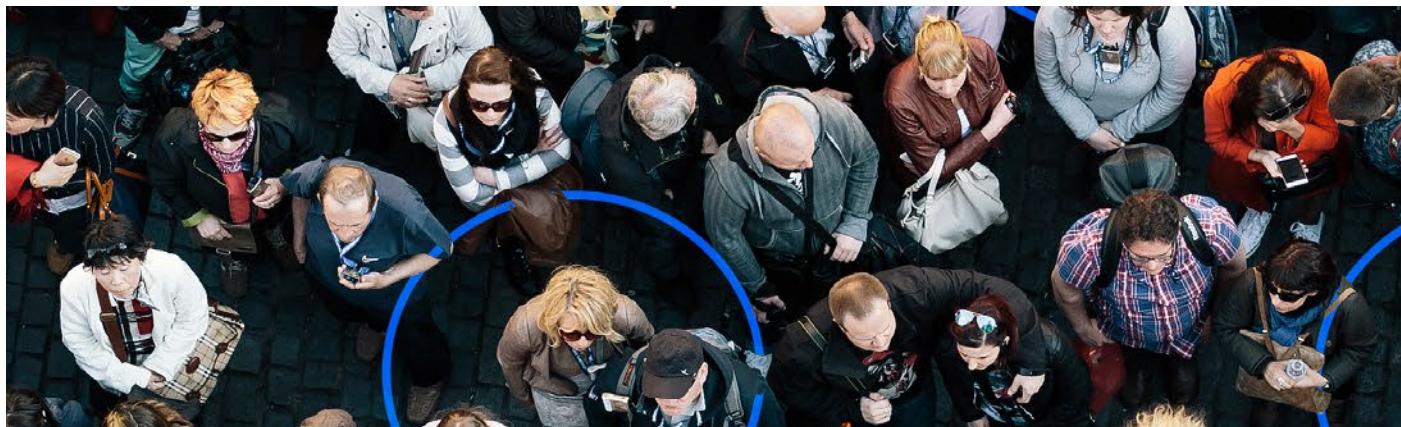
¿A quién aplica?

Tanto a los agentes públicos como privados, de dentro y fuera de la UE, en tanto que el sistema de inteligencia artificial entre en el mercado de la Unión o su uso afecte a personas establecidas en ella.

- Proveedores
- Implementadores
- Importadores
- Proveedores de modelos de inteligencia artificial de uso general, incluidos los grandes modelos generativos de inteligencia artificial

² Inteligencia artificial: preguntas y respuestas* Comisión Europea
https://ec.europa.eu/commission/presscorner/detail/es/qanda_21_1683





¿Cuáles son las categorías de riesgo?

• **Riesgo mínimo:** son aquellos que tienen un impacto mínimo o nulo en los derechos fundamentales de las personas, y que se usan en ámbitos o situaciones poco críticos. Por ejemplo, los sistemas de IA que se usan para el entretenimiento o el ocio.

• **Alto riesgo:** los que se usan en sectores o ámbitos especialmente sensibles, donde un mal funcionamiento o un uso indebido puede causar daños graves a las personas o a la sociedad. Estos sistemas de IA deben cumplir una serie de requisitos exigentes, como la transparencia, la trazabilidad, la supervisión humana, la calidad de los datos, la robustez o la seguridad.

Consulta el anexo para ver las categorías de alto riesgo.

• **Riesgo inadmisible:** Se trata de un conjunto muy limitado de usos especialmente nocivos de la inteligencia artificial que se oponen a los valores de la UE porque vulneran los derechos fundamentales y, en consecuencia, quedarán prohibidos.
Consulta el anexo para ver las categorías de riesgo inadmisible.

• **Riesgo específico para la transparencia:** En el caso de determinados sistemas de inteligencia artificial se imponen obligaciones específicas de transparencia, por ejemplo, cuando existe un riesgo claro de manipulación o suplantación (por ejemplo, mediante el uso de robots conversacionales). Los usuarios deben ser conscientes de que están interactuando con una máquina.

• **Riesgos sistémicos derivados de los modelos de inteligencia artificial de uso general:** Por ejemplo, los modelos más potentes, con riesgos sistémicos, podrían causar accidentes graves o ser utilizados indebidamente para ciberataques de amplio alcance. Muchas personas podrían verse afectadas si un modelo propaga sesgos nocivos en múltiples aplicaciones.

¿Cuándo entrará en vigor? Consúltalo en el anexo.

Nuestro código de conducta de la IA

¿Cuáles son nuestros principios éticos?

Nuestros principios de IA, tienen como objetivo promover la adopción de una inteligencia artificial confiable y centrada en el ser humano, así como garantizar un alto nivel de protección de la salud, la seguridad, los derechos fundamentales, la inclusión de todas las personas y el medio ambiente. Estos principios se aplican en todos nuestros productos y servicios desde el diseño y se extienden a nuestros proveedores y socios comerciales.



IA justa

¿Qué significa una IA justa?

- Una IA sin discriminaciones o sesgos por razón de raza, origen étnico, religión, género, orientación e identidad sexual, discapacidad, edad o cualquier tipo de condición de las personas.
- Una IA representativa, diversa y accesible para todos por igual.



¿Cómo puedes hacer que tu IA sea justa?

- Si utilizas datos o variables sensibles para entrenar tus modelos, debes asegurarte de que su uso es estrictamente necesario, está justificado y que no conlleva a ningún tipo de discriminación.
- Considerando las variables proxy.
- Comprobando que tus datos de entrenamiento son representativos para la población o el grupo objetivo.
- Garantizando que la proporción de falsos positivos y negativos no conlleva a ningún tipo de discriminación.
- Asegurándote de que la precisión de tu modelo es la adecuada para el caso de uso en el que estás trabajando para tomar decisiones justas y confiables.
- Evitando cualquier barrera que pueda impedir su uso en igualdad de condiciones por parte de usuarios con discapacidad.



IA transparente y explicable

¿Qué significa una IA transparente y explicable?

- Una IA que comunica de manera entendible su propósito y los datos que utiliza.
- Una IA de la que entendemos y podemos explicar las conclusiones o decisiones para permitir que los usuarios puedan confiar en sus resultados.



¿Cómo puedes hacer que tu IA sea transparente y explicable?

- Comunicando de manera clara los datos que el sistema utiliza y con qué finalidad.
- Siendo capaz de comprender y poder explicar de forma adecuada los resultados a los que llegan nuestros sistemas de IA.
- En el caso de los sistemas que tienen interacción directa con los usuarios (ej. Chatbots), informándoles de forma explícita que se trata de una IA y no un humano.
- Proporcionando transparencia en los contenidos que han sido generados con IA.
- Prestando especial atención si la IA va a ser utilizada por menores de edad y/o personas que no poseen las competencias digitales necesarias para reconocer la interacción con la IA o los contenidos que ésta genera.



IA centrada en las personas

¿Qué significa una IA centrada en las personas?

- Una IA que respeta y protege los derechos humanos, el bienestar y salud de las personas, su autonomía y libertad de decisión.



¿Cómo puedes hacer que tu IA sea centrada en las personas?

- Reflexionando y pensando en los posibles impactos negativos, así como en los usos secundarios o indebidos que puedan hacerse del sistema de IA.
- Calibrando las posibles desviaciones del modelo y los falsos positivos y negativos teniendo en cuenta el ámbito de aplicación.
- Dando la libertad a los usuarios de revocar o revertir las decisiones basadas en el sistema.
- Ofreciendo una supervisión humana adecuada al caso de uso del sistema.
- Asegurando que se puede utilizar de forma sencilla e intuitiva por parte de cualquier persona que carezca de las habilidades digitales necesarias para hacer un correcto uso de la tecnología de forma autónoma.



IA con privacidad y seguridad desde el diseño

¿Qué significa una IA con privacidad y seguridad desde el diseño?

- Una IA que respeta el derecho a la privacidad de las personas y sus datos personales.
- Una IA con un nivel de seguridad adecuado, conforme a las medidas de seguridad establecidas en la legislación vigente en los países donde trabajamos.
- Una IA sólida y robusta frente a los ataques e intentos de manipulación.



¿Cómo puedes hacer que tu IA tenga en cuenta la privacidad y seguridad desde el diseño?

- Cumpliendo con nuestra política de privacidad y con el gobierno de la privacidad de Telefónica.
- Cumpliendo la normativa de seguridad de Telefónica.
- Asegurando la trazabilidad y documentación necesaria para identificar y corregir posibles errores en el desarrollo del sistema.



IA comprometida con el medio ambiente

Este principio está en proceso de incluirse como uno de nuestros principios de IA.

¿Qué significa una IA comprometida con el medio ambiente?

- Una IA medible en cuanto a impacto medioambiental, lo que permite que puedan aplicarse medidas para mejorar su rendimiento y, por tanto, sus impactos nocivos.



¿Cómo puedes hacer que tu IA sea más eficiente y con menor impacto en huella de carbono?

- Haciendo seguimiento y medición del consumo energético y huella de carbono de tus sistemas de IA.
- Buscando medidas para optimizar y eficientar los sistemas de IA en los que trabajas.
- Haciendo seguimiento del impacto que generan los proveedores de sistemas con los que trabajas, en términos de consumo y huella de carbono.
- Entendiendo que no siempre es necesario recurrir a las soluciones más potentes o de mayor necesidad de computación. Dependerá del caso de uso.

¿Y qué pasa cuando trabajas con socios, partners o terceros?

En muchos de nuestros productos y servicios, trabajamos con socios o proveedores o incluso únicamente actuamos con un rol de distribuidor o comercializador. Otras veces, integramos en nuestros sistemas componentes de terceros u open source. Por eso, tienes que asegurarte de integrar nuestros principios a lo largo de toda la cadena de valor de la IA, que en ocasiones es compleja.

Aunque los sistemas de inteligencia artificial de nuestros productos y servicios sean creación de partners o terceros, tienes que asumir un papel activo en la gobernanza de la IA, asegurándote de estén alineados con nuestros principios éticos corporativos, y que se cumplan las normativas aplicables.



¿Cómo puedes asegurarte de integrar nuestros principios cuando colaboras con terceros/socios o integras componentes de terceros en tus sistemas de IA?

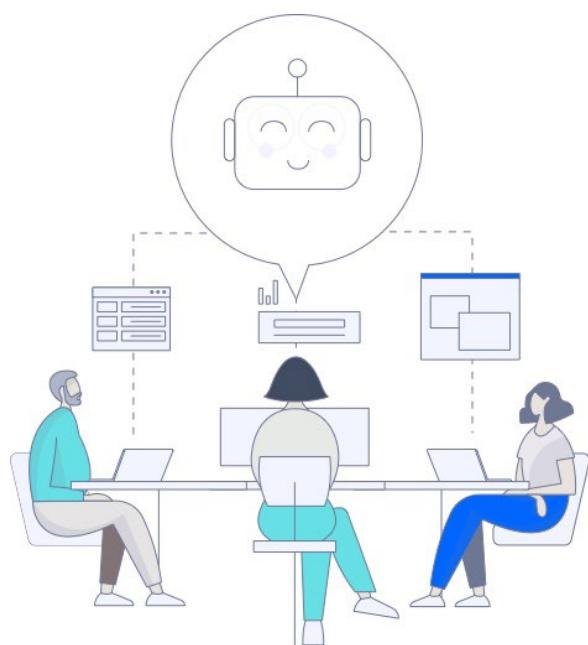
- Incluyendo en los contratos y modelos de compras con terceros los aseguramientos y requisitos que sean necesarios.
- Preguntando a los proveedores de los sistemas con los que trabajas información acerca de si han realizado algún tipo de análisis de sesgos, si tienen documentación disponible del sistema y de su entrenamiento, de los datos que utilizan y de cómo funciona el sistema, de su precisión, etc.
- Asegurando la ausencia de controversias públicas que afecten a potenciales socios o proveedores antes de la firma del contrato.

Gobernanza de IA en Telefónica

¿Qué responsabilidades tengo en la gobernanza de la IA de Telefónica como responsable de producto?

Como responsable de productos y servicios que hacen uso de IA, debes seguir estos puntos:

1. Identificar y [registrar los casos de uso en la herramienta de Gobernanza de IA](#) de Telefónica con el soporte del RAI Champion de tu área, que además solicitará tu acceso a la herramienta.
2. Implementar los [requisitos derivados del análisis de riesgos](#) del caso de uso registrado en Gobernanza de IA. Para ello, contarás con el apoyo de los equipos implicados, como privacidad, legal, ciberseguridad, etc., así como con tu RAI Champion.
3. En el caso de que el caso de uso de IA que has registrado sufra algún tipo de modificación, como, por ejemplo:



- Modificaciones en los datos y variables utilizadas,
- Cambio en los proveedores con los que trabajas,
- Diferente propósito con el que registraste el caso de uso,
- Nuevas funcionalidades no contempladas desde un inicio que puedan tener un impacto en las personas, medio ambiente o la sociedad;

tendrás que revisitar el registro y hacer los cambios pertinentes.

Equipos y responsabilidades de la gobernanza de la IA de Telefónica



RAI Champion

Figura encargada de velar por el uso responsable de la IA en tu área. Podrás hacerle cualquier consulta relacionada con el uso de la IA en tus productos y servicios. También te dará soporte a la hora de completar el registro de tus casos de uso de IA en la herramienta de Gobernanza de IA y en la implementación de los requisitos necesarios.



Equipo de Coordinación de IA

Equipo responsable de promover un uso responsable de la IA en Telefónica, a través de la implementación del modelo de gobierno. Lidera y coordina a los RAI Champions y Coordinadores de IA en las OBs. Se encarga del análisis de riesgos de los casos de uso de IA y de promover la IA Responsable mediante un plan de cultura. .



Grupo de expertos de IA

Expertos en Inteligencia Artificial y otras disciplinas, como Filosofía, Sociología, Privacidad, Medio ambiente, Legal, etc., cuya misión es evaluar los casos de riesgo que supongan un conflicto ético o de interés.



Compliance

Área responsable de promover la implantación del Modelo de Gobierno de IA, además de supervisar y verificar el análisis de riesgos y la correcta implementación de los requisitos derivados del registro.

Nuestro trabajo consiste en mejorar la vida de las personas, innovando de manera responsable y sostenible. Cada uno de nosotros jugamos un papel crucial en la gobernanza de la IA;



asumamos juntos la responsabilidad de garantizar que la IA sea utilizada de manera que beneficie a la sociedad, respete la dignidad humana y fomente un futuro sostenible.

Sigamos aprendiendo, compartiendo y aplicando las mejores prácticas para que la IA sea nuestra aliada en la construcción de un futuro mejor.



ANEXO²

2 Inteligencia artificial: preguntas y respuestas*

https://ec.europa.eu/commission/presscorner/detail/es/qanda_21_1683



Ten en cuenta que este anexo es un resumen conciso del texto legal y, aunque se pretenden cubrir los aspectos cruciales, puede haber casos específicos que no se encuentren reflejados en su totalidad. En caso de requerir información adicional, acude al equipo de Coordinación o de Compliance, quienes te proporcionarán el soporte necesario.

Categorías de alto riesgo

- Determinadas infraestructuras críticas, por ejemplo, en los ámbitos del tráfico por carretera y el suministro de agua, gas, calefacción y electricidad.
- La educación y la formación profesional, por ejemplo, para evaluar los resultados de la enseñanza, dirigir el proceso de aprendizaje y controlar que no se hagan trampas.
- El empleo, la gestión del personal laboral y el acceso al trabajo el acceso a servicios y prestaciones públicos y privados esenciales (por ejemplo, asistencia sanitaria), la evaluación de la solvencia de las personas físicas, y la evaluación de riesgos y la fijación de precios en relación con los seguros de vida y de enfermedad.
- Determinados sistemas utilizados en los ámbitos de la policía, el control de fronteras, la administración de justicia y los procesos democráticos.
- La evaluación y clasificación de las llamadas de emergencia.
- Los sistemas de identificación biométrica, categorización y reconocimiento de emociones (fuera de las categorías prohibidas).

Categorías de riesgo inadmisible

- Puntuación social con fines públicos y privados;
- Aprovechamiento de los puntos vulnerables de las personas, uso de técnicas subliminales;
- Identificación biométrica remota, en tiempo real, en lugares de acceso público por parte de las fuerzas y cuerpos de seguridad, con excepciones limitadas (véase más adelante);
- Categorización biométrica de las personas físicas basada en datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, u orientación o identidad sexual; seguirá siendo posible filtrar conjuntos de datos basados en datos biométricos en el ámbito policial;
- Actuación policial predictiva individual;

- Reconocimiento de emociones en el lugar de trabajo y en las instituciones educativas, a menos que sea por razones médicas o de seguridad (por ejemplo, control del grado de cansancio de un piloto);
- Extracción no selectiva de internet o circuito cerrado de televisión para imágenes faciales con el fin de crear o ampliar bases de datos.

¿Cuándo entrará en vigor?

Tras su adopción por el Parlamento Europeo y el Consejo, la Ley de IA entrará en vigor a los veinte días de su publicación en el Diario Oficial. Será plenamente aplicable veinticuatro meses después de su entrada en vigor, de forma gradual como se explica a continuación:

- seis meses después de la entrada en vigor, los Estados miembros irán eliminando progresivamente los sistemas prohibidos;
- doce meses después serán aplicables las obligaciones de la gobernanza de la inteligencia artificial de uso general;
- veinticuatro meses después serán aplicables todas las normas de la Ley de IA, incluidas las obligaciones relacionadas con los sistemas de alto riesgo definidos en el anexo III (lista de casos de usos de alto riesgo);
- treinta y seis meses después serán aplicables las obligaciones relacionadas con los sistemas de alto riesgo definidos en el anexo II (lista de la legislación de armonización de la UE).