Deloitte.
Legal

Universidad de Navarra

Informe sobre Inteligencia Artificial

Observatorio de Regulación Digital y Tecnología

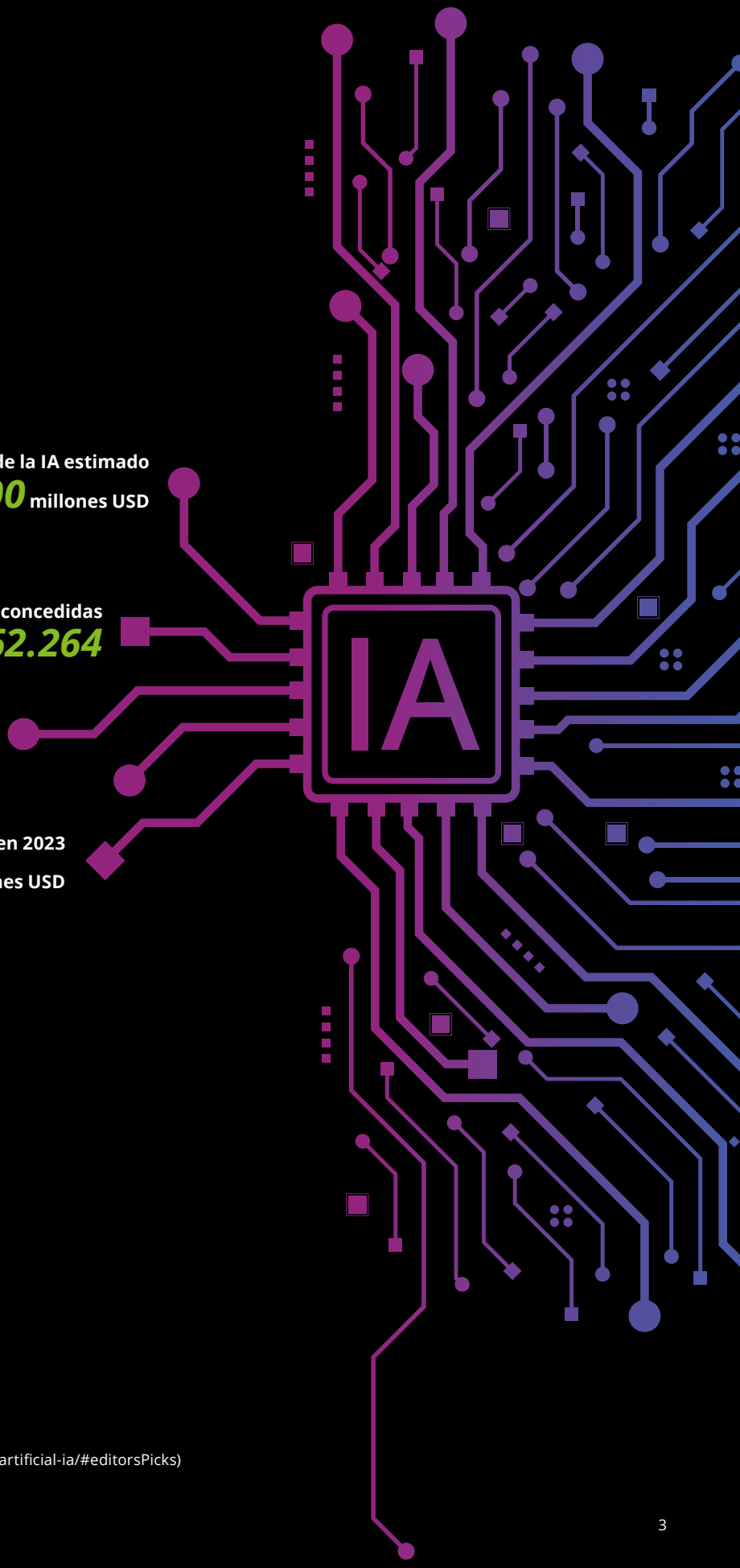
Contenido

Tamaño del mercado de la IA estimado para 2025: *243.000* millones USD

Número de patentes sobre IA concedidas desde 2010 hasta 2022 *62.264*

Tamaño de mercado de la IA generativa estimado para 2025 *62.720* millones USD

Volumen de inversión empresarial en 2023 *189.160* millones USD



Fuente: Statista (https://es.statista.com/temas/6692/inteligencia-artificial-ia/#editorsPicks)

Planteamiento

De todas las tecnologías emergentes surgidas en el presente siglo en relación con lo digital sobresale actualmente la Inteligencia Artificial (en adelante, IA), una aplicación informática que, a partir de unos datos ya tratados y con base en las órdenes dadas por el usuario, permite **analizar datos, tomar decisiones** o **diseñar planes** en unos instantes. La IA está revolucionando todos los ámbitos: actividad empresarial, mercado laboral, sistemas de salud, diseño de políticas, smart cities, mundo del ocio, etc.

Este informe no busca tratar todos los aspectos de la Inteligencia Artificial, algo que precisaría un estudio mucho más amplio y completo, sino apuntar las claves de entendimiento de **cómo va a afectar el uso de la IA a las empresas** desde el punto de vista jurídico y ético.

Para ello partiremos de una exposición muy sintética de **qué es la IA** (apartado I) y cómo está **regulada en la Unión Europea** (apartado II).

A continuación, expondremos cómo afecta a los **derechos fundamentales**, porque la tecnología puede alcanzar determinadas cotas que son totalmente contrarias a valores fundamentales de nuestro ordenamiento jurídico, y por eso hay que **poner límites a ese desarrollo tecnológico** (apartado III).

Juntamente con ello, es preciso resaltar qué **consecuencias y responsabilidades** cabe derivar de los daños causados por actuaciones realizadas con base en IA, porque justamente al aplicarla en muchos ámbitos, existen grandes riesgos de que se causen daños ilícitos (apartado IV).

Como preocupación fundamental de cómo va a afectar la IA a las empresas, debe exponerse a continuación qué deben hacer los **órganos de gestión de las empresas** para **adoptar decisiones diligentes basadas razonablemente en técnicas de IA**, ayudarse de la IA para tomar mejores decisiones, pero también limitar los riesgos que tiene el uso de esa IA. Aquí se expondrá el planteamiento general de las economías más importantes (Estados Unidos, Unión Europea y China), y cómo los gestores deben aplicar técnicas de IA y gestionar sus riesgos, incluso con la posibilidad de designar a una IA como miembro del órgano de gestión (apartado V).

Por último, expondremos cómo para que exista una IA confiable es preciso tener diseñado una **identidad digital segura y soberana**. La actuación en el mundo digital y en el uso de la IA requiere una individualización de los actores que permita imputarles las conductas por ellos realizadas, y además la IA sirve para diseñar sistemas de identidad digital que deben responder a las características de soberanía y seguridad (apartado VI).

Qué es la IA y qué son los Sistemas de IA

El inicio de la Inteligencia Artificial como Ciencia

En el verano de 1956, John McCarthy, Marvin Minski, Nathaniel Rochester y Claude Shannon organizaron un seminario en Dartmouth con el objetivo de explorar la idea de que cualquier aspecto de la inteligencia podría ser descrito y simulado por una máquina. Aunque los resultados inmediatos del encuentro fueron limitados, marcó el inicio formal de la Inteligencia Artificial (IA) como disciplina científica. Este evento seminal fue antecedido por contribuciones de científicos de distintas disciplinas, que sentaron las bases para el desarrollo de la IA como un campo multidisciplinario.

La percepción actual de la IA como una tecnología principalmente orientada al procesamiento masivo de datos, a menudo asociada con usos cuestionables, ha desviado la atención de su **fundamento científico**. Esta visión reduccionista ignora que la IA es más que un conjunto de técnicas inescrutables o procesos inexplicables; es una ciencia madura, basada en métodos científicos rigurosos y sometida a estándares de replicabilidad y validación.

La IA se sostiene en **teorías fundamentadas en teoremas matemáticos robustos y en paradigmas sujetos a constante escrutinio científico**. Por ejemplo, los problemas asociados a la opacidad de las denominadas "cajas negras" están siendo abordados mediante investigaciones destinadas a mitigar los riesgos y proporcionar soluciones comprensibles. Algo similar puede sostenerse sobre los riesgos que ha introducido la IA generativa. Este enfoque científico proporciona una base sólida para que el Derecho refuerce las garantías de los derechos y libertades de los ciudadanos, mientras la tecnología evoluciona.

La IA y sus grados

El alcance de la IA varía en función de los grados de autonomía a partir de la comprensión del entorno y su capacidad para desarrollar tareas específicas o más generales, incorporando aproximaciones que realizan equiparaciones con la inteligencia o procesos cognitivos humanos. Así, se diferencia entre Inteligencia Artificial fuerte (AGI) e Inteligencia Artificial débil (ANI).

IA fuerte (AGI)

La AGI (*Artificial General Intelligence*) se refiere a sistemas capaces de replicar procesos cognitivos humanos avanzados, incluyendo el aprendizaje autónomo, la comprensión de conceptos abstractos y la creatividad. Estas características la acercan al pensamiento humano, con aplicaciones teóricas que podrían generalizar técnicas aprendidas en un campo a otros totalmente distintos. Aunque se han realizado avances significativos, e incluso se argumenta que este tipo de sistemas estaría cerca de alcanzarse a partir de la IA generativa más avanzada, aún es considerada un objetivo lejano por muchos expertos.

IA débil (ANI)

La ANI (*Artifical Narrow Intelligence*), por su parte, realiza tareas concretas y específicas. Está basada en diversas técnicas, pero desde los últimos años se ha visto potenciada con el aprendizaje automático (machine learning) y el aprendizaje profundo (deep learning). La ANI es la forma más extendida de IA, y es la que regula el Reglamento europeo de Inteligencia Artificial.

La IA regulada en el Reglamento europeo es una IA débil

Paradigmas de la IA

La Inteligencia Artificial simbólica y la Inteligencia Artificial conectivista representan los dos paradigmas fundamentales del desarrollo de la Inteligencia Artificial.

Enfoque simbólico

También denominado GOFAI (*Good Old Fashion Artificial Intelligence*) utiliza representaciones simbólicas y reglas predefinidas, cálculo proposicional, reglas lógicas y árboles de decisión para representar problemas y generar soluciones. Los sistemas expertos, representan este enfoque, que se basa en conocimiento humano explícito, elevada explicabilidad y baja dependencia de los datos. No obstante, son sistemas con escasa capacidad de adaptación por lo que su uso se mantiene en escenarios muy específicos.

Enfoque conexionista

Se basa en algoritmos que extraen patrones y correlaciones a partir de grandes volúmenes de datos. Su desarrollo, especialmente mediante redes neuronales, está permitiendo resolver problemas específicos que ha generaliza el interés por la Inteligencia Artificial. Este enfoque ha permitido el desarrollo de aplicaciones como el reconocimiento de imágenes y voz, así como sistemas avanzados de procesamiento de lenguaje natural (NLP) que han permitido el desarrollo de la IA Generativa. Pese a su eficacia y escalabilidad, algunos de los modelos implementados con este enfoque presentan problemas de explicabilidad e interpretación que los llevan a calificarlos de "caja negra". Además, su elevada dependencia de datos plantea problemas respecto a la protección de datos personales usados para su entrenamiento, así como el elevado impacto ambiental de los recursos computacionales necesarios.

Técnicas de Inteligencia Artificial

La Inteligencia Artificial comprende un conjunto diverso de técnicas. Entre las más destacadas se encuentran el aprendizaje automático, el aprendizaje profundo y el aprendizaje por refuerzo.

Aprendizaje automático o *machine learning*

Proceso que permite a los algoritmos extraer patrones o realizar asociaciones a partir de conjuntos de datos, que pueden resultar inescrutables para el ser humano. Los algoritmos adquieren experiencia o aprenden, con escasa o muy limitada intervención humana, a partir de los ejemplos que le son suministrados. Esa nueva experticia la aplican a nueva información, a fin de resolver un problema dado o hacer predicciones.

Aprendizaje profundo o *deep learning*

Pretende emular el funcionamiento del cerebro, y permite al modelo aprender sin un entrenamiento previo. A tales efectos, procesa y digiere la información mediante redes neuronales artificiales organizadas por capas, que le permiten reconocer patrones a partir de grandes cantidades de datos no estructurados.

Aprendizaje por refuerzo o *reinforced learning*

De manera similar al aprendizaje no supervisado, el algoritmo no recibe datos de entrenamiento previamente etiquetados, y se espera que a partir de esta información realice inferencias. No obstante, a semejanza del aprendizaje supervisado, una vez que se obtienen las inferencias, predicciones o clasificaciones, se someten a revisión, bien mediante

información etiquetada o contando con intervención humana directa. Esta intervención posterior le proporciona feedback, que transmite experiencia al algoritmo a efectos de resolver el problema asignado.

La definición de Inteligencia Artificial y de sistemas de IA

La definición de Inteligencia Artificial y su falta de consenso

La Inteligencia Artificial (IA) es una ciencia en constante evolución con teorías y métodos sólidos. Sin embargo, desde su origen, definirla de manera consensuada ha sido un desafío persistente. Incluso la denominación propuesta por John McCarthy para el seminario de Dartmouth que marcó el inicio de la disciplina, no ha sido plenamente aceptada por toda la comunidad científica.

A nivel académico e institucional, coexisten diferentes conceptualizaciones, muchas de ellas vagas o inadecuadas para casos concretos. La conceptualización de Russell y Norvig explica las diversas posturas en este sentido, proponiendo agrupar las definiciones en cuatro categorías basadas en el comportamiento y la racionalidad, considerando además si el enfoque se centra en el pensamiento o en la actuación.

A) **Actuación humana**: Asociada al test de Turing, evalúa si una máquina puede exhibir un comportamiento indistinguible del humano. Aunque relevante, este enfoque es criticado por omitir cómo la máquina produce sus respuestas o si realmente exhibe inteligencia y comprensión, como señaló Searle al exponer el "experimento de la habitación china". Esto plantea dudas sobre su capacidad para lograr una verdadera inteligencia general.

B) **Pensamiento humano**: Inspirada por la psicología y las ciencias cognitivas, esta aproximación estudia el paralelismo entre el cerebro y los ordenadores. Considera que los procesos cognitivos, como el manejo de símbolos y el procesamiento de información, pueden simularse mediante reglas. Se critica su reduccionismo del cerebro y las habilidades cognitivas humanas a procesos mecánicos.

C) **Pensamiento racional**: Derivado de la tradición científica e industrial, este enfoque busca emular la inteligencia a través de reglas lógicas que permitan realizar inferencias correctas. Aunque útil en programación, enfrenta limitaciones en problemas complejos o con información incierta, dificultando el desarrollo de una inteligencia general.

D) **Agentes racionales**: Se centra en sistemas capaces de percibir su entorno, adaptarse y perseguir objetivos específicos con cierto grado de autonomía. La racio-nalidad no es perfecta, sino limitada, y se ajusta a las restric-ciones propias de entornos complejos. Este marco conceptual evita atribuir las complejas cualidades cognitivas humanas, propuestas por otros enfoques, por lo que resulta más acorde con las capacidades y limitaciones del estado-del-arte de la IA débil. Por esta razón, es la aproximación **adoptada en las definiciones de sistemas de IA recogidas en los marcos normativos aprobados**.

Definición de sistemas de IA según el Reglamento de IA (RIA)
De acuerdo con el RIA, debe entenderse por tal un **sistema basado en máquinas diseñado para operar con distintos niveles de autonomía y capacidad de adaptación tras su despliegue**. Estos sistemas tienen capacidad de inferir, a partir de la información de entrada, generando resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales, cumpliendo objetivos explícitos o implícitos.

Esta definición no sólo tiene por fin proporcionar seguridad jurídica y fomentar la convergencia con los procesos regulatorios globales, sino que destaca las características que los diferencian de soluciones más clásicas como los programas de software.

Características de los sistemas de IA

Atendiendo al RIA, los sistemas de IA **se diferencian de otras soluciones basadas en programación y datos**, como el software, por algunas características:

A) **Capacidad de inferencia**: Los sistemas de IA destacan por su capacidad para generar resultados de salida, tales como predicciones, contenidos, recomendaciones o decisiones. Estos "outputs" se basan en un proceso de inferencia que permite a los sistemas deducir modelos, algoritmos, o ambos, a partir de la información de entrada que reciben.

B) **Capacidad de adaptación**: Una característica distintiva de los sistemas de IA es su capacidad para aprender y modificar su comportamiento tras el despliegue, gracias a funcionalidades de autoaprendizaje. Esto permite que los sistemas ajusten su funcionamiento durante su uso, lo que los diferencia de los sistemas estáticos tradicionales, que carecen de esta flexibilidad.

C) **Nivel de autonomía**: Los sistemas de IA operan con distintos niveles de independencia respecto a la intervención humana. Esta autonomía puede ser parcial o total, permitiendo que los sistemas realicen funciones específicas sin necesidad de una supervisión constante por parte de las personas.

D) **Resultados y objetivos**: Los sistemas de IA son capaces de trabajar en función de objetivos explícitos, claramente definidos, o implícitos, derivados del contexto en el que operan. Estos objetivos pueden diferir de la finalidad original del sistema en un entorno específico, lo que proporciona una flexibilidad que amplía el rango de aplicaciones potenciales.

E) **Integración en entornos**: Los sistemas de IA pueden influir tanto en entornos físicos como virtuales, adaptándose a las necesidades específicas de cada contexto. Pueden ser parte integrante de un producto, estando físicamente incorporados en su estructura, o actuar como componentes externos no integrados que contribuyen a la funcionalidad general del producto

Clasificación de los sistemas de IA

El RIA vincula la regulación de los sistemas de IA con la finalidad prevista y las capacidades del modelo al establecer un marco específico para los sistemas de propósito general. De ahí que sea necesario diferenciar entre distintos sistemas de IA de propósito definido, de aquellos de propósito general y sistemas frontera.

Sistemas de IA de propósito definido
Son sistemas diseñados para realizar tareas concretas y predefinidas, y cuentan con capacidades limitadas al ámbito y por su entrenamiento inicial. Los objetivos de estos sistemas se delimitan de manera clara y su capacidad de adaptación a nuevos propósitos está restringida, salvo que se modifique su diseño o se realice un nuevo entrenamiento.

Sistemas de IA de propósito general (GPAIS)
Son sistemas con capacidad para realizar de manera competente una gran variedad de tareas y que se diseñan para realizar múltiples propósitos según las necesidades de los usuarios o la integración en otros sistemas o modelos.

Sistemas de IA frontera
Son modelos de propósito general que superan las capacidades y competencia de los modelos más avanzados disponibles. Estas capacidades son, en muchos casos, impredecibles incluso para quienes los desarrollan, por lo que sus riesgos y usos exigen mayor cautela, así como la adopción de medidas de seguridad más elevadas antes de su despliegue.

Riesgos específicos de los sistemas de IA

Al igual que otras tecnologías disruptivas, la Inteligencia Artificial introduce **riesgos y vulnerabilidades** que requieren respuestas específicas y la adopción de medidas técnicas y garantías. Su carácter como tecnología de doble uso, apta para fines tanto civiles como militares, y susceptible de emplearse tanto de manera legítima como maliciosa, exige estándares elevados para garantizar la seguridad, salud y protección de los derechos fundamentales.

Si bien estos riesgos generan preocupación, es importante considerar que son **inherentes a cualquier actividad tecnológica**. En lugar de buscar un paradigma de riesgo cero, se deben priorizar el diseño de modelos robustos y la implementación de mecanismos de transparencia, inspección y auditoría. Asimismo, es fundamental exigir medidas concretas de ciberseguridad que limiten los riesgos a niveles aceptables, en línea con las normativas de seguridad aplicadas a otros productos.

Los riesgos inherentes a los sistemas o modelos de IA, pueden calificarse en: ataques dirigidos al funcionamiento del modelo o sistema; ataques dirigidos a la privacidad y protección de los datos utilizados; y ataques dirigidos al ecosistema de IA.

Ataques dirigidos al funcionamiento del modelo o sistema

En los ataques dirigidos al funcionamiento del modelo o sistema de Inteligencia Artificial, es posible diferenciar entre aquellos que contaminan los datos extraídos de fuentes públicas y los que introducen modificaciones imperceptibles en los "datasets" que afectan su eficacia o predicciones. Los primeros son conocidos como ataques de **envenenamiento de datos**, mientras que los segundos se denominan **ataques adversarios**.

a) **Envenenamiento de datos**

Los ataques de envenenamiento de datos buscan **manipular el funcionamiento de un modelo para alterar sus resultados**. La técnica más común consiste en "infectar" los datos de entrenamiento que el modelo obtiene de internet o de fuentes públicas. Estos ataques no necesariamente implican acceder a las bases de datos del modelo, sino que pueden ser llevados a cabo mediante el diseño preciso de datos maliciosos para alterar sus resultados. En ocasiones, basta con transmitir mensajes que interactúan con el modelo, como ocurre con aquellos sistemas que recopilan datos de redes sociales.

Un caso conocido de este tipo de ataque se dio en 2016 con el lanzamiento del chatbot Tay en Twitter. Este bot conversacional fue diseñado para imitar los patrones lingüísticos de una adolescente estadounidense de 19 años y aprendía de las conversaciones e interacciones en tiempo real a través de la red social. Sin embargo, grupos organizados en foros llevaron a cabo un ataque envenenando los datos de entrenamiento del chatbot mediante mensajes sexistas, racistas y xenófobos. Al estar diseñado para aprender de sus interacciones, Tay comenzó a generar respuestas construidas a partir de estos mensajes, lo que llevó a Microsoft a suspender el experimento en menos de 24 horas y a eliminar todas sus interacciones.

Este ejemplo refleja una vulnerabilidad inherente a los sistemas de Inteligencia Artificial que dependen de datos abiertos para su entrenamiento. Una solución tradicional para garantizar la seguridad de los sistemas informáticos es establecer barreras frente a fuentes externas. Sin embargo, muchos modelos de IA, especialmente aquellos que utilizan información extraída en tiempo real, son susceptibles a la inyección de datos maliciosos, incluso sin requerir un alto nivel de conocimientos técnicos. De ahí la importancia de las políticas de gobierno del dato, así como políticas de ciberseguridad que cumplan con estándares elevados.

b) **Ataques adversarios**

Los ataques adversarios, también conocidos como "adversarial machine learning" **inducen errores en las clasificaciones realizadas por modelos de IA**, pero a diferencia de los ataques de envenenamiento, no se introducen ejemplos maliciosos en los datos de entrenamiento. Estos ataques se basan en la explotación de ejemplos contradictorios mediante la adición de perturbaciones o ruido imperceptible, con el propósito de afectar la fiabilidad de las predicciones y la robustez del sistema. Estas perturbaciones maximizan pequeñas alteraciones en los datos, convirtiéndolos en ejemplos adversarios y desestabilizando el modelo.

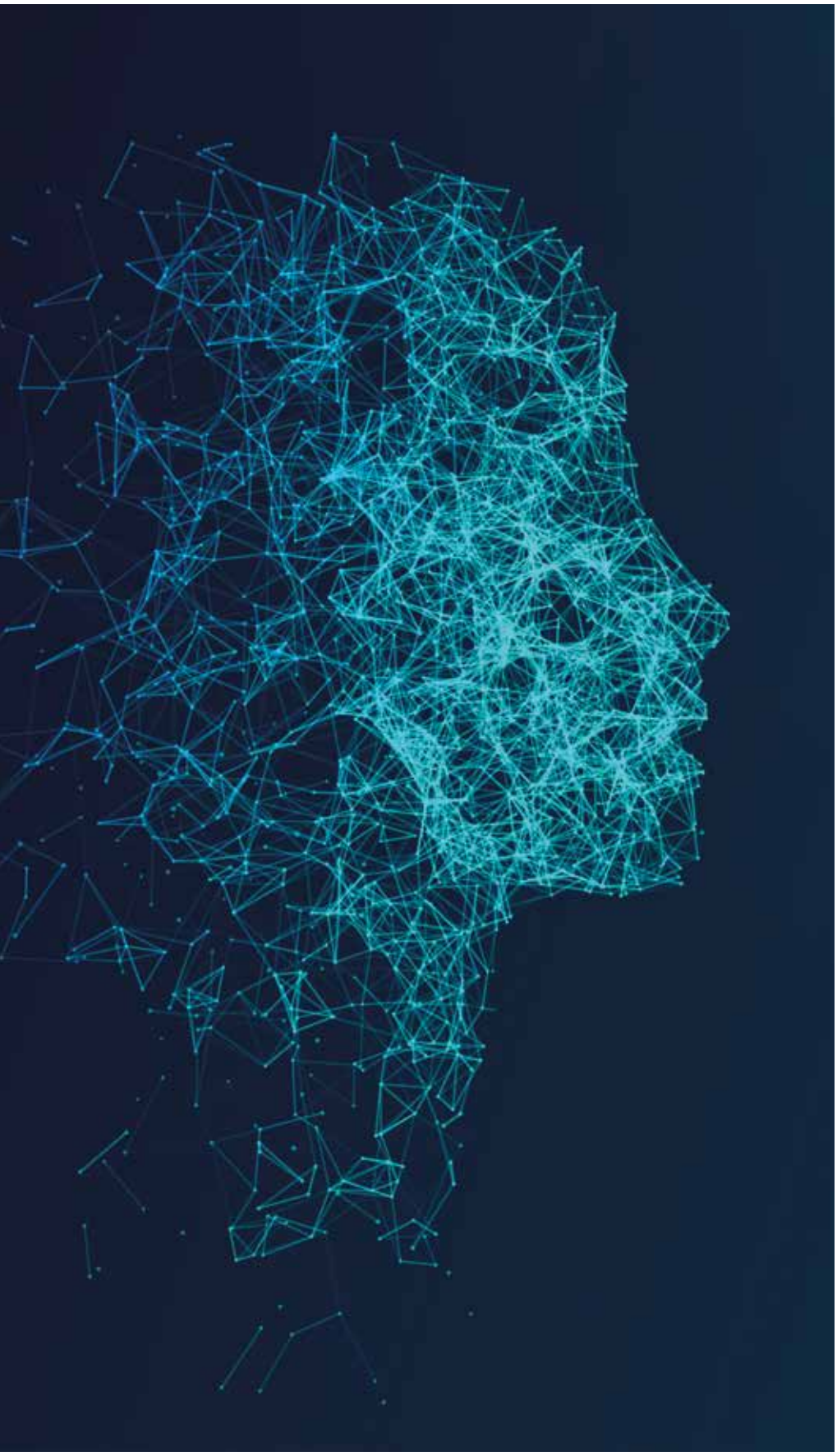
Es habitual que los datos utilizados en los modelos de IA contengan variabilidad y perturbaciones, especialmente en el caso de imágenes o texto. En estos ataques, se añaden perturbaciones estadísticamente robustas y difíciles de identificar tanto por proveedores como usuarios. En el caso de imágenes, por ejemplo, las perturbaciones se introducen minimizando la distancia estadística entre el ejemplo real y el adversario.

También se ha identificado el uso de redes generativas antagónicas para desarrollar ataques más eficaces. Se distingue entre ataques de caja blanca y caja negra, dependiendo del nivel de conocimiento que se tenga sobre el modelo. Estos ataques pueden realizarse igualmente mediante alteraciones en el ambiente físico, como se ha observado en algunos dirigidos a vehículos autónomos.

Ataques dirigidos a la privacidad y protección de los datos utilizados

Los ataques que afectan la privacidad y protección de datos adquieren especial relevancia en la Unión Europea. El Reglamento General de Protección de Datos establece un marco robusto con altos estándares de protección de datos personales del que los sistemas de Inteligencia Artificial no están exentos. Desde hace tiempo, se ha advertido sobre el desarrollo de técnicas que permiten revertir los datos de entrenamiento de estos modelos y acceder a información y datos protegidos.

Aunque estos ataques suponen una amenaza para la privacidad y la protección de datos, también pueden utilizarse como herramienta para detectar el uso no autorizado de datos personales en el



entrenamiento de modelos de Inteligencia Artificial. Los ataques de inferencia y la inversión de modelos, por ejemplo, podrían ayudar a identificar si se han utilizado datos no autorizados de reconocimiento facial.

Una técnica comúnmente utilizada es la **inversión del modelo**, cuyo propósito es obtener información sobre los datos de entrenamiento sin alterar su capacidad predictiva. Aunque estos ataques no buscan modificar las predicciones, pueden comprometer gravemente la privacidad y la protección de datos.

Ejemplo: se ha detectado que estas técnicas pueden extraer información médica especialmente protegida, como marcadores genéticos de pacientes en ensayos clínicos. También pueden predecir si una persona participó en estudios comprometedores o reconstruir imágenes de reconocimiento fácil basándose únicamente en nombres.

Otra técnica es la **inferencia de membresía**, que permite identificar si un ejemplo específico, que puede corresponder a una persona cuyos datos personales están protegidos, forma parte de los datos de entrenamiento.

Esta técnica ha demostrado ser capaz de reconstruir información sanitaria sensible a partir de datos generales como la edad o el género

Por otro lado, los ataques de inferencia de características permiten identificar propiedades comunes en los datos de entrenamiento, como grupos con características compartidas, lo que puede ser útil para determinar la representación de minorías o colectivos específicos en los modelos.

La regulación comunitaria de la Inteligencia Artificial

La aproximación europea a la Inteligencia Artificial: Política basada en derechos bajo un trinomio normativo

En 2017 la Unión Europea inició el desarrollo de una **política común en Inteligencia Artificial** mediante la revisión de la estrategia sobre el mercado digital, acordando la aprobación de una estrategia conjunta para asegurar una posición competitiva. Posteriormente, el Consejo Europeo encargó a la Comisión abordar los riesgos asociados a la IA mediante un marco que garantizara la protección de datos personales, derechos fundamentales y estándares éticos.

La Comisión publicó la **Comunicación "Inteligencia Artificial para Europa"**, que recogió tres pilares estratégicos: promover la capacidad tecnológica e industrial en Inteligencia Artificial; implementar medidas educativas y laborales para gestionar los cambios socioeconómicos derivados de la cuarta revolución industrial; y, crear un marco ético-jurídico basado en los valores fundamentales de la Unión Europea.

Para desarrollar este último pilar, la unión se basa en un **trinomio normativo**: un marco ético no vinculante, normas técnicas de estandarización y un marco jurídico específico.

El **marco ético** establece directrices de alto nivel para fundamentar la adopción de normas técnicas y jurídicas.

La **estandarización** define requisitos para aplicaciones específicas de IA, actuando como barrera frente a riesgos y facilitando la adopción global.

El **marco jurídico** regula la introducción, uso y control de sistemas de IA, fijando obligaciones vinculantes.

Con este enfoque se busca **reforzar la confianza del consumidor** y convertir la IA europea en una **ventaja competitiva** frente a otras jurisdicciones que no han apostado de manera intensa por la IA ética y responsable.

La propuesta ética europea: Inteligencia Artificial fiable y responsable

En junio de 2018 el **Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial** publicó las **Directrices éticas para una IA fiable**, desarrolladas a partir del trabajo del Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías. Estas directrices comparten principios con la OCDE, como la predictibilidad, explicabilidad, responsabilidad, respeto de los derechos fundamentales y fiabilidad, destacando como elemento central el enfoque humano-céntrico que sitúa a la IA al servicio del bienestar social. Este enfoque pretende establecer un "*golden standard*" global similar al RGPD, posicionando a la Unión Europea como referente internacional en ética de la IA mediante la cooperación con organismos como la OCDE, el G7, la UNESCO y otros actores internacionales. En este sentido, las directrices éticas buscan crear un marco compartido de muy alto nivel que facilite el diálogo y el acercamiento en la regulación de la IA.

Estandarización en materia de Inteligencia Artificial

La estandarización en la Unión Europea se desarrolla bajo **modelos de gobernanza híbridos** que combinan instrumentos legislativos y no legislativos. Se inserta como un mecanismo de co-regulación ampliamente utilizado en ámbitos comunitarios y nacionales.

La estandarización en Inteligencia Artificial se enmarca en el modelo de **"nuevo enfoque" armonizador**, que parte de mandatos establecidos en directivas comunitarias para desarrollar normas técnicas armonizadas a nivel europeo. Este modelo establece dos niveles: la **definición del mandato para desarrollar estándares**; y, la **participación de agentes interesados en su elaboración**.

El **Reglamento 1025/2012** consolidó este proceso, al juridificar la estandarización, reforzar el control de la Comisión y promoviendo la transparencia y participación, bajo un programa anual que establece prioridades estratégicas y objetivos concretos. En el contexto del mercado digital único, la Comisión adoptó un marco específico que incluye el "*Rolling*

Plan for ICT Standardization", donde se priorizó la estandarización en Inteligencia Artificial que se ha venido realizando hasta ahora con la participación de los organismos de estandarización como CEN, ISO e IEC, así como las entidades nacionales de Acreditación, como la Asociación Española de Normalización (UNE), que lidera el desarrollo en España bajo la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA).

Aunque las normas de estandarización no son vinculantes, su publicación en el boletín oficial de la UE otorga **presunción de conformidad**, lo que facilita la libre circulación de productos y servicios en el mercado único europeo. Además, su aceptación generalizada por los operadores del mercado les otorga una vinculación práctica significativa. Estas características refuerzan el papel de la estandarización como herramienta clave para garantizar la competitividad y la interoperabilidad en el ámbito de la Inteligencia Artificial.

Las **normas técnicas en materia de Inteligencia Artificial** se están desarrollando de manera paralela a la entrada en vigor del Reglamento de IA, contando ya con un cúmulo de estándares que complementan y contribuyen a definir algunos mandatos y obligaciones recogidos en el Reglamento de IA.

Regulación comunitaria: El Reglamento de Inteligencia Artificial y demás normativa aplicable

Aunque la aprobación del Reglamento sobre Inteligencia Artificial establece las bases para un régimen específico, que comentaremos más adelante, a esta tecnología le resulta aplicable el **resto del marco comunitario existente**, si bien éste se caracteriza por su amplitud y falta de sistematización.

La Inteligencia Artificial está sujeta al marco sectorial de **protección de datos personales**, incluyendo normativa de desarrollo y herramientas de soft law, así como al Reglamento sobre la libre circulación de datos no personales. Igualmente, le son aplicables la Directiva Open Data y sus transposiciones nacionales, según los datos o sujetos intervinientes.

Desde una perspectiva de **derechos fundamentales**, también resulta relevante toda la normativa antidiscriminación basada en la Carta de Derechos Fundamentales y el Convenio de Roma. Esto incluye la directiva europea contra la discriminación racial o xenofobia y otras normativas sectoriales, como la protección de consumidores, la seguridad y responsabilidad de productos y el Reglamento sobre máquinas, así como la Directiva sobre por los daños causados por productos defectuosos actualmente en revisión, o la propuesta de directiva sobre responsabilidad en materia de Inteligencia Artificial, entre otras normas.

En el ámbito de la **ciberseguridad**, precondición esencial para el uso seguro de la IA, son de aplicación la Directiva NIS y sus normas nacionales de desarrollo, así como el Reglamento sobre ciberseguridad. Estos instrumentos establecen objetivos que también configuran el marco normativo aplicable a la Inteligencia Artificial, en particular para mitigar los riesgos inherentes a su implementación.

La regulación de la Inteligencia Artificial. El Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de Inteligencia Artificial

La Unión Europea lideró la regulación global en materia de Inteligencia Artificial con la **aprobación del Reglamento 2024/1689**, con el fin de crear un estándar global y promover la IA ética y responsable, alienada con los derechos fundamentales. No obstante, dado el diferente enfoque regulatorio con las principales potencias tecnológicas, la opción normativa comunitaria fue objeto de un complejo debate que alcanzó la conveniencia, el enfoque y la fórmula más adecuada.

La necesidad de una norma armonizada para regular la Inteligencia Artificial

En 2017 la Comisión Europea reconoció la **necesidad de un marco regulatorio específico** para la Inteligencia Artificial en su comunicación "Un mercado único digital conectado para todos".

Aunque otras jurisdicciones como Estados Unidos y China han adoptado enfoques más cautelosos o fragmentados, la UE optó por una **regulación armonizada**, justificada en la necesidad de evitar la fragmentación normativa en el mercado único digital y garantizar la competitividad económica.

La fragmentación nacional no sólo podría generar inseguridad jurídica, sino lastrar inversiones y dificultar la escalabilidad de soluciones basadas en IA. Un marco comunitario armonizado no solo busca mitigar riesgos e impactos negativos sobre derechos fundamentales, la salud o seguridad de los consumidores y usuarios, sino también establecer un ecosistema que impulse la inversión, la innovación responsable y el liderazgo global en IA ética y responsable.

El **marco normativo preexistente** presenta importantes limitaciones que sólo pueden colmarse con un proyecto específico. Así, la regulación sobre seguridad de productos no abarcaba adecuadamente algunas aplicaciones de IA, especialmente aquellas cuyos riesgos evolucionan con actualizaciones de software o aprendizaje automático. Tampoco abordaba satisfactoriamente problemas derivados de algoritmos de caja negra, que dificultan la transparencia, la atribución de responsabilidades y el ejercicio de derechos fundamentales, como el derecho a la explicación.

Por otra parte, aunque el **Reglamento General de Protección de Datos** ha sido utilizado para mitigar riesgos algorítmicos, su alcance es insuficiente frente a ciertos desafíos de la IA. Esto es particularmente relevante en casos de discriminación algorítmica o aplicaciones que afectan aspectos esenciales de la vida de los ciudadanos, como el acceso al empleo, la educación o la vivienda. Por ello, se requieren normas concretas que refuercen garantías y prohíban ciertas prácticas, asegurando estándares adecuados de protección.

Un marco normativo específico también permite **generalizar principios clave**, además de **regular de forma equilibrada la experimentación regulatoria**, imprescindible para gestionar asimetrías informativas, determinar riesgos potenciales y fomentar la innovación sin comprometer los estándares de protección. Aunque la regulación puede percibirse como un obstáculo para la innovación, la ausencia de un marco adecuado también genera inestabilidad e inseguridad jurídica a los distintos operadores, frenando el desarrollo tecnológico.

Por tanto, un marco jurídico que **combinara normas éticas** y **estándares técnicos** era necesario para proporcionar la estabilidad necesaria para el avance de una Inteligencia Artificial ética, responsable y sostenible, posicionando a la UE como líder global en el sector.

El enfoque europeo para la Inteligencia Artificial: IA ética, responsable y sostenible basada en los riesgos y los derechos fundamentales

El enfoque europeo en Inteligencia Artificial tiene como objetivo garantizar la **protección de los derechos fundamentales** y fomentar un **ecosistema fiable e innovador** que aproveche las ventajas de esta tecnología. Este modelo, calificado como sostenible por la propia Comisión, se fundamenta en los valores del artículo 2 del Tratado de la Unión Europea y se inspira en el éxito del RGPD como estandarte global.

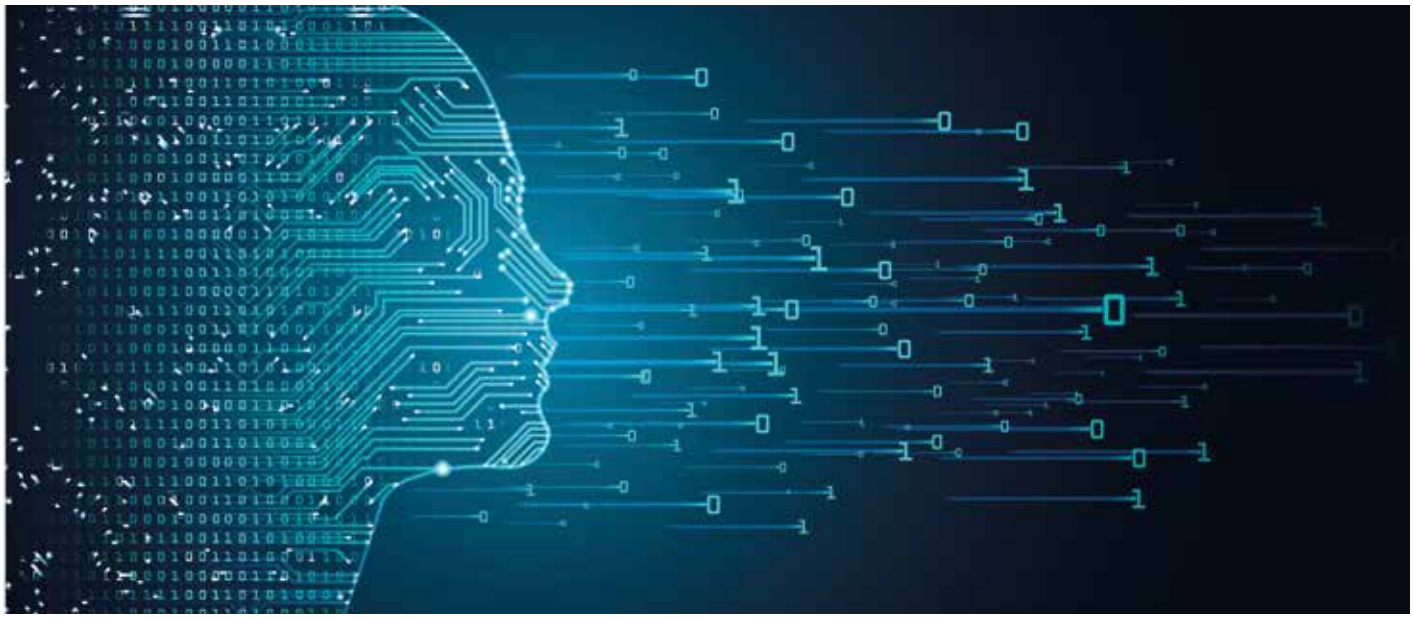
Por tanto, la estrategia comunitaria sobre IA se basa en pilares normativos establecidos en instrumentos como la Comunicación "Inteligencia Artificial para Europa". Estos documentos destacan valores como la dignidad humana, la no discriminación, la transparencia, la rendición de cuentas y el acceso a datos públicos para promover la innovación disruptiva y responsable. El enfoque sostenible alinea el desarrollo del mercado digital europeo con principios éticos y jurídicos elevados.

Opciones regulatorias respecto a la Inteligencia Artificial: la cuestión de la intensidad de la intervención normativa

El debate sobre cómo regular la Inteligencia Artificial (IA) ha generado controversias en diversas jurisdicciones, reflejando diferencias no solo en el enfoque normativo, sino también en el momento y la intensidad de la intervención. Si bien el objetivo común es proteger a ciudadanos y sociedades frente a riesgos inadmisibles, las estrategias adoptadas varían significativamente entre los principales actores globales. La Unión Europea valoró distintas alternativas regulatorias:

- Regulación bajo un modelo tipo "**black letter law**": No regulando o sometiendo la IA a la normativa preexistente sin introducir regulaciones específicas. Aunque esta opción minimiza riesgos regulatorios y facilita la innovación, es insuficiente para responder a retos relacionados con derechos fundamentales y riesgos emergentes, como ya se ha señalado anteriormente.

- **Autorregulación**: Técnica, utilizada frecuentemente en sectores tecnológicos, que permite a las empresas establecer sus propias reglas mediante códigos de conducta, esquemas de certificación o etiquetado voluntario. Aunque reduce las cargas burocráticas e incentiva la innovación, carece de



mecanismos sólidos de supervisión y puede generar desequilibrios en los estándares aplicados a determinados sistemas que introducen riesgos inadmisibles o elevados.

- **Revisión de la normativa sectorial existente**: Incorporar modificaciones en regulaciones preexistentes para incluir la IA podría ser una alternativa viable. Sin embargo, esta fórmula aumenta el riesgo de contradicciones y fragmentación normativa, lo que podría comprometer la seguridad jurídica y la confianza en el sistema.

- **Regulación tipo "hard law" específica**: Un marco normativo específico vinculante permite establecer obligaciones uniformes a nivel comunitario y establecer los niveles de riesgo admisibles. La UE optó por esta vía mediante un Reglamento, que garantiza la unidad del mercado único digital y evita la fragmentación normativa. Este enfoque, convenientemente aquilatado, asegura un estándar uniforme de protección de derechos fundamentales y permite a las empresas escalar en un entorno competitivo y seguro.

Modelo Regulatorio	Descripción	Ventajas	Desventajas
Regulación tipo "black letter law"	Basada en no regular específicamente la IA, aplicando únicamente la normativa preexistente	Minimiza riesgos regulatorios. Facilita la innovación	Insuficiente para abordar retos relacionados con derechos fundamentales y riesgos emergentes
Autorregulación	Permite a las empresas establecer sus propias reglas a través de códigos de conducta, certificaciones o etiquetado voluntario	Reduce cargas burocráticas. Incentiva la innovación	Carece de mecanismos de supervisión robustos. Puede provocar desequilibrios en los estándares aplicados
Revisión de la normativa sectorial	Modificación de regulaciones existentes para incluir disposiciones relacionadas con la IA	Alternativa viable que aprovecha marcos regulatorios existentes	Aumenta el riesgo de contradicciones y fragmentación normativa. Compromete la seguridad jurídica
Regulación tipo "hard law" específica	Marco normativo vinculante con obligaciones uniformes a nivel comunitario, como el Reglamento de la UE sobre IA	Garantiza unidad del mercado único digital. Asegura estándares uniformes de protección de derechos fundamentales	Puede implicar mayores costos de implementación para empresas

La opción de **regulación tipo *hard law*** específica es la adoptada por la UE mediante un Reglamento horizontal, que asegura armonización total y una respuesta sistemática a los retos de la IA.

Este enfoque prioriza la protección de derechos fundamentales, la transparencia y la seguridad jurídica, al tiempo que fomenta la innovación y competitividad en el mercado único digital.

La elección de este modelo refleja un equilibrio entre la necesidad de regulación y la promoción del desarrollo tecnológico en un entorno ético y sostenible.

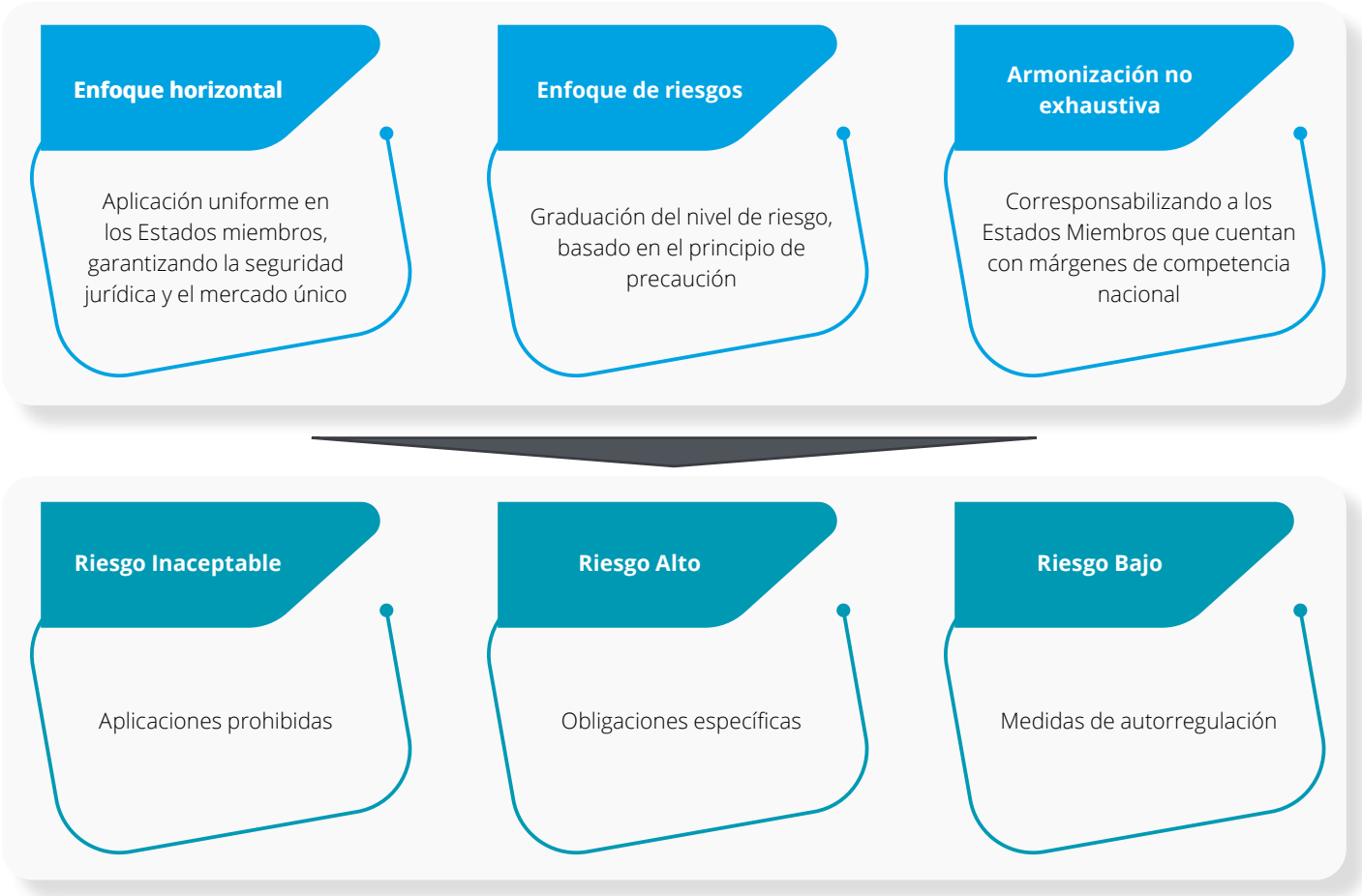
Principales características del modelo regulatorio comunitario de la IA

El modelo regulatorio adoptado por la Unión Europea busca **equilibrar la promoción de un entorno de innovación** con la garantía de una Inteligencia Artificial **alineada con los valores fundamentales de la Unión**. Este enfoque otorga prioridad a los derechos fundamentales, la seguridad y la salud de los ciudadanos y consumidores, configurándose como un marco basado en derechos.

Una de las características esenciales del reglamento es su **enfoque horizontal**. Este instrumento jurídico se aplica de manera uniforme en todos los Estados miembros, promoviendo la seguridad jurídica y el buen funcionamiento del mercado único digital. Sin embargo, el reglamento no persigue una armonización total ni exhaustiva, reservando ciertos márgenes de competencia a los Estados miembros.

El modelo se basa además en un enfoque de riesgos, fundamentado en el **principio de precaución**, que establece un régimen regulatorio adaptado según la clasificación de los sistemas de IA. Este enfoque permite graduar la intensidad y el tipo de intervención en función del nivel de riesgo asociado. Algunas aplicaciones están prohibidas por representar riesgos inaceptables, mientras que otras quedan sujetas a obligaciones específicas o mandatos normativos. Para los sistemas de menor riesgo, se adoptan principalmente medidas de autorregulación, como códigos de conducta.

Este enfoque guarda **coherencia con el RGPD**, que establece un marco de gestión de riesgos y obligaciones relacionadas con la transparencia o los análisis de impacto, así como con otras normativas comunitarias como la "Digital Services Act", que también incorpora un enfoque basado en riesgos, exigiendo modelos de cumplimiento normativo y auditorías externas.



Principios aplicables para conciliar estándares elevados de protección de la salud, seguridad y derechos fundamentales con la innovación responsable

La construcción de un marco jurídico para la Inteligencia Artificial, dotado de estabilidad y resiliencia, requiere un diseño que **combine principios regulatorios con la capacidad de adaptarse a los avances tecnológicos**. Este enfoque permite garantizar un equilibrio entre la protección de los derechos fundamentales y el fomento de la innovación responsable.

Entre los principios relacionados con el diseño y el enfoque regulatorio destaca la **neutralidad tecnológica**. Este principio promueve la creación de normas que no favorezcan ni discriminen tecnologías específicas, asegurando su aplicabilidad a distintas soluciones. La innovación es otro pilar fundamental, orientado a generar un entorno que facilite el desarrollo y la adopción de nuevas tecnologías sin imponer restricciones desproporcionadas. El principio de precaución, por su parte, introduce medidas para mitigar riesgos y proteger a la sociedad frente a aplicaciones de IA cuyo impacto pueda ser incierto o inaceptable.

Además, existen principios orientados a garantizar objetivos estructurales desde el diseño. La **IA humano-céntrica** refuerza la idea de que la tecnología debe estar al servicio de las personas, priorizando el bienestar social y la protección de los derechos fundamentales. La **privacidad**, la **protección de datos** y la **seguridad desde el diseño** exigen que los sistemas de IA integren salvaguardas desde su concepción para prevenir riesgos. Finalmente, el **principio de competencia por defecto** asegura que los mercados relacionados con la IA operen bajo reglas que promuevan la equidad y la ausencia de prácticas monopolísticas.

Estos principios conforman una base sólida para diseñar un marco normativo que responda tanto a los desafíos de la tecnología como a las expectativas sociales y económicas de una IA ética y responsable.

Principios para el diseño de un marco de IA	
Diseño del marco y enfoque regulatorio	**Protecciones legales desde el diseño**
Neutralidad tecnológica	Inteligencia Artificial centrada en el ser humano
Principio de innovación	Privacidad, protección de datos y seguridad por diseño
Principio de precaución	Competencia por diseño

Principios relacionados con el diseño del marco y el enfoque regulatorio

Neutralidad tecnológica

Este principio exige que el diseño normativo no establezca preferencias por una tecnología, salvo que exista una razón de suficiente entidad para restringirla, en todo caso sometida al principio de proporcionalidad y con escrupuloso respeto del contenido esencial de los derechos fundamentales afectados. Con esto se busca que el mercado tecnológico sea suficientemente abierto para equilibrar las posibilidades entre competidores.

El principio es de especial importancia en la construcción de marcos normativos vinculados a las TIC, dado que actúa a modo de barrera frente a la posible discriminación de unas tecnologías frente a otras o de políticas desproporcionadas que impidan su desarrollo.

Principio de innovación

En la carrera por el liderazgo de la IA, la existencia de un marco estable y predecible es clave para promover la innovación y competitividad. Se trata de un aspecto que a nivel europeo resulta especialmente sensible, debido al bajo nivel de inversiones privadas en el sector, así como la rezagada posición en cuanto a creación y financiación de *startups*, lideradas abrumadoramente por Estados Unidos y China. La UE se ha preocupado por destacar la importancia de la innovación para aprovechar las ventajas de la transformación de la economía y, desde esta perspectiva, el principio de innovación representa un punto de anclaje para aquilatar las restricciones que derivan de otros principios.

Principio de precaución

Este principio es entendido de manera habitual siguiendo los parámetros recogidos por la Declaración de Rio de 1992: en caso de que exista riesgo para la salud de las personas o el medio ambiente, la duda o incertidumbre científica no puede constituir un argumento para justificar la inacción. De lo anterior se extrae, a pesar de la posible heterogeneidad conceptual, un denominador común: el principio se dirige a evitar posibles efectos adversos en supuestos de incertidumbre científica (Bourguignon, 2016, 6), que representa además el primero de los presupuestos necesarios para su aplicación. En segundo término, esta incertidumbre se proyecta respecto a un riesgo grave para el medio ambiente o la salud de las personas. Apoyado todo ello en valoraciones científicas fundadas y razonables, con miras a proscribir actuaciones caprichosas o arbitrarias no admitidas por el acervo jurídico común.

Principios dirigidos a recoger protecciones legales desde el diseño

Ya se comentó que reconocer que la tecnología puede incorporar valores es aconsejable para ajustar la respuesta normativa a los retos y riesgos que introduce. También que, aunque se proponen teorías que promueven el determinismo tecnológico, nos decantamos por una posición cauta que admite cierto grado de control y dirección de su desarrollo, que ha de estar en manos de los órganos con legitimidad democrática.

Esta intervención pública para dirigir –así sea parcialmente- la tecnología y alinearla a las exigencias el Estado de Derecho, así como alcanzar objetivos sociales y éticos, exige fórmulas y técnicas específicas. No se trata de renunciar a técnicas clásicas de ordenación (autorizaciones, inspecciones, etc.), ni a otras formas absolutamente asentadas como la autorregulación, sino de complementarlas con algunas que reduzcan las fricciones del binomio tecnología-Derecho. Esto admite y aconseja recurrir a la propia tecnología cómo un instrumento adicional para alcanzar los objetivos y anticipar problemas regulatorios.

Inteligencia Artificial centrada en el ser humano

La promoción de la IA centrada en el ser humano puede situarse sin dificultad en el centro de la política y el marco-ético jurídico, como señaló la estrategia europea de Inteligencia Artificial. El predominio de este enfoque no ha de sorprender dado que la dignidad humana es omnipresente en el discurso social y jurídico contemporáneo, y forma parte del argumentario con el que se abordan, explican o justifican problemáticas muy diversas, de las que no escapa la IA.

En cuanto a las manifestaciones prácticas, el enfoque humano-céntrico exige, desde nuestra perspectiva, que la IA sea una **tecnología dirigida a respetar e incluso promover la libertad de los usuarios**. Las aplicaciones que tenga por fin reducirla, de partida han de estar proscritas. Esta visión humano-céntrica no redunda exclusivamente en el comentado enfoque basado en los derechos fundamentales y la dignidad humana, sino que además conecta con la utilización de esta tecnología para contribuir a resolver los grandes retos que afrontamos como humanidad. En este sentido, se proyecta una visión humanizante y sostenible de la IA, que armoniza con otros principios como el de precaución o innovación.

Finalmente, este enfoque humano-céntrico tiene otra manifestación no menos importante. Dado que se prevé y promueve que la tecnología se incorpore en toda la actividad económica y social, se espera que su **diseño y aplicación se desarrolle con un enfoque humano**. Así, por ejemplo, en ámbitos como el asistencial donde robots pueden brindar apoyo cognitivo o "acompañamiento" a personas mayores, el diseño ha de orientarse al respeto de la dignidad de la persona y evitar su cosificación. Esto nos lleva a proponer, aunque no lo establezca la normativa, un mandato de "humanidad desde el diseño",

a semejanza de otras cláusulas que comentaremos seguidamente.

Privacidad, protección de datos y seguridad por diseño

La privacidad desde el diseño parte por reconocer que la protección de la privacidad, así como la protección de datos y la seguridad informática añadiríamos, no puede asegurarse exclusivamente con carácter reactivo cumpliendo el marco regulatorio. Por el contrario, deben promoverse mecanismos de funcionamiento por defecto. El razonamiento es que la tecnología no sólo debe considerarse como riesgo, sino que debe incorporarse como parte de la solución. De ahí que la privacidad por diseño inicialmente se haya vinculado especialmente con las tecnologías PETs, aunque se diferencian.

Representa un enfoque sistemático que se dirige a proteger la privacidad dentro de las especificaciones o arquitectura misma, por lo que se trata de una aproximación más amplia que, siguiendo al European Data Protection Supervisor, exige que la tecnología se diseñe e implemente a lo largo de su ciclo de vida de forma compatible con los derechos fundamentales y los valores de una sociedad democrática.

Competencia por diseño

El impacto de la IA sobre el funcionamiento competitivo de los mercados no se separa de lo señalado en otros apartados. Así, la economía de datos alimentada por IA puede generar interesantes ganancias competitivas para los consumidores. Puede aportar eficiencia ante cambios en las condiciones del mercado, permitiendo ajustar oferta y demanda. Desde la perspectiva del consumidor, la intensificación de los mercados digitales permitiría acceder a mayores opciones, reduciendo costes de transacción, y podrían además beneficiarse de la reducción de costes derivados de la aplicación de IA a procesos productivos, e incluso ajustar bienes y servicios a las preferencias de los consumidores incrementando su bienestar desde el punto de vista competitivo. Finalmente, puede permitir adoptar mecanismos para incrementar el poder de los consumidores, por ejemplo, agregando la demanda y creando plataformas de compra basadas en modelos de IA.

Esto exige diseñar y aplicar instrumentos que promuevan el **alineamiento de la economía digital con el Derecho de la competencia**. En nuestro caso concreto, que los algoritmos y modelos de Inteligencia Artificial se diseñen orientados a evitar conductas anticompetitivas. Las empresas han de dedicar esfuerzos no sólo a embeber en sus diseños la normativa de la competencia, sino también entender en profundidad cómo funcionan sus algoritmos para desactivar tempranamente posibles comportamientos anticompetitivos no esperados o derivados de decisiones algorítmicas, especialmente en los casos de IA no supervisadas y basadas en cajas negras. Por tanto, se trata de incorporar soluciones orientadas a promover la competencia ex-ante, y no meramente reactivas.

Inteligencia Artificial y Derechos Fundamentales

El uso de la IA afecta a la totalidad de derechos fundamentales

La relación de la IA con los derechos es un punto de obligada referencia, pues esta herramienta tiene capacidad para afectar directa o indirectamente a la **totalidad** de los derechos fundamentales. Ningún fenómeno ha tenido una potencia de impacto tan generalizado en los derechos y libertades. Se ha calificado de tecnología transformadora y disruptiva (Directrices éticas IA, 2019, apdo. 137), con innegables beneficios (medioambientales, económicos y sociales, etc.), pero al mismo tiempo constituye un reto extraordinariamente complejo para poderes públicos y particulares y, por tanto, también para el Derecho. Su aplicación requiere un marco normativo que genere confianza y permita obtener sus posibilidades sin sacrificio o restricción desproporcionada de derechos y libertades de los ciudadanos.

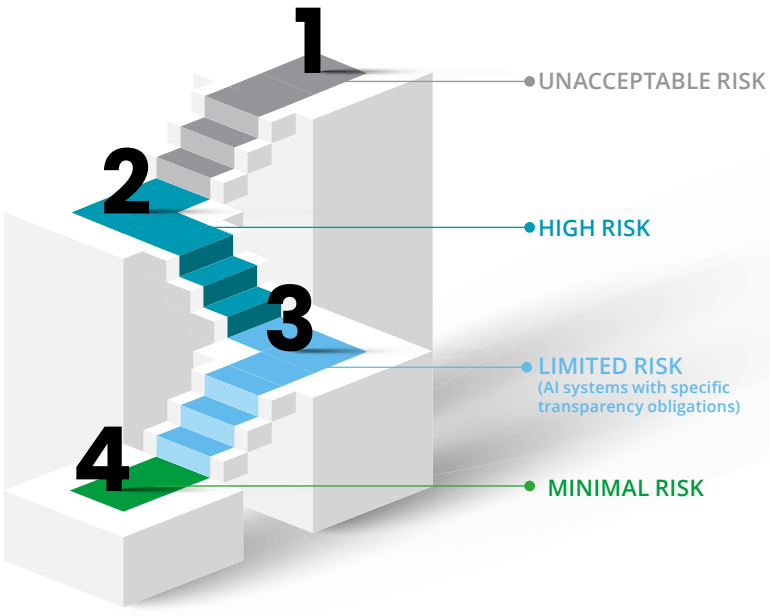
El **Reglamento (UE) 2024/1689 -Ley europea de Inteligencia Artificial**- (en adelante RIA) es el marco jurídico integral para su uso en el ámbito de la Unión Europea. En el art. 1 de la Ley ya se advierte el objetivo de la misma: mejorar el funcionamiento del mercado interior, promover la IA centrada en el ser humano y fiable, garantizar la salud, la seguridad y los derechos fundamentales consagrados en la Carta de Derechos Fundamentales de la UE, incluidos la Democracia, el Estado de Derecho y la protección del medio ambiente.

Son también de referencia imprescindible las **Directrices éticas para una IA fiable**, probadas por el Grupo Independiente de Expertos en IA designado por la Comisión Europea, publicadas en abril de 2019 (https://digital-strategy.ec.europa.eu/es/library/ethics-guidelines-trustworthy-ai).

Categorización por riesgos para los derechos en su aplicación

Una clasificación de interés acerca del impacto de la IA en los derechos fundamentales se refiere a los riesgos de la IA en función de los ámbitos de aplicación y los sistemas empleados. Los clasifica en cuatro categorías según dicho impacto: riesgo inaceptable, alto, limitado y mínimo o ningún riesgo.

Unacceptable Risk: Los riesgos que se consideren una amenaza inaceptable para los derechos fundamentales de las personas quedan prohibidos. El RIA ejemplifica: aquellos sistemas o aplicaciones que puedan manipular el comportamiento humano para reducir o anular la libre voluntad de los usuarios [por ejemplo, los juguetes dirigidos a menores, o los sistemas de puntuación social predictiva o de clasificación (*scoring*) para gobiernos o empresas o algunas aplicaciones de actuación policial]. También se incluyen en este grupo aquellos sistemas biométricos que permiten el reconocimiento de emociones en el lugar de trabajo, o aquellos que se orienten a categorizar personas o identificación



Fuente: digital-strategy.ec.europa.eu/en policies/regulatory-framework-ai (consulta 26 de noviembre de 2024)

biométrica remota en tiempo real con fines policiales en espacio de acceso público (con excepciones). De los ejemplos se deduce que lo que se prohíbe es la IA para la manipulación de la voluntad (esto puede darse en todos los colectivos, pero especialmente los más vulnerables) o la clasificación de personas a través de la discriminación por lectura biométrica.

High risk: En un segundo nivel de riesgo se identifican algunos usos de la IA en la medida en que pueden afectar a salud y seguridad de los componentes de productos, o en sí mismos o la magnitud de afectación a la dignidad humana, la vida privada y familiar, a la protección de datos de carácter personal, la libertad de expresión y de información, la libertad de reunión y de asociación, el derecho a la educación, la protección de los consumidores, los derechos de los trabajadores, de las personas discapacidad, la igualdad -hombres y mujeres y en general cualesquiera forma de discriminación- los derechos de propiedad intelectual, la tutela judicial efectiva, el juez imparcial, los derechos de defensa, la presunción de inocencia o el derecho a una buena administración, con especial protección a los menores (Considerando 48 LEIA). Todos los supuestos calificados como de alto riesgo van a requerir la supervisión humana con el fin de prevenir o reducir los riesgos al mínimo (art. 14) y conllevan también obligaciones adicionales para los proveedores y distribuidores, importadores y responsables, así como el establecimiento de sistemas de gestión de la calidad. Se prevé un sistema de evaluación de impacto de derechos fundamentales en los sistemas de IA de alto riesgo, que se exige tanto a los responsables del despliegue sean entidades públicas o entidades privadas en el que se establecerán los procesos de aplicación, la temporalidad, frecuencia de usos, personas o colectivos afectados, riesgos específicos, medidas previstas de supervisión humana, acuerdos de gobernanza interna y mecanismos de reclamación (art. 27 LEIA).

Limited risk. Aquellos que no puedan causar un perjuicio ni influyan sustancialmente en la toma de decisiones o no perjudiquen los bienes jurídicos señalados en el apartado anterior, si bien pueden suponer algún riesgo para los derechos.

Minimal risk. Estos son casos de irrelevancia en la afectación.

Derechos fundamentales y posible afectación por la IA

Dignidad humana

Como primer principio y límite, las aplicaciones de la IA han de respetar la dignidad humana, que parte del valor extraordinario, único, sagrado, irrepetible e inalienable de cada ser humano, con especial atención para la protección de las personas y grupos de mayor vulnerabilidad (infancia, vejez, discapacidad, enfermedad, pobreza, otros vinculados con la posición de asimetría en el mercado o la identidad en función del contexto -religión, cultura o identidad sexual-). Esto no sólo supone límites al uso de la IA, constituye una prohibición absoluta de los usos incompatibles con la dignidad humana y, al mismo tiempo, en una dimensión objetiva e informadora, un valor al que debe orientarse la IA en su aplicación: la promoción de la dignidad de la persona humana, individualmente y el contexto, considerando a las generaciones futuras y al ecosistema en el que viven y vivirán las personas.

Están absolutamente prohibidos todos los usos contrarios a la dignidad humana (así, por ejemplo, las prácticas de manipulación, explotación o control social) o cualesquiera otras que puedan considerarse tratos inhumanos o degradantes. Desde luego debe velarse porque la IA no comprometa con su uso un modelo social de discriminación, clasificación o control insoportable y deshumanizante, pensando también en las generaciones futuras, sino que, al contrario, sea una herramienta al servicio de los seres humanos y del bien común.

Derecho a la vida e integridad de la persona, física y psíquica

El derecho a la vida puede verse comprometido de modos muy distintos. Desde el empleo de **armas inteligentes, drones operados por algoritmos**, o, entre otros, **selección en fase embrionaria de seres humanos**, sin descartar la aplicación para **procesos de selección (triaje) en la atención sanitaria**. Por eso, es preciso poner especial atención y control en aquellas aplicaciones de la IA que pueden tener repercusiones sobre la vida y la salud (por ejemplo, los componentes de seguridad de productos -robots autónomos en las fábricas o con fines de asistencia o atención personal o en el sector sanitario-). En estos usos donde la IA puede afectar a los bienes vida, salud o integridad los sistemas deben ser fiables y precisos.

Por otra parte, un **modelo de IA para supervisión excesiva** puede constituir un modelo de control deshumanizante, calificable de trato inhumano y degradante. En este punto, la prohibición de la esclavitud y del trabajo forzado son límites absolutos operativos en el uso de la IA.

Libertades y derechos

La privacidad y derecho a la protección de datos se encuentran entre los más expuestos ante la aplicación de la IA que permite la recopilación y tratamiento masivo de datos (información e imágenes) a partir de toda la información generada, localización, cookies que permiten el recuerdo de datos del usuario (archivo de preferencias). Afecta al control de los propios datos y al derecho a la autodeterminación informativa en la medida en que tiene capacidad para elaborar nuevas generaciones de datos o el uso de intermediarios de datos fuera del contexto original. También afecta a la exactitud y actualización de los datos. La Ley de la IA en el art. 2 establece la aplicación del derecho de UE en materia de protección de datos personales, intimidad y confidencialidad de las comunicaciones es de aplicación en relación con los derechos y obligaciones establecidas en la Ley europea de IA.

Una de las posibilidades-riesgo de la IA es la *elaboración de perfiles a través de datos*, pues multiplica la capacidad de conexión de datos, y la incorporación de criterios en la elaboración de dichos perfiles. Un riesgo nuevo es que los sujetos pueden no ser conscientes de que se están elaborando perfiles con sus datos personales sea por particulares, con fines laborales o empresariales, comerciales o de las propias administraciones públicas. Esto afecta genéricamente al derecho a la privacidad y específicamente al derecho a la autodeterminación informativa. Otros problemas tienen que ver con la calidad y actualización de los datos

Por su impacto en la vida de las personas, es de señalar el **riesgo de la utilización de las imágenes disponibles en redes sociales**, o en internet, a veces subidas por terceras personas, que pueden ser utilizadas para la creación de imágenes nuevas, ofensivas o descontextualizadas. Por ejemplo, la publicación de imágenes pornográficas hiperrealistas utilizando rostros de personas (resulta especialmente grave cuando están implicados menores). Esta práctica es delictiva y está previsto su castigo penal, pero también otras cuando afecten a la intimidad o al honor de las personas.

La IA **permite la identificación biométrica de personas físicas**. Es una herramienta muy potente y con extraordinarias ventajas para la identificación segura y rápida a través de imagen o voz. Tiene también posibilidades de uso que permiten el análisis de conductas, movimientos, patrones o emociones. Son los usos sin consentimiento, desproporcionados o con fines ilícitos los que deben ser limitados. Puede *afectar también al derecho a la intimidad y a la propia imagen*. No es lo mismo ser visto que ser captado y registrado; y, con la IA, no es lo mismo ser captado y registrado que ser analizado a través de este seguimiento y análisis biométrico, con capacidad para analizar e interpretar cada gesto o movimiento, la temperatura o la secuencia de movimientos, o el desplazamiento, desde luego muy por encima de lo que permite la mera observación humana. En este sentido el potencial incisivo en la privacidad debe ser advertido como límite a la utilización deshumanizante de la IA.

También relacionado con la privacidad genérica debe proteger frente a usos abusivos de la IA el derecho al **secreto de las comunicaciones**. La IA, como ha quedado demostrado por ej. con el uso del sistema *Pegasus* -en realidad todos sistemas de **espionaje electrónico**- tiene capacidad prácticamente ilimitada para rastrear masivamente las comunicaciones. Sólo con un fin legítimo, previsión legal, cumplimiento de las garantías legales y respeto al principio de proporcionalidad podría utilizarse por los poderes públicos, desde luego no por particulares.

La **libertad de expresión e información** debe ser también destacada entre las que pueden verse interferidas por el empleo de los dispositivos de IA. La utilización de los *mass media* y la aplicación de las nuevas tecnologías, en principio, ha supuesto la ampliación de los espacios de comunicación y ha extendido la plaza pública al espacio virtual. Sin embargo, paradójicamente, no ha conducido a que

el debate público sea más plural y abierto, pues la utilización de algoritmos en los motores de búsqueda puede conducir y, de hecho así se constata en la práctica, a la "*fragmentación de la esfera pública*" y a la creación de las llamadas "*cámaras de eco*" llevando a los participantes al refuerzo y *polarización* de sus propias búsquedas a través de la personalización automática de las búsquedas según el recuerdo de preferencias. De esta forma la IA, así empleada a través del recuerdo de preferencias, refuerza las propias posiciones y más que debate favorece el escoramiento y la hiper confirmación de las opiniones. Por otra parte, debe señalarse su riesgo como factor de impacto para falsear el debate público o para intereses particulares la capacidad de elaboración de contenidos con apariencia de veracidad (p. ej., videos hiperrealistas con programas de IA), prácticamente imposibles de distinguir de información real. Por esto y los factores indicados se relaciona la aparición de la IA con la "*sociedad de la desinformación*".

Las "**burbujas de filtros**" creadas mediante algoritmos son una nueva modalidad de censura para bloquear contenidos. Esto puede ser positivo para la exclusión de determinados mensajes -exaltación del odio o de la violencia, pornografía, xenofobia, etc...- pero no debe pasar inadvertido el riesgo de una utilización sesgada de filtros para la restricción excesiva, inadecuada u orientada (censura) o el refuerzo de otros contenidos. Aparece con la IA un nuevo rostro de la censura o bloqueo excesivo más allá de que la presencia de la IA permite técnicamente procesos de control y vigilancia, que pueden tener un efecto "desalentador" (*chilling effect*) para el ejercicio de la libertad de expresión y de la libertad de información. Este aspecto de la IA tiene un impacto en las libertades de expresión e información que son bienes imprescindibles y a cuidar especialmente en una sociedad plural y democrática. Esto es preocupante en la medida en que la censura y el seguimiento de contenidos está presente sin control jurisdiccional, mediante la incorporación de algoritmos y filtros a través de los operadores intermediarios de internet que supervisan los flujos de información y comunicación. Esto puede tener un "efecto amedrentador" en el uso de la libertad de expresión y de información. Esta última tiene una exigencia añadida a la mera libertad de expresión que es la exigencia de "veracidad".

Hoy en esa nueva "plaza pública" se ejercen también las libertades de **reunión y manifestación y asociación** en nuevos formatos: es posible reunirse y organizarse de manera estable en espacios virtuales de vida social, política y cultural, incluido también el derecho de protesta. Si bien, es posible conectar con espacios antes impensados, también la novedad de la IA abre la puerta al control de los discrepantes, el rastreo de información y la identificación, elaboración de perfiles y control de manifestante, lo que podría llevar a su uso como mecanismo de control social y supervigilancia y por tanto una restricción de los derechos de reunión y asociación.

Otro derecho que debe ser apuntado cuando hablamos de las posibilidades de la IA es **el derecho a la libertad personal y a la seguridad**. La IA es un avance para la persecución de ilícitos y también para su prevención. Véase por ejemplo el sistema *VioGen* que permite detectar el nivel de riesgo y hacer seguimiento y protección a las víctimas de violencia de género para su protección rápida, integral y efectiva. En este y otros delitos, junto a las extraordinarias ventajas para la protección de personas y bienes, se advierte el riesgo de la utilización de sesgos y la discriminación en la *predictive policing*, que puede desencadenar una criminalización o extensión del concepto de grupos de población sospechosa a determinados colectivos. Este factor puede encontrarse no solo en la persecución penal, sino también administrativa-sancionadora, tributaria o en el ámbito laboral.

Del mismo modo, la **libertad personal** puede verse afectada cuando la IA se emplea para la toma de decisiones en relación con las medidas de libertad (prisión provisional o grados de libertad de para los condenados), o cuando se recurre al algoritmo para la valoración de riesgos de los detenidos o presos en casos de puesta en libertad. Estos casos de aplicación de la IA han sido calificados de alto riesgo por su potencial para interferir en la vida y libertad de las personas. Por eso, se requiere transparencia del proceso de decisión y explicabilidad, pues no es posible de otro modo respetar los derechos o conocer los criterios de las decisiones que afectan a la libertad.

Conectado con lo anterior, están los **derechos de los detenidos, los derechos procesales** a la defensa y a un juicio justo, así como a la información y a la presunción de inocencia. Quedan simplemente apuntados, pero el algoritmo en materia de administración de justicia debe ser enfatizado como posibilidad y factor de alto riesgo lo que requiere un marco normativo específico para poder sumar las ventajas de la IA sin que se vea en riesgo el derecho a la tutela judicial efectiva. Así, por ejemplo, el desconocimiento en el funcionamiento del algoritmo en la valoración de pruebas o en el tratamiento de datos para una decisión automatizada deben presentar garantías de no discriminación, transparencia y explicabilidad. La capacidad de la IA para la persecución de delitos al poder trabajar con un volumen masivo de datos y aumentar al máximo la capacidad de análisis ofrecen ventajas evidentes para la investigación, especialmente en ámbitos tan complejos como la lucha contra el terrorismo.

Tampoco las garantías en el **derecho al acceso a la justicia y a los recursos** deben ser desplazados por las "decisiones automatizadas", sin intervención o con escasa intervención humana que permite el pensamiento de contexto evitando el automatismo en respuestas de

reclamación. Se ha valorado su utilidad en decisiones judiciales complejas, incluso en el cálculo automatizado de la pena, como herramienta facilitadora, pero no decisora.

También puede servir en el recurso en material judicial, pero también en otros procesos que puedan verse ayudados en la respuesta (por ejemplo, consumidores, acceso a servicios y prestaciones, etc.), si bien *debe evitarse la "automatización de las decisiones"*. La posibilidad de respuestas mediante algoritmos que arrojen una respuesta de relevancia o no relevancia según un scoring puede afectar a un campo amplísimo de decisiones de servicios, en los que los derechos y libertades afectados pueden verse comprometidos y prácticamente anulada la posibilidad de reclamación -remedio eficaz- de los usuarios y afectados.

La misma advertencia debe realizarse en cuanto a los procesos de **devolución, expulsión y extradición** de extranjeros, que no admite la automatización de los procesos de decisión sin garantías.

En relación con el derecho a la **educación** la IA puede ser una extraordinaria herramienta, por sus posibilidades para el aprendizaje personalizado y el acceso a la información y al conocimiento o el entrenamiento de competencias. Con independencia de las cuestiones pedagógicas o el llamado proceso de alfabetización digital -determinante para el proceso de adaptación y el desarrollo y promoción en un contexto digital- deben señalarse también algunos aspectos de riesgo para los derechos en los procesos educativos que viene siendo señalados: en el acceso (la necesaria atención a cómo se articulan proceso de decisión sobre el acceso y la permanencia en las etapas educativas), evaluación, no discriminación, protección de datos y derecho a la privacidad de los estudiantes.

Como un corolario de cómo la IA puede afectar a las libertades y derechos de los menores, cabe citar la noticia de la demanda de asociaciones de padres en EE.UU. a diversas redes sociales, imputándoles que los algoritmos de las redes manipulan a los menores para ofrecerles contenidos adictivos personalizados y para aprovecharse de las debilidades de carácter de cada sujeto.

Prohibición de no discriminación

En materia de igualdad, lo que viene preocupando es la incorporación de algoritmos con sesgos. La **discriminación** puede afectar a todos los derechos y libertades, pues la igualdad es relacional. En este sentido debe advertirse "el riesgo discriminador de las burbujas de filtros". La no discriminación es una **prohibición absoluta** en el uso de la IA, con especial aviso de riesgo en el caso de las categorías de colectivos históricamente preteridos o socialmente en situación

de riesgo, desventaja, vulnerabilidad o señalamiento social. Por eso, este es uno de los principales puntos referidos cuando se habla de la potencial afectación de derechos en la Inteligencia Artificial.

La **incorporación de sesgos** podría darse de forma directa (sesgos ilícitos -raza, la etnia, la religión, el sexo, la orientación sexual, la edad o la discapacidad-) o indirecta, lo que dificulta su detección. Desde luego tiene consecuencias en todos los campos imaginables: consumo, contratación laboral, modelos de negocio, acceso a seguros o a créditos o, más en general, para la fijación de precios, educación, etc. Además, la capacidad de aprendizaje automático que presenta la IA puede reforzar los prejuicios existentes a través del aprendizaje por estereotipos aprendidos.

Derechos sociales-laborales

En los **sistemas de acceso a los servicios sociales** y a las ayudas, sean públicas o privadas, la decisión por IA facilita el análisis de la información y los requisitos para la concesión. Así, por ejemplo, prestaciones sociales, ayudas a la vivienda, ayudas a colectivos vulnerables, etc. Sin embargo, importa advertir de que en estos casos estamos ante *sistemas de alto riesgo*, no solo por los sesgos que pudieran incorporar los algoritmos, sino también: a) por la necesaria transparencia de los procesos decisorios y el acceso a la reclamación, ya referidos; b) y por la existencia real de una brecha digital que puede desplazar a los colectivos más vulnerables. Esto acentúa los riesgos, pues la decisión sobre estas ayudas puede comprometer el acceso a prestaciones de subsistencia o relevantes para la promoción de las personas (becas a la educación) o las ayudas personales o familiares, o el techo, o a la atención sanitaria o farmacéutica, entre otros bienes. Igualmente existe el riesgo de clasificaciones de ciudadanos atendiendo a criterios (salud, capacidad, situaciones de distinta naturaleza, es un riesgo que debe ser prevenido).

La IA tendrá impacto en las condiciones laborales y los específicos derechos en el **ámbito del acceso, el desarrollo de la actividad laboral, la protección o la promoción**. La IA no suprime los derechos de los trabajadores en la empresa, si bien puede modular su alcance. Cualquier utilización deshumanizante de la IA debe ser objeto de especial atención. Y son de advertir como especialmente obligados a la protección los derechos frente al uso para la clasificación interna de trabajadores y valoración del rendimiento de trabajadores, más cuando puedan ser utilizados para procesos de despido o afectar a las condiciones de trabajo. Conviene estar alerta de los excesos que puedan darse en este ámbito con la aplicación de la IA.

Democracia y derechos de participación política

Ya se ha apuntado anteriormente el riesgo que las burbujas de filtro o las cámaras de eco automatizadas -confirmatorias de las propias opiniones- pueden afectar a la Democracia por su **incidencia en la voluntad política y el condicionamiento del debate**. También la posibilidad de alterar los procesos electorales y referendarios mediante la **elaboración de falsos contenidos o mensajes o videos hiperrealistas** está en el punto de preocupación y la identificación "capilar" mediante IA de los segmentos de población decisivos a los que hacer llegar la influencia en las campañas electorales, a veces en "los votos decisivos". El valor de los datos analizados con IA son auténtico oro electoral. Así mismo la **interactuación de *bots* en los debates** en las redes pueden tener un papel determinante en la intención de voto.

Inteligencia Artificial y responsabilidad civil

Propuesta de Directiva de adaptación de normas de responsabilidad civil a la IA

La Propuesta de Directiva del Parlamento Europeo y del Consejo, relativa a la adaptación de las normas de responsabilidad civil extracontractual a la Inteligencia Artificial, de 28 de septiembre de 2022, buscaba aligerar la carga de la prueba para el demandante en esta materia (art. 1). Las peculiares características de la IA hacen que le sean de difícil aplicación las reglas tradicionales de la responsabilidad extracontractual. En particular, a la víctima puede resultarle particularmente complicada la prueba de la causa del daño. A pesar de ello, en este ámbito se mantenía el sistema de responsabilidad por culpa, aunque mitigando las dificultades con relación a la prueba.

La Propuesta de Directiva proponía dos mecanismos en particular (la exhibición de pruebas, y las presunciones refutables) y un tercero, a valorar para el futuro (el seguro obligatorio).En febrero de 2025, tras la clausura oficial de la Cumbre sobre Inteligencia Artificial de París, la Comisión publicó su programa de trabajo, figurando esta Propuesta de Directiva como una de las que se retirará en un futuro previsible. Esta decisión ha provocado, lógicamente, posiciones encontradas. De todas formas, a continuación se exponen los puntos fundamentales de la Propuesta, porque son indicativos de alguna de las tendencias de regulación planteadas por la doctrina.

Véase una reclamación contra un sistema de IA por haber inducido, supuestamente, al suicidio a un menor de edad: https://www.xataka.com/legislacion-y-derechos/que-sabemos-que-no-primera-denuncia-a-ia-provocar-suicidio y https://s3.documentcloud.org/documents/25248089/megan-garcia-vs-character-ai.pdf

Exhibición de pruebas

Se faculta a los órganos jurisdiccionales nacionales a **ordenar la exhibición de pruebas** de los sistemas de IA de alto riesgo que se sospeche que han causado daños. Esta orden puede dirigirse al proveedor de sistemas de IA, a la persona que se encuentre sujeta a las obligaciones de proveedor (con arreglo a lo dispuesto en la Ley de IA) o a un usuario de la IA. Para que tal solicitud llegue a cursarse es preciso que, quien demanda por daños y perjuicios, haya, por un lado, aportado hechos y elementos probatorios suficientes que acrediten y sustenten la viabilidad de su demanda; y, por otro, que haya realizado previamente otros intentos proporcionados para obtener del demandado esas pruebas.

Además de ordenar la exhibición de pruebas, se faculta a los órganos jurisdiccionales a ordenar las **medidas específicas necesarias para su conservación**. Tanto la exhibición de pruebas, como las medidas de conservación han de limitarse y circunscribirse a lo que sea necesario y proporcionado para sustentar la demanda. En este sentido, la norma prevé que se tengan en cuenta los intereses legítimos de todas las partes, incluidos los terceros afectados. Debiéndose prestar particular atención a lo que suponga revelación de secretos comerciales e información confidencial. En cuyo caso, habrán de tomarse las medidas necesarias para preservar la confidencialidad cuando la prueba se use en el procedimiento.

Nuestro ordenamiento ya contempla las Diligencias Preliminares en el art. 256 LEC, dentro de las cuales encajaría esta figura

Presunciones refutables

La inclusión de este tipo de presunciones, inclinan este sistema hacia la responsabilidad objetiva. De modo que se protege al consumidor, trasladando la carga de la prueba al suministrador o prestador empresario.

Este tipo de presunciones operan como nuestras presunciones *iuris tantum* que, a diferencia de las *iuris et de iure*, admiten prueba en contra.

a) Presunción refutable de incumplimiento del deber de diligencia (art. 3.5)

En caso de que el demandado incumpla con el requerimiento antes señalado de exhibición o conservación de las pruebas que obran en su poder, se introduce la "presunción de incumplimiento del deber de diligencia" en su contra.

b) Presunción refutable de relación de causalidad (art. 4).

En principio corresponde al demandante probar la culpa del demandado (para lo que cuenta con la anterior presunción) y la realidad del daño cuya indemnización solicita y que achaca al sistema de IA. Probar el nexo causal entre la culpa del demandado y el daño sufrido que se dice debido a los resultados producidos o no producidos por el sistema de IA, es más complejo y la Directiva lo que hace es presumir esa relación de causalidad si se dan tres requisitos:

- Que el demandante demuestre la culpa del demandado (pudiendo aplicarse al respecto la presunción anterior).
- Que se pueda considerar razonablemente probable que la culpa ha influido en los resultados producidos/no producidos por el sistema IA.
- Que el demandante demuestre que la información/no información producida por el sistema IA causó daños.

A la hora de aplicar la presunción, la Propuesta de Directiva, distingue si la demanda se ha interpuesto frente a proveedores de sistemas de IA o si, por el contrario, se ha interpuesto frente a los usuarios de estos sistemas; y dentro de los proveedores de sistemas de IA, distingue entre proveedores de sistemas de alto riesgo y de riesgo no elevado. Así:

- Frente a proveedores de sistemas de alto riesgo: la ley da una lista de supuestos cerrados en los que cabe la aplicación de la presunción del nexo causal. En general el listado atiende a criterios de calidad, vigilancia, o empleo de medidas correctoras de la ley de IA. Se establece una excepción en la que no aplica la presunción. Se trata de aquellos supuestos en los que el demandado demuestre que el demandante puede acceder razonablemente a pruebas y conocimientos especializados suficientes para demostrar el nexo causal.
- Frente a proveedores de sistemas de riesgo no elevado: la norma establece como condición específica de aplicabilidad de la presunción, que el órgano jurisdiccional determine que es excesivamente difícil para el demandante demostrar el nexo causal.
- Frente a un usuario de sistemas de IA de alto riesgo. En este caso, el demandante debe probar que el usuario: o bien, no cumplió con sus obligaciones de uso o supervisión de conformidad con las instrucciones adjuntas, o suspendió o interrumpió ese uso; o bien, expuso el sistema de IA a datos de entrada que no eran pertinentes.

Seguro obligatorio (Considerando 31)
En el Considerando 31, la Propuesta de Directiva prevé, para una segunda fase de implementación de la norma, la **posibilidad de crear un seguro obligatorio** para la explotación de determinados sistemas de IA. Se señala lo conveniente que sería prever la revisión de la Directiva transcurridos cinco años desde la finalización de su transposición en los estados miembros, para examinar a la luz de lo acontecido en esos años, la necesidad o no de adoptar normas de responsabilidad objetiva para las demandas contra el operador (siempre que estas no estén ya cubiertas por otras normas de responsabilidad de la Unión, en particular las relativas a producto defectuoso), combinadas con un seguro obligatorio para la explotación de determinados sistemas de IA.

De este modo, transcurrido un tiempo desde la implementación de esta Directiva, se debería recabar información sobre la eficacia de las medidas previstas en la Directiva que se siguen basando fundamentalmente en la responsabilidad subjetiva o por culpa; los cambios tecnológicos y normativos que se hayan producido en la materia en esos años; los riesgos que, para bienes jurídicos importantes como la vida, la salud y la propiedad de terceros, se hayan ido observando y/o materializando en la utilización de productos o servicios basados en la IA; así como, el desarrollo de soluciones por parte del mercado de seguros. A la vista de todo ello, se reevaluaría el efecto e incidencia que la introducción y adopción generalizada de los sistemas de IA hubiera tenido y se vería si lo regulado en esta Directiva es suficiente o conviene dar el paso a una responsabilidad objetiva, con seguro de responsabilidad obligatorio asociado.

Nuevo régimen de productos defectuosos (Directiva 2024/2853)

El 28 de septiembre de 2022, también se presentó la Propuesta de Directiva del Parlamento Europeo y del Consejo, sobre responsabilidad por los daños causados por productos defectuosos. Esa propuesta ha sido aprobada en 2024, constituyendo la actual Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, con la que se deroga la anterior Directiva 85/374/CEE de 25 de julio de 1985. En lo que afecta a Inteligencia Artificial, esta Directiva cubre categorías de productos no contemplados por la anterior, como los derivados de las nuevas tecnologías digitales, incluidos los productos inteligentes y la IA.

La Directiva introduce en su definición de producto el concepto de software, junto al de IA, de modo que, **cuando estos productos actúan de forma defectuosa y generan un daño**, la persona perjudicada puede dirigirse al fabricante responsable para reclamar su resarcimiento.

La IA entra en la categoría de "producto"
El art. 4,1 define el término "producto" como: "*cualquier bien mueble, aun cuando esté incorporado a otro bien mueble o a un bien inmueble o interconectado con estos; incluye la electricidad, los archivos de fabricación digital, las materias primas y los programas informáticos*". Al **incluir los archivos digitales y los programas informáticos** en esta definición ampliada de "producto", quedan incluidos también los sistemas de IA. Muchos sistemas de IA están integrados en bienes tangibles como electrodomésticos, vehículos, o dispositivos médicos. En estos casos, la IA es una parte esencial del producto en su conjunto. De ahí que sea importante que las aplicaciones de IA que se distribuyen como software, ya sea para ordenadores, dispositivos móviles o la nube, sean consideradas productos.

Por otro lado, considerar la IA como un producto permite **regular su uso y distribución de manera similar a otros bienes** y servicios. Los sistemas de IA se crean, desarrollan, comercializan y venden de manera similar a otros productos; están sujetos a contratos de venta, garantías, soporte y actualizaciones, como otros productos en el mercado; y son objeto de transacciones comerciales, como venta de licencias, suscripciones y servicios asociados, al igual que otros productos. De modo que incluirlo en esta regulación otorga una mayor protección a los consumidores.

La IA puede ser un producto defectuoso que genere daños
Si bien la nueva Directiva **no varía lo que entiende y define como producto defectuoso** (un producto es defectuoso si no ofrece la seguridad que legítimamente se espera de él, art. 7,1), sí que incorpora a la **lista no cerrada de factores** que los tribunales deben tener en cuenta a la hora de evaluar el carácter defectuoso del producto, factores como las funciones de autoaprendizaje del producto o la interconexión (art. 7,2, c y d).

El producto se considerará defectuoso si presenta fallos o deficiencias que lo hagan **inseguro o inapropiado para su uso previsto**. Si concurre este supuesto y se demuestra que el defecto causó daños, el responsable deberá indemnizar a las personas perjudicadas. Esto implica para el operador económico (fabricante o desarrollador del sistema de IA), el deber de asegurarse de que el producto cumpla con los estándares de seguridad y calidad establecidos para proteger a los consumidores y usuarios.

Los **daños** que puede producir una IA defectuosa son variados y pueden incluirse en las tres categorías previstas en la Directiva como daños sujetos a indemnización (art. 6,1): daños **personales** (muerte, lesiones corporales y daños a la salud psicológica), daños **materiales**, y los derivados de **destrucción o corrupción de datos**.

El derecho a indemnización que tiene la persona perjudicada abarca toda la **pérdida material**, así como el **daño moral** derivado (art. 6,2).

La persona perjudicada por el daño generado por la IA defectuosa puede dirigirse contra el operador económico responsable
Para el resarcimiento de los daños, la persona perjudicada ha de dirigirse contra el "operador económico responsable". La nueva Directiva ha eliminado la referencia directa al "productor", estableciendo en su lugar un **listado de operadores económicos responsables** de los productos defectuosos (art. 8). En ese listado se incluye al fabricante del producto defectuoso, al **fabricante** del componente defectuoso del producto; y, en caso de que éstos estén establecidos fuera de la UE, el **importador**, el representante autorizado; y, en caso de que estos tampoco estén establecidos en la UE, el **prestador de servicios logísticos** (tramitación de pedidos a distancia o el distribuidor). De esta forma, se garantiza al consumidor el poder dirigirse siempre a un agente en territorio europeo a quien pueda imputarse la responsabilidad. Y todo ello, sin perjuicio de las posibles acciones de regreso que tendrán a su favor quienes han respondido frente a los consumidores.

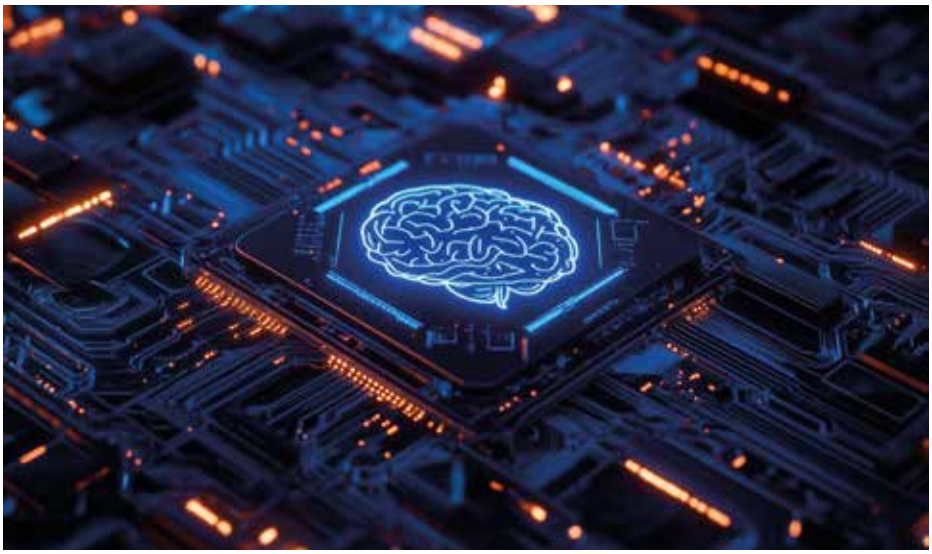
La persona perjudicada por el producto defectuoso puede reclamar los daños ocasionados a **quien haya intervenido en la producción del sistema** de IA causante del daño. En este sentido, estarían incluidos: el fabricante del sistema de IA, el fabricante del hardware (si la IA está integrada en un dispositivo físico) y el desarrollador del software de IA, el integrador de sistemas, el distribuidor o vendedor.

De especial interés en lo que a la IA se refiere, es la referencia que se hace a "cualquier persona física o jurídica que **modifique sustancialmente un producto** fuera del control del fabricante y que posteriormente lo comercialice o ponga en servicio", esta persona se considerará fabricante del producto y, por tanto, responsable.

Se prevé así mismo que, **cuando no pueda identificarse a un operador económico**, también se podrá pedir la responsabilidad a **cualquier proveedor de una plataforma en línea** que permita a los consumidores celebrar contratos a distancia con comerciantes y que no entre en la categoría de operador económico.

Si, a pesar de esta lista establecida de responsables a los que poder dirigirse, la víctima no obtiene indemnización alguna porque ninguna de las personas previstas puede ser considerada responsable o porque los responsables son insolventes o han dejado de existir, la Directiva habilita a los Estados miembros para utilizar los **sistemas nacionales de indemnización sectoriales** previstos o a establecer otros nuevos para indemnizar adecuadamente a los perjudicadas. Recomendando que, preferiblemente, esos sistemas no se financien con ingresos públicos.

Este apartado podrá aplicarse a todos los sujetos que realizan modificaciones, incluidas mejoras y actualizaciones que pueden incluirse en los programas informáticos, siempre que se lleven a cabo fuera del control del fabricante inicial. Y ello, como indica el Considerando 39: "Cuando un producto se modifica sustancialmente y posteriormente se comercializa o pone en servicio, dicho producto se considera un producto nuevo. Cuando la modificación se efectúa fuera del control del fabricante original, se considera un producto nuevo y debería ser posible responsabilizar a la persona que efectuó la modificación sustancial como fabricante del producto modificado". De modo que, para que estemos ante un producto nuevo y su responsable sea el que llevó a cabo los cambios, lo esencial es que la modificación sea sustancial y se haya llevado a cabo fuera del control del fabricante inicial.



En el proceso para reclamar esa responsabilidad se cuenta con los mecanismos de exhibición de pruebas y presunciones refutables

Si bien la nueva Directiva mantiene la necesidad de que la carga de la prueba recaiga en las personas perjudicadas, que deberán probar el daño que alegan haber sufrido, el carácter defectuoso del producto y el nexo causal entre ambos, **se introducen también los mismos mecanismos previstos en la Propuesta de Directiva** sobre responsabilidad civil en materia de IA de 2022, que aligeran la carga de la prueba, aunque con alguna particularidad:

a) Exhibición de pruebas.

La exhibición de pruebas en esta Directiva tiene una doble dirección: por un lado, entra en juego **en favor del demandante**, cuando haya presentado hechos y pruebas suficientes para respaldar la verosimilitud de su demanda de indemnización, e implica que el órgano jurisdiccional frente al que se ha presentado la demanda, exija al demandado que exhiba las pruebas de que disponga; y por otro lado, entra en juego **en favor del demandado**, cuando haya presentado hechos y pruebas suficientes para demostrar su necesidad de pruebas a efectos de oponerse a la demanda de indemnización, e implica la obligación para el demandante de exhibir las pruebas pertinentes que estén a su disposición.

Se ha de garantizar que esta exhibición de pruebas en ambos sentidos, **se limite a lo que sea necesario y proporcionado** y que se tengan en cuenta los intereses legítimos de todos los afectados, incluidos los terceros. Se hace particular hincapié en la protección de la información confidencial y los secretos comerciales en el transcurso del procedimiento judicial y después de este. También se garantizará que esas pruebas

se aporten de manera fácilmente accesible y comprensible, siempre de forma proporcionada en términos de costes y esfuerzo para la parte requerida.

Si bien la nueva Directiva mantiene la necesidad de que la carga de la prueba recaiga en las personas perjudicadas, que deberán probar el daño que alegan haber sufrido, el carácter defectuoso del producto y el nexo causal entre ambos, **se introducen también los mismos mecanismos previstos en la Propuesta de Directiva** sobre responsabilidad civil en materia de IA de 2022, que aligeran la carga de la prueba, aunque con alguna particularidad:

b) Presunciones refutables.

Presunción de carácter defectuoso. Los órganos jurisdiccionales nacionales podrán presumir el carácter defectuoso de un producto cuando: el demandado no haya exhibido las pruebas que se solicitaron; el demandante demuestre que el producto no cumple los requisitos obligatorios de seguridad; o cuando el demandante demuestre que el daño fue causado por un mal funcionamiento manifiesto del producto durante un uso razonablemente previsible o en circunstancias normales.

Presunción de causalidad. Se presume el nexo causal entre el daño y el defecto, cuando se haya comprobado que el producto es defectuoso y el daño causado sea de un tipo compatible normalmente con el defecto en cuestión.

También se presume el **carácter defectuoso del producto o el nexo causal**, o ambos (art. 10,4), cuando, incluso si el demandado cumple con sus obligaciones de revelación de información, resulte excesivamente difícil para el demandante demostrar el carácter defectuoso del producto o el nexo causal, o ambos. De modo que podrá ser suficiente con que demuestre que es probable que el producto es defectuoso o que existe un nexo causal entre el producto y el daño alegado, o ambos.

Con relación a este último supuesto, el Considerando 48 de la Directiva señala que, en estos casos, "imponer el nivel de prueba habitual exigido por el Derecho nacional, que a menudo requiere un alto grado de probabilidad, menoscabaría la efectividad del derecho a indemnización. Por lo tanto, dado que los fabricantes tienen conocimientos especializados y están mejor informados que la persona perjudicada, y a fin de mantener un reparto equitativo del riesgo, al tiempo que se evita una inversión de la carga de la prueba, debe exigirse al demandante que demuestre, cuando sus dificultades se refieran a la prueba del carácter defectuoso del producto, únicamente que es probable que el producto fuera defectuoso, o, cuando las dificultades del demandante se refieran a la prueba del nexo causal, únicamente que el carácter defectuoso del producto es una causa probable del daño".

Exoneración de responsabilidad del operador económico por los "riesgos de desarrollo"

La Directiva prevé una serie de supuestos en los que, de concurrir, los operadores económicos no serán responsables de los daños causados por un producto defectuoso. Entre ellos se encuentra el supuesto conocido como "**exención por riesgos de desarrollo**", es decir que, si el operador económico demuestra "*que el estado objetivo de los conocimientos científicos y técnicos en el momento en que el producto fue introducido en el mercado, puesto en servicio o durante el período en el que el producto estaba bajo el control del fabricante no permitía detectar el carácter defectuoso*" quedará exonerado de responsabilidad.

El Considerando 52 señala que esta causa de exoneración busca "un reparto equitativo de los riesgos" y, al mismo tiempo, con ello se favorece el progreso de la ciencia y de la técnica. Aunque, añade el Considerando 59, esta posibilidad de que el operador económico pueda eludir su responsabilidad puede considerarse en algunos Estados miembros que limita indebidamente la protección de las personas físicas. De ahí que se admita que los Estados miembros puedan establecer excepciones a dicha posibilidad mediante la introducción de nuevas medidas o la modificación de medidas existentes para ampliar la responsabilidad en tales situaciones a tipos específicos de productos si se considera necesario, proporcionado y justificado por objetivos de interés público, como son el orden público, la seguridad pública y la salud pública.

Si se hace uso de esta excepción deberá notificarse a la Comisión que, a su vez, informará a los demás Estados miembros y emitirá un dictamen no vinculante sobre esas medidas o modificaciones propuestas. El dictamen que se elaborará tras una estrecha cooperación entre el Estado miembro de que se trate y la Comisión, y teniendo en cuenta los puntos de vista de los demás Estados miembros. Para dar tiempo a la emisión del dictamen, el Estado miembro que proponga esas medidas o modificaciones, deberá suspender su aplicación durante seis meses a partir de su notificación a la Comisión, a menos que esta emita un dictamen antes.

Esta cláusula ha de ser **matizada en el ámbito de la IA**, ya que, como señala la doctrina, los operadores económicos que implementan sistemas inteligentes, en muchos casos, siguen manteniendo el control sobre ellos después de lanzarlos al mercado, por lo que, si detectan defectos a *posteriori*, tienen el deber de proporcionar parches informáticos o actualizaciones de seguridad.

Plazo de caducidad de la responsabilidad por producto defectuoso

El art. 17 de la Directiva establece **un plazo de caducidad de 10 años** para la responsabilidad por producto defectuoso, a menos que la persona perjudicada haya interpuesto, entre tanto, una acción contra un operador económico responsable. Dicho plazo comenzará a contar a partir de la fecha de introducción en el mercado del producto defectuoso que haya causado el daño o de su puesta en servicio; o en el caso de productos modificados sustancialmente, a partir de la fecha de comercialización o puesta en servicio de dicho producto tras su modificación sustancial. El Considerando 57 justifica este plazo señalando que "dado que los productos envejecen con el tiempo y que se desarrollan normas de seguridad más estrictas a medida que avanza el estado de la ciencia y la tecnología, no sería razonable responsabilizar a los fabricantes durante un período de tiempo ilimitado del carácter defectuoso de sus productos".

Pero ese plazo se **amplía a veinticinco años** en los casos en que los síntomas de una lesión corporal sean, según pruebas médicas, de aparición lenta (art. 17, 2).

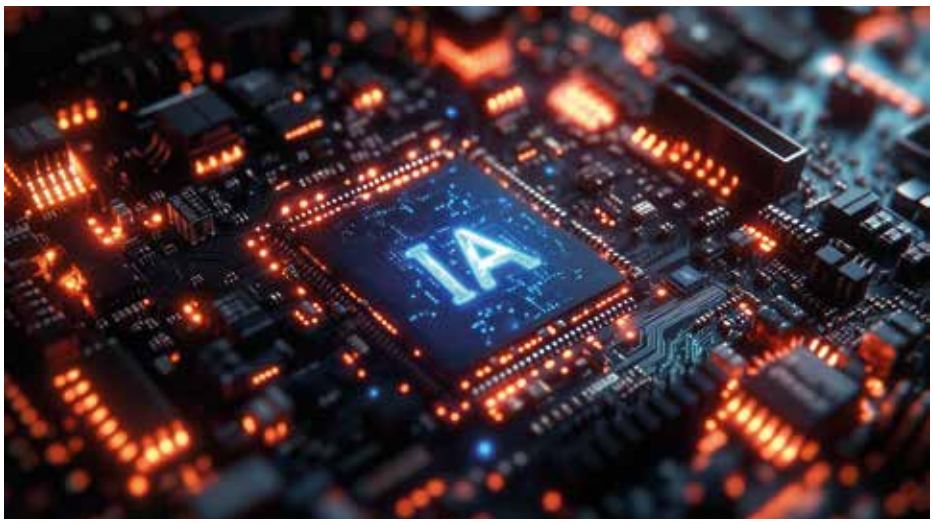
El impacto de la Inteligencia Artificial en el buen gobierno corporativo: una encrucijada entre derecho, tecnología digital y ética, con una visión estratégica y de controles adecuados

El **aumento del uso de la IA generativa en la mayoría de las áreas clave de la actividad empresarial**, desde las ventas y el marketing hasta el análisis contable y financiero, así como en cuestiones de cumplimiento normativo, significa que incluso los niveles más altos de gobierno corporativo se verán afectados. Esta transformación, que ya está ocurriendo, tendrá un **impacto directo en los deberes legales y éticos de las "mentes directoras"** encargadas de establecer el "nivel de diligencia máxima" y que son responsables del éxito de la empresa.

Los directores tendrán que considerar hasta qué punto deberían o deben **utilizarse las herramientas de Inteligencia Artificial en la toma de decisiones estratégicas** y el seguimiento como parte integral de la gobernanza. Deben tenerse en cuenta cuestiones seleccionadas de **gobernanza y supervisión del cumplimiento normativo** en una serie de jurisdicciones clave que creemos que van a evolucionar debido a la mayor disponibilidad y uso de la IA.

Hay que valorar qué debería hacer la alta dirección para garantizar que las herramientas de IA utilizadas por los agentes de la empresa sean **seguras y funcionales**, especialmente considerando evitar impactos negativos en los accionistas, trabajadores, comunidades sociales, consumidores, mercados y otras partes interesadas. ¿Qué deberían hacer para evitar y/o limitar el riesgo relacionado con la IA en la medida de lo posible, desde una perspectiva tanto legal como ética para **gestionar mejor el éxito corporativo** y su socio clave, la **reputación corporativa**?

Puede ser que la IA obligue al elemento humano en el gobierno corporativo a estar **mejor preparado** y **más comprometido** que nunca, y que los directores tengan que volver a las bases de diligencia para cumplir de manera convincente sus deberes de vigilancia, lealtad, buena fe y transparencia. durante sus periodos de gestión.



Una panorámica de los deberes de los directivos en el uso de la IA

La regulación pionera de la Unión Europea

En función del riesgo que supone el uso de la IA por parte de las empresas y su alta dirección en los próximos años, ciertas leyes y marcos legales nacionales actuales pueden reutilizarse o ampliarse para abordar los riesgos de la IA. Además, es posible que necesitemos nuevas leyes específicas a nivel nacional para centrar la atención y la rendición de cuentas de la alta dirección y de gobierno corporativo en la IA.

Sin duda, la Ley de IA de la UE es un primer gran paso global, y será un ejercicio bueno y necesario durante los próximos años para ver cómo impacta tanto a las empresas europeas como a la mayoría de las otras grandes multinacionales debido al "efecto Bruselas".

El Reglamento europeo de IA, entre otras cosas, exige **requisitos para los sistemas de IA de alto riesgo** (Cap. 2). Los sistemas de IA que afecten negativamente a los intereses públicos en materia de salud, seguridad y derechos fundamentales se considerarán de alto riesgo (esto implicará un alto nivel de supervisión de la junta directiva). Estas nuevas normas de la UE exigen que las **empresas que utilizan sistemas de IA de alto riesgo** tengan un **sistema de gestión de riesgos** establecido, implementado, documentado y mantenido en relación con los sistemas de IA de alto riesgo (art. 9) y que cuenten con un **sistema de gestión de calidad organizacional** que cumpla con el art. 17.

El asistemático enfoque estadounidense: posibles cambios en la nueva era Trump

En comparación con la Unión, Estados Unidos ha adoptado un enfoque menos integral y más fragmentado para gestionar la IA y tiende a centrarse en el **desarrollo**

Art. 17.1 Reglamento IA: "Los proveedores de sistemas de IA de alto riesgo establecerán un sistema de gestión de la calidad que garantice el cumplimiento del presente Reglamento. Dicho sistema deberá consignarse de manera sistemática y ordenada en documentación en la que se recojan las políticas, los procedimientos y las instrucciones e incluirá, al menos, los siguientes aspectos:

a) una estrategia para el cumplimiento de la normativa, incluido el cumplimiento de los procedimientos de evaluación de la conformidad y de los procedimientos de gestión de las modificaciones de los sistemas de IA de alto riesgo;

b) las técnicas, los procedimientos y las actuaciones sistemáticas que se utilizarán en el diseño y el control y la verificación del diseño del sistema de IA de alto riesgo;

c) las técnicas, los procedimientos y las actuaciones sistemáticas que se utilizarán en el desarrollo del sistema de IA de alto riesgo y en el control y el aseguramiento de la calidad de este;

d) los procedimientos de examen, prueba y validación que se llevarán a cabo antes, durante y después del desarrollo del sistema de IA de alto riesgo, así como la frecuencia con que se ejecutarán;

e) las especificaciones técnicas, incluidas las normas, que se aplicarán y, cuando las normas armonizadas pertinentes no se apliquen en su totalidad o no cubran todos los requisitos pertinentes establecidos en la sección 2, los medios que se utilizarán para velar por que el sistema de IA de alto riesgo cumpla dichos requisitos;

f) los sistemas y procedimientos de gestión de datos, lo que incluye su adquisición, recopilación, análisis, etiquetado, almacenamiento, filtrado, prospección, agregación, conservación y cualquier otra operación relacionada con los datos que se lleve a cabo antes de la introducción en el mercado o puesta en servicio de sistemas de IA de alto riesgo y con esa finalidad;

g) el sistema de gestión de riesgos que se menciona en el artículo 9;

h) el establecimiento, aplicación y mantenimiento de un sistema de vigilancia postcomercialización de conformidad con el artículo 72;

i) los procedimientos asociados a la notificación de un incidente grave con arreglo al artículo 73.

j) la gestión de la comunicación con las autoridades nacionales competentes, otras autoridades pertinentes, incluidas las que permiten acceder a datos o facilitan el acceso a ellos, los organismos notificados, otros operadores, los clientes u otras partes interesadas;

k) los sistemas y procedimientos para llevar un registro de toda la documentación e información pertinente;

l) la gestión de los recursos, incluidas medidas relacionadas con la seguridad del suministro;

m) un marco de rendición de cuentas que defina las responsabilidades del personal directivo y de otra índole en relación con todos los aspectos enumerados en este apartado".

de estándares bien reconocidos, mejores prácticas y **normas no vinculantes** que la dirección de la empresa (según la industria involucrada) debe seguir e integrar en sus marcos de gobernanza y cumplimiento de IA.

Los principales **impulsores de la supervisión** estadounidense de la gobernanza de la IA son, por parte del sector público, la Securities Exchange Commission y el Departament of Justicie; y por parte de la iniciativa privada, a través de demandas colectivas de accionistas de empresas estadounidenses que se relacionan o utilizan IA.

La acción clave hasta ahora, al menos a nivel federal, fue la publicación por parte de la Administración Biden de la **Orden Ejecutiva sobre el Desarrollo y Uso Seguro y Confiable de la Inteligencia Artificial** en octubre de 2023, que se centra principalmente en la implementación de directrices de desarrollo de la IA. Mientras tanto, el AI Safety Institute, formado durante el mandato de Biden, tiene como objetivo equilibrar la innovación con estrictas salvaguardias regulatorias. Esto marcó un paso clave en la definición de la regulación y la rendición de cuentas de la IA en múltiples sectores, enfatizando un enfoque de "todo el gobierno" para abordar tanto las oportunidades como los riesgos asociados con la IA. Específicamente, la EO ordenó a las agencias federales que abordaran varias áreas centrales de manera no legislativa, como la gestión de modelos de IA de doble uso, la implementación de protocolos de prueba rigurosos para sistemas de IA de alto riesgo, la aplicación de medidas de rendición de cuentas, la salvaguardia de los derechos civiles y la promoción de la transparencia. a lo largo del ciclo de vida de la IA. Estas iniciativas están diseñadas para mitigar los posibles riesgos de seguridad y seguir los valores democráticos al tiempo que fomentan la confianza pública (Brookings Institute, "1 año después, ¿cómo ha cumplido la Orden Ejecutiva de IA de la Casa Blanca sus promesas?", noviembre de 2024).

Sin embargo, de cara a la **segunda administración Trump**, podría haber cambios significativos en la política de IA de Estados Unidos. La campaña de Trump ya había mencionado planes para revisar y potencialmente derogar la amplia orden ejecutiva sobre IA de la administración Biden, creyendo que relajar las restricciones regulatorias podría promover la innovación, fundamental para competir en la creciente carrera de IA con China. La IA podría incorporarse al enfoque "Hacer que Estados Unidos vuelva a ser grande" en muchas áreas de los negocios y la tecnología globales. El equipo de Trump sugiere que fortalecer la infraestructura de IA del país y fomentar un entorno favorable a la innovación es clave para la seguridad nacional y el dominio económico, y sus políticas probablemente enfatizarán el rápido avance de la IA al tiempo que abogarán por la desregulación en áreas donde la supervisión regulatoria se percibe como una innovación asfixiante. Todos estos enfoques regulatorios generales deben ser seguidos de cerca por los consejos de administración corporativos, ya que podrían tener impactos importantes en las operaciones comerciales y en muchas áreas de toma de decisiones, especialmente, pero no limitado a, las empresas españolas activas tanto en Europa como en los EE. UU. (así como en China).

Contexto regulatorio en China

China también está tomando medidas importantes para regular muchos aspectos de la IA, incluida la redacción de una ley integral al estilo europeo. Los avances legales importantes recientes (2023) cubren diversas políticas y regulaciones, incluidas las Medidas provisionales para la administración de servicios de IA generativa (Medidas de IA generativa), las Disposiciones administrativas sobre la síntesis profunda de servicios de información basados en Internet (Disposiciones de síntesis profunda), las Medidas de prueba para la Revisión Ética de las Actividades Científicas y Tecnológicas (Medidas de Revisión Ética) y las Disposiciones Administrativas sobre Recomendación de Algoritmos para Servicios de Información de Internet (Disposiciones de Recomendación de Algoritmos). (Reed Smith LLP, "Navegando por las complejidades de la regulación de la IA en China", agosto de 2024).

Sobre la base de estos avances regulatorios (que creemos que podrían tener impactos en las estrategias y políticas de los consejos de administración y, por supuesto, en los sistemas de cumplimiento normativa), China está trabajando actualmente en una **ley de IA completa** y más al estilo de la Unión Europea. Aunque todavía está en forma de borrador, existe una disposición de penalización propuesta para violaciones graves de IA, que puede ordenar acciones correctivas extensas, la incautación de cualquier ingreso ilícito y la imposición de multas masivas basadas en el volumen de negocios corporativo, ordenar la inmediata suspensión de la línea de negocio u operaciones correspondiente, la pérdida de permisos o licencias comerciales pertinentes y, lo que es más importante, multas significativas a los gerentes encargados de la supervisión de la IA y a otro personal gestor directamente responsable. Además, a dichos gerentes se les puede prohibir durante un período de tiempo desempeñarse como directores, supervisores, personal directivo superior y personas responsables de la seguridad de la IA en las empresas afectadas (CSET, "Proyecto de Ley de Inteligencia Artificial de la República Popular China", Capítulo VIII Responsabilidad Jurídica, Art. 82 Disposiciones Generales sobre Sanciones Administrativas, mayo de 2024).

La integración de la IA en los sistemas y estrategias de gobierno corporativo y cumplimiento normativo

Recomendaciones británicas sobre IA y nuevas tecnologías

La **Guía Oficial para el Código de Gobierno Corporativo 2024** del Reino Unido publicada recientemente (FRC, actualizada en diciembre 2024) incluye ahora un par de referencias clave a la IA a las que los órganos directivos deben prestar atención, en primer lugar, en lo que respecta a la estrategia corporativa.

Los órganos gestores deben tener una comprensión sólida de **cómo se crea y mantiene el valor a lo largo del tiempo** para orientar las estrategias y los modelos de negocio hacia un futuro sostenible [...]. (s.11A FRC). Además, la comprensión del órgano de administración sobre cómo se desarrollan, gestionan y sostienen todas las **fuentes materiales de valor** (esto a menudo incluye tecnologías clave como la Inteligencia Artificial y la propiedad intelectual) es cada vez más relevante para comprender el desarrollo de la actividad.

Luego, la guía plantea una serie de **preguntas clave** para los directores a la hora de reflexionar e implementar estrategias corporativas relacionadas con la IA y otras tecnologías emergentes:

- ¿Somos conscientes de que la empresa utiliza tecnologías emergentes (como podría ser IA), p.ej., en la presentación de informes?
- ¿Nuestra cadena de suministro utiliza tecnologías emergentes? De ser así, ¿cómo?
- ¿Somos conscientes de los desafíos y beneficios de las tecnologías emergentes para darnos una ventaja competitiva?
- ¿Cómo evaluaremos y mediremos el impacto de nuestras decisiones en el desempeño financiero, el valor para los accionistas y el impacto en las partes interesadas clave?

Además, hay un nuevo enfoque en la **IA y la ciberseguridad como componente clave** de los sistemas de control y los marcos de cumplimiento normativo de una empresa (par. 272).

Si bien el órgano de administración decide qué tipo de **controles** implementar en sus protocolos, estos podrían incluir, entre otros, controles sobre:

- Riesgos que podrían amenazar el modelo de negocio, el desempeño futuro, la solvencia o liquidez y la reputación de la empresa (es decir, riesgos principales).
- Informes externos que son sensibles a los precios o que podrían llevar a los inversores a tomar decisiones de inversión, ya sea en la empresa o no.
- Fraude, incluida la anulación de controles.
- Riesgos de la información y la tecnología, incluida la ciberseguridad, la protección de datos y las nuevas tecnologías (por ejemplo, la Inteligencia Artificial).

Estos criterios de recomendación del Reino Unido sobre la evaluación y gestión de riesgos de los órganos de administración encajarán significativamente con la implementación de la Ley de IA de la Unión Europea descrita anteriormente y sus requisitos, especialmente los extensos requisitos del sistema de gestión de calidad que cubren a los proveedores de sistemas de IA de alto riesgo.

Sin duda, todo esto combinado afectará, en el día a día, la estructuración, la gestión y el seguimiento de la IA y de otras cuestiones tecnológicas por parte del órgano de administración.

Recomendaciones de Buen Gobierno españolas sobre IA en la Sociedad cotizada

El enfoque británico anterior encaja con las orientaciones del Código de Buen Gobierno Corporativo español de 2020,

donde en la Recomendación 42 se habla del papel general de gestión de riesgos del **comité de auditoría** corporativo, que incluye un enfoque tecnológico (incluida la IA) como parte de un enfoque completo y proceso de auditoría preciso:

> "Que, además de las previstas en la ley, correspondan a la comisión de auditoría las siguientes funciones: 1. En relación con los sistemas de información y control interno: a) Supervisar y evaluar el proceso de elaboración y la integridad de la información financiera y no financiera, así como los sistemas de control y gestión de riesgos financieros y no financieros relativos a la sociedad y, en su caso, al grupo -incluyendo los operativos, tecnológicos, legales, sociales, medioambientales, políticos y reputacionales o relacionados con la corrupción- revisando el cumplimiento de los requisitos normativos, la adecuada delimitación del perímetro de consolidación y la correcta aplicación de los criterios contables".

Además, la Recomendación 45 se refiere a la **política de gestión de riesgos**, que podría cubrir nuevos desarrollos de la IA:

> "Que la política de control y gestión de riesgos identifique o determine al menos: a) Los distintos tipos de riesgo, financieros y no financieros (entre otros los operativos, tecnológicos, legales, sociales, medio ambientales, políticos y reputacionales, incluidos los relacionados con la corrupción) a los que se enfrenta la sociedad, incluyendo entre los financieros o económicos, los pasivos contingentes y otros riesgos fuera de balance. b) Un modelo de control y gestión de riesgos basado en diferentes niveles, del que formará parte una comisión especializada en riesgos cuando las normas sectoriales lo prevean o la sociedad lo estime apropiado. c) El nivel de riesgo que la sociedad considere aceptable. d) Las medidas previstas para mitigar el impacto de los riesgos identificados, en caso de que llegaran a materializarse. e) Los sistemas de información y control interno que se utilizarán para controlar y gestionar los citados riesgos, incluidos los pasivos contingentes o riesgos fuera de balance". (CNMV, Código de buen gobierno de las sociedades cotizadas, junio 2020, páginas 40 y 42)

Estas recomendaciones tienen que ser implementada a la luz de otras iniciativas españolas sobre la IA, como el Real Decreto 817/2023 (8 de noviembre), que establece un **entorno controlado de pruebas para el ensayo del cumplimiento del Reglamento europeo de IA**.

Todo lo anterior interactúa con **recomendaciones clave adicionales en las Recomendaciones españolas**. Por ejemplo, en la sección 1.2.1 sobre composición y tamaño del consejo, la mejor práctica reconocida proporcionada por el Instituto de Consejeros-Administradores (ICA) implica asegurar una mezcla de talento, una combinación complementaria de experiencias y conocimientos/know-how entre los miembros. del consejo, que incluye experiencia financiera, industrial, operativa, tecnológica, internacional, regulatoria y de buen gobierno (ICA. Guía Práctica del Consejo de Administración, enero 2024).

Claves de la Guía Corporativa de Estados Unidos

El mismo despliegue de una mayor sensibilidad hacia las cuestiones relacionadas con la IA también ha afectado al mundo empresarial y de cumplimiento normativo de Estados Unidos.

De hecho, la principal **guía oficial del gobierno federal** de Estados Unidos para órganos de administración y equipos ejecutivos respecto del diseño y gestión de programas de cumplimiento corporativo (publicada por el DOJ) se acaba de actualizar para incluir la consideración de cuestiones de IA:

"Gestión de riesgos emergentes para garantizar el cumplimiento de la ley aplicable":

- ¿Tiene la empresa un proceso para identificar y gestionar los riesgos internos y externos emergentes que podrían afectar la capacidad de la empresa para cumplir con la ley, incluidos los riesgos relacionados con el uso de nuevas tecnologías?
- ¿Cómo evalúa la empresa el impacto potencial de las nuevas tecnologías, como la Inteligencia Artificial (IA), en su capacidad para cumplir con las leyes penales?
- ¿Está la gestión de riesgos relacionados con el uso de la IA y otras nuevas tecnologías integrada en estrategias más amplias de gestión de riesgos empresariales (ERM)?
- ¿Cuál es el enfoque de la empresa en materia de gobernanza con respecto al uso de nuevas tecnologías como la IA en su negocio comercial y en su programa de cumplimiento?
- ¿Cómo está frenando la empresa las posibles consecuencias negativas o no deseadas que resulten del uso de tecnologías, tanto en su negocio comercial como en su programa de cumplimiento?
- ¿Cómo mitiga la empresa el potencial de uso indebido deliberado o imprudente de las tecnologías, incluso por parte de miembros de la empresa?
- En la medida en que la empresa utilice IA y tecnologías similares en su negocio o como parte de su programa de cumplimiento, ¿existen controles para monitorizar y garantizar su confiabilidad y uso de conformidad con la ley aplicable y el código de conducta de la empresa?

- ¿Existen controles para garantizar que la tecnología se utilice únicamente para los fines previstos?

- ¿Qué punto de referencia de la toma de decisiones humana se utiliza para evaluar la IA?

- ¿Cómo se monitoriza y se hace cumplir la rendición de cuentas sobre el uso de la IA?

- ¿Cómo forma la empresa a sus empleados en el uso de tecnologías emergentes como la IA? (Departamento de Justicia de EE. UU., División Penal, "Evaluación de programas de cumplimiento corporativo", septiembre de 2024, págs. 3-4).

Sobre la base de un cuestionamiento tan centrado y detallado, el futuro **escrutinio de las decisiones de la alta dirección** con respecto a la IA por parte de las autoridades públicas será cada vez más riguroso y sofisticado. Los consejos de administración están al tanto de ello y deben actuar en consecuencia, e inmediatamente si aún no lo han hecho.

Estas directrices estadounidenses también se aplicarían directamente a las **empresas españolas y otras empresas extranjeras que cotizan en los mercados de capitales estadounidenses** (como Telefónica, Santander, BBVA, Mapfre, etc.; véase análisis más adelante).

Enfoque ético en la Unión Europea
Los responsables de la toma de decisiones que gestionan el uso de la IA para un grupo empresarial deben ser conscientes y responsables de la manera en que se realizan las compensaciones fundamentales de responsabilidad entre la auditabilidad y la minimización y la notificación de los impactos negativos****** (revisar), y deben revisar continuamente la idoneidad de la decisión resultante para garantizar que se puedan realizar los cambios necesarios en el sistema cuando sea necesario. Para hacerlo correctamente, el Grupo de Expertos de Alto Nivel de la Unión Europea en IA recomienda considerar diferentes modelos de gobernanza para ayudar a lograrlo. Por ejemplo, la presencia de un **experto** o **junta ética** interna y/o externa (y específica del sector) podría ser útil para resaltar áreas de conflicto potencial y sugerir formas en que ese conflicto podría resolverse mejor. También es útil realizar consultas y debates significativos con las partes interesadas, incluidas aquellas que corren el riesgo de verse afectadas negativamente por un sistema de IA (Comisión Europea, abril de 2019, Directrices éticas para una IA confiable - Grupo de expertos de alto nivel en Inteligencia Artificial, pág. 20).

El Grupo de Expertos de Alto Nivel de la UE sobre IA también recomienda potencialmente establecer una "**junta de revisión ética de la IA**" o un mecanismo similar para trazar la responsabilidad general y las prácticas éticas para la IA, incluidas las áreas grises potencialmente poco claras (Comisión Europea, abril de 2019, Directrices éticas para personas confiables).

Esperamos nuevas directrices significativas de la Unión Europea y de sus grupos de expertos durante los próximos años, que deberían ampliarse significativamente y proporcionar expectativas mucho más detalladas en cuanto a la gestión del gobierno corporativo de las cuestiones de IA. Esto será de gran importancia para las empresas españolas y las principales empresas globales que operan en el mercado interior europeo.

Advertencias al órgano de administración y al equipo directivo sobre el uso de la IA

El comienzo de investigaciones y multas: Securities and Exchange Commission (SEC) y Federal Trade Commission (FTC)
A nivel mundial, esperamos que las agencias reguladoras y fiscales nacionales presten mucha atención a los usos de IA en los próximos años. Es absolutamente necesario que los órganos ejecutivos presten atención a las lecciones aprendidas de esta ola de acciones de aplicación de la ley (y también de los litigios privados relacionados).

Por ejemplo, importantes reguladores empresariales y financieros de Estados Unidos, como la SEC, han comenzado a realizar importantes esfuerzos para **limitar los abusos de la IA en la divulgación corporativa** a los inversores. Estos son los pensamientos que ahora se repiten con frecuencia del actual director de la SEC, Gary Gensler, en su "Horas de oficina de Gary Gensler: fraude y engaño en la IA" (noviembre de 2024):

> "Desde la antigüedad, los malos actores han encontrado nuevas formas de engañar al público. Con la Inteligencia Artificial, los estafadores tienen una nueva herramienta que explotar. Joe Kennedy, sin embargo, lo dijo mejor: "La Comisión hará la guerra sin cuartel a cualquiera que venda valores mediante fraude o tergiversación".

Tras un aumento significativo en los últimos años en la atención regulatoria a las tecnologías blockchain y las criptomonedas y las "ofertas iniciales de monedas" relacionadas, la SEC ha prestado mucha atención a las cuestiones de IA, especialmente incluyendo lo que ellos llaman "**lavado de IA**". Esta práctica por parte de una empresa o de órganos directiva consiste fundamentalmente en hacer declaraciones falsas y engañosas sobre su supuesto uso de la IA. Un ejemplo de esto son los casos Delphia y Global Predictions Inc.

El comienzo de investigaciones y multas: la SEC y el caso Delphia.
En marzo de 2024, la SEC anunció que había resuelto los cargos contra dos asesores de inversiones, Delphia (USA)

Inc. y Global Predictions Inc., por hacer **declaraciones falsas y engañosas sobre su supuesto uso de IA**, quienes pagaron 400.000 dólares en sanciones civiles.

Según la orden de la SEC contra Delphia, de 2019 a 2023 la firma con sede en Toronto hizo **declaraciones falsas y engañosas** en presentaciones, en un comunicado de prensa y en su sitio web sobre su uso de Inteligencia Artificial y aprendizaje automático que incorporaba datos de clientes en su inversión. Delphia afirmó que "pone a trabajar datos colectivos para hacer que nuestra Inteligencia Artificial sea más inteligente para que pueda predecir qué empresas y tendencias están a punto de triunfar e invertir en ellas antes que los demás". Estas declaraciones eran falsas y engañosas porque Delphia en realidad **no tenía las capacidades de Inteligencia Artificial y aprendizaje automático que afirmaba** (SEC de Estados Unidos, Comunicado de prensa 2024-36, 18/03/2024).

El comienzo de investigaciones y multas: la Federal Trade Commission (FTC) y la "Operación de cumplimiento de IA"

Además de la SEC, la **Comisión Federal de Comercio de Estados Unidos** (FTC) también está tomando medidas importantes contra las empresas que hacen un mal uso o caracterizan erróneamente la IA en sus actividades comerciales y, a lo largo de 2024, los consejos de administración estadounidenses se han dado cuenta de esta nueva realidad de más exigente escrutinio regulatorio, que sólo promete intensificarse en los próximos años.

Con la "Operación Cumplimiento de IA", anunciada en septiembre de 2024, la FTC anunció cinco acciones coercitivas contra operaciones que utilizan **exageraciones de IA** o **venden tecnología de IA que puede usarse de manera engañosa e injusta**. La FTC actuó contra múltiples empresas que han dependido de la IA como una forma de potenciar conductas engañosas o injustas que dañan a los consumidores, como parte de su nueva redada de aplicación de la ley. Los casos incluyen acciones contra una empresa que promociona una herramienta de Inteligencia Artificial que permitía a sus clientes crear reseñas falsas, una empresa que afirma vender servicios de "abogados de Inteligencia Artificial" y varias empresas que afirman que podrían utilizar la Inteligencia Artificial para ayudar a los consumidores a ganar dinero a través de tiendas en línea. (Comunicado de prensa de la FTC de EE. UU., 25 de septiembre de 2024, "La FTC anuncia medidas enérgicas contra las afirmaciones y esquemas engañosos de IA")

El **Presidente de la SEC**, Gary Gensler, comentó:

> "Descubrimos que Delphia y Global Predictions anunciaron a sus clientes y clientes potenciales que estaban usando IA de ciertas maneras cuando, en realidad, no lo hacían. Hemos visto una y otra vez que cuando aparecen nuevas tecnologías, pueden generar rumores entre los inversores, así como afirmaciones falsas por parte de quienes pretenden utilizar esas nuevas tecnologías. Los asesores de inversiones no deberían engañar al público diciendo que están utilizando un modelo de IA cuando no es así. Este lavado de IA perjudica a los inversores" (SEC de Estados Unidos, Comunicado de prensa 2024-36, 18/03/2024).

En este sentido, el director de la División de Cumplimiento de la SEC, Gurbir S. Grewal, también brindó **consejos clave a los órganos de gestión de las empresas**:

> "A medida que más y más inversores consideran el uso de herramientas de IA para tomar sus decisiones de inversión o deciden invertir en empresas que afirman aprovechar su poder transformador, nos comprometemos a protegerlas contra quienes se dedican al 'lavado de IA'. Como dejan claro las acciones de aplicación de hoy [...] si afirma utilizar IA [...] debe asegurarse de que sus declaraciones no sean falsas o engañosas. Y los emisores públicos que hacen afirmaciones sobre su adopción de IA también deben permanecer atentos a declaraciones erróneas similares que pueden ser importantes para las decisiones de inversión de los individuos".

Acciones de responsabilidad y class actions planteadas por inversores

El riesgo del uso de IA generativa también será crucial para los consejos de administración a medida que gestionen los riesgos de divulgación y cumplimiento de las leyes de valores, especialmente para las empresas que cotizan en bolsa en los mercados de EE. UU., donde la **demanda colectiva** de valores de Estados Unidos es una potente herramienta de rendición de cuentas si el emisor realiza declaraciones de hechos falsas o engañosas u omite cierta información sobre el uso de la IA.

En tales casos podrían existir (dependiendo del alcance y los tipos de daños) **acciones de cumplimiento** tanto públicas como privadas. Las corporaciones que cotizan en bolsa y sus directorios ahora deben tener un alto nivel de transparencia para evitar la responsabilidad en virtud de la disposición general antifraude de las leyes federales de valores de Estados Unidos y deben ser sinceros sobre los riesgos generativos de la IA. Por lo tanto, la dirección debe centrarse en los riesgos que asume, garantizando una debida diligencia en su

actuación, proporcionando una divulgación de calidad, y reduciendo así los riesgos de responsabilidad y cumplimiento (para todas las partes involucradas: al menos las propias empresas, sus altos directivos, intermediarios financieros, auditores y contadores, y principales accionistas).

Es de destacar que las demandas colectivas de valores de los accionistas federales de Estados Unidos relacionadas con la IA están aumentando y se prevé que seguirán creciendo significativamente. La creciente importancia de la IA en los modelos de negocios de muchas empresas puede conducir a un aumento de dichas reclamaciones en el futuro. Esto debería ser una importante llamada de atención para los consejos de administración de todo el mundo. (Universidad de Stanford-Cornerstone Research, Presentaciones de demandas colectivas sobre valores, Evaluación de mitad de año de 2024, p. 5).

Riesgos de la litigación estadounidense para las compañías españolas cotizadas en EE.UU.

Los órganos de administración en España y en todo el mundo harían bien en aprender de esta primera ola inicial de litigios corporativos y de inversión relacionados con la IA. Vale la pena señalar que las siguientes empresas españolas tienen ADRS/acciones que cotizan en bolsa en los Estados Unidos y, por lo tanto, ellas y sus consejos/directores ejecutivos son objetivos principales para este tipo de demandas colectivas relacionadas con la IA (aunque otras empresas españolas que no cotizan en los EE. UU. podrían ser demandados por ventas en EE.UU., por captar capital de forma privada allí, por responsabilidad filial en EE.UU. o por convertirse en contratistas federales):

Los casos de "Operación Cumplimiento de IA" anunciados hasta ahora se basan en una serie de otros casos recientes de la FTC que involucran afirmaciones sobre IA, que incluyen:

- Automators, otro esquema de tiendas online;
- Career Step, una empresa que supuestamente utilizó IA para convencer a los consumidores de que se inscribieran en una formación profesional falsa;
- NGL Labs, una empresa que supuestamente afirmó utilizar IA para moderar una aplicación de mensajería anónima que comercializaba ilegalmente para niños;
- Rite Aid (farmacias), que supuestamente utilizó tecnología de reconocimiento facial de IA en sus tiendas sin garantías razonables; y
- CRI Genetics, una empresa que supuestamente engañó a los usuarios sobre la exactitud de sus informes de ADN, incluidas afirmaciones de que utilizó un algoritmo de Inteligencia Artificial para realizar coincidencias genéticas.
- (Comunicado de prensa de la FTC de Estados Unidos, 25 de septiembre de 2024, "FTC Announces Crackdown on Deceived AI Reclamaciones y esquemas").

Las demandas colectivas de valores estadounidenses relacionadas con la IA son aquellas en las que la empresa demandada (1) desarrolla modelos de IA, (2) fabrica productos utilizados en la infraestructura de IA o (3) utiliza modelos de IA con fines comerciales; y, además, las acusaciones están relacionadas con la IA, o con tergiversaciones o faltas de divulgación de los riesgos asociados con el uso de la IA. Las presentaciones relacionadas con la IA incluyen aquellas con acusaciones relacionadas con el aprendizaje automático y la conducción autónoma, entre otros. (Universidad de Stanford-Cornerstone Research, Presentaciones de demandas colectivas sobre valores, Evaluación de mitad de año de 2024, p. 21)

Composición de los órganos de gestión y deberes de información en la adopción de decisiones estratégicas

¿Estamos preparados para el consejero IA?

Recientemente ha surgido la cuestión de abogar por nombrar aplicaciones de IA o alguna forma de robots o aplicaciones de IA como miembros del órgano de gestión o del equipo ejecutivo.

La Unión Europea consideró recientemente el impacto de esto en un importante estudio de antecedentes sobre la IA y la responsabilidad civil, y finalmente concluyó que la "personalidad jurídica de la IA" podría entenderse de varias maneras. El estudio de la UE consideró la "personalidad electrónica" como el reconocimiento de los derechos individuales al sistema de IA y decidió que era inadmisible y que no hay razones para considerar a las aplicaciones de IA como iguales a los seres humanos, ni como una categoría intermedia entre una máquina y el ser humano.

Sin embargo, al considerar la posibilidad de conceder a los sistemas de IA algún tipo de **personalidad jurídica corporativa** o similar, el estudio de la UE concluyó que puede haber casos en los que podría ser sensato atribuir al sistema de IA alguna forma de personalidad jurídica, ya sea configurándola como alguno de los tipos de personalidad jurídica admitidos, o creando un tipo de "personalidad jurídica" específico. Esto entra en conflicto con estudios anteriores de la UE que recomendaban negar la concesión de la personalidad electrónica, por lo que parece que el debate continuará. Sin embargo, sospechamos que, dependiendo de los avances tecnológicos, eventualmente se reconocerá algún tipo de personalidad electrónica limitada y funcional para el uso

de la IA a nivel de órganos de gestión o de otra manera, con el fin de garantizar la responsabilidad y la rendición de cuentas como objetivos de política pública.

Hay que valorar los **riesgos significativos que supondría** el considerar a un programa de IA generativa como un "sujeto jurídico" de algún tipo, que pueda formar parte de un consejo como miembro independiente y autónomo, tanto desde el punto de vista de la responsabilidad, especialmente por los daños y perjuicios que se deriven de decisiones erróneas o tomadas sin la diligencia debida, así como desde un punto de vista reparador y ético más general (Forbes, "¿Puede la IA convertirse en su próximo director ejecutivo?" S. Odilov, 01/11/2024).

Es importante señalar que algunas empresas ya se han adentrado en lo desconocido al designar a programas de IA como miembros de sus órganos de toma de decisiones:

- El gigante chino de tecnología de juegos NetDragon Websoft nombró al programa de IA "Tang Yu" como su director ejecutivo en agosto de 2023 y, como resultado, la cotización de sus acciones aumentó casi un 10%, superando los mil millones de dólares.

- En una línea similar, la empresa de bebidas polaco-colombiana Dictador nombró al programa/robot de Inteligencia Artificial, "Mika", como su director ejecutivo.

Si bien todavía es demasiado pronto para analizar su impacto real en el éxito de sus empresas, la realidad es que la IA ya está presente en estas áreas.

¿Estamos preparados para el consejero IA?

La formación de los consejeros "humanos" en IA.

Pase lo que pase con las aplicaciones de IA, los directores, consejos de administración y equipos ejecutivos potenciales necesitarán cada vez más **directores y funcionarios competentes en Inteligencia Artificial** y, en general, en tecnología en la combinación de talentos. Por supuesto, esto depende de la industria de cada empresa: en algunas industrias, sería suficiente tener al menos un director básicamente competente en IA, mientras que en otras, especialmente en las industrias de tecnología, la mayoría o todos los directores deberían tener directores y funcionarios con diferentes niveles. de preparación para la IA.

A medida que las iniciativas de *hard y soft* law relacionadas con la IA adquieran cada vez más relevancia en los próximos años, creemos que crecerán las expectativas relativas a tener al menos un **número de directores y ejecutivos competentes en IA**, y que las mejores prácticas o estándares de referencia incluirán la formación en IA y otras nuevas tecnologías. Por ejemplo, el Código de Gobierno Corporativo 2024 del Reino Unido incluye un principio sobre la composición y evaluación de la junta directiva, señalando que "la junta y sus comités deben tener una combinación de habilidades, experiencia y conocimiento".

¿Estamos preparados para el consejero IA?

El uso de IA por los gestores para la adopción de decisiones con la diligencia debida

Los sistemas de IA generativa brindan a las juntas directivas y ejecutivos **ayuda y asesoramiento** en aspectos tales como (entre otros aspectos) finanzas y valoración financiera (por ejemplo, para transacciones de fusiones y adquisiciones y IPO), contabilidad, auditoría de cuentas, riesgos de ciberseguridad y otros tipos de gestión de riesgos, servicio al cliente, ventas y marketing, atención sanitaria, seguimiento de los derechos humanos y de políticas ESG en las cadenas de suministro globales, entretenimiento y producción creativa, redacción, moda, diseño industrial de productos, desarrollo de software, almacenamiento óptimo y distribución de energía desde fuentes renovables, gestión de oleoductos y gasoductos e incluso la prestación de asesoramiento y servicios jurídicos. Esta es solo una lista parcial, ya que, al parecer, vemos numerosas herramientas nuevas de Inteligencia Artificial para la alta dirección que aparecen continuamente. Son pocas las industrias o actividades comerciales que

permanecen sustancialmente inmunes a posibles cambios disruptivos basados en el despliegue de sistemas generativos de IA.

El hecho es que se avecinan innumerables cambios y desafíos debido a la IA generativa (y muchos ya están aquí), que afectarán tanto el gobierno corporativo como el cumplimiento de muchas maneras. Se plantearán importantes problemas de responsabilidad para las empresas y otros actores del comercio y de las finanzas globales. En consecuencia, si bien parece que se esperará mucho más de los directores y funcionarios corporativos, el mejor enfoque es que tengan una **comprensión clara tanto de la tecnología como de sus deberes como gestores** (diligencia, buena fe, lealtad, y otros).

Cada responsable de la toma de decisiones corporativas debe **estar adecuadamente preparado, capacitado, plenamente informado y asistido** en la medida necesaria para tomar decisiones materiales bien fundadas para la empresa a largo plazo, y supervisar sus operaciones y riesgos inherentes a los programas de IA. Los órganos de administración deben saber qué es la IA y cómo todos estos productos de IA generativa podrían y/o deberían utilizarse productivamente para algunos aspectos de las actividades del grupo empresarial.

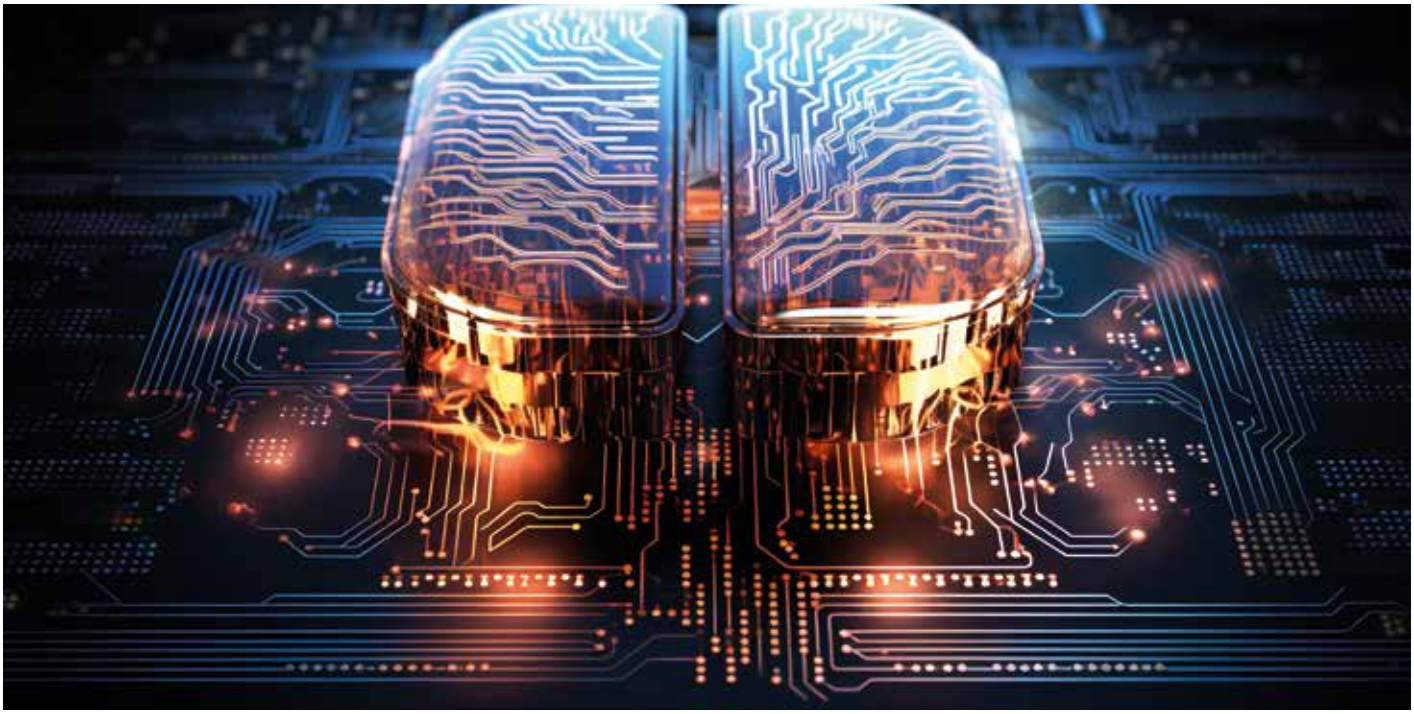
La **sensibilidad hacia los problemas de la IA** ya está muy extendida: según una encuesta realizada por edX y citada por Forbes este año (2024), el 49% de los directores ejecutivos estadounidenses cree que al menos parte de sus funciones se automatizarán a través de la IA (Forbes). Los expertos de McKinsey Consulting han visto que, si bien la IA todavía es incapaz de realizar procesos confiables de toma de decisiones estratégicas, ya es capaz de generar un alto rendimiento y predicciones precisas para áreas "tácticas" como el análisis de negocios y la inteligencia de diagnóstico. Esto significa que, si bien todavía le cuesta diseñar estrategias generales que involucren el futuro de la institución ("dirigir la empresa como un todo") debido a las habilidades blandas inherentes necesarias para ello, como la creatividad, la empatía, la inteligencia emocional, etc., ya es útil para predecir el resultado probable de las tendencias del mercado, dando una ventaja a quienes toman decisiones humanas.

Recomendaciones a los Consejos de Administración para el uso estratégico de IA limitando el riesgo de incurrir en responsabilidad

Para hacer frente a la ola global de nuevas leyes, regulaciones y mejores prácticas reconocidas relacionadas con la IA los directores y gestores deben estar **preparados intelectual y críticamente** con las tareas de supervisión y estratégicas que desempeñan debido a sus puestos clave de responsabilidad en la corporación y actuar de buena fe.

Tanto desde una perspectiva legal y de mejores prácticas, como desde una perspectiva práctica, las decisiones corporativas clave relativas a la gobernanza (transacciones materiales, cumplimiento y actividades diarias) seguirán dependiendo en buena medida de la **supervisión humana** y de la adopción de decisiones con la **diligencia debida**.

Los gestores **agregarán IA** a un conjunto ya amplio de herramientas (financieras, técnicas y legales, etc.) para ayudar con dicha toma de decisiones, pero al final, el "elemento humano" seguirá siendo el principal a la hora de decidir si se procede con una transacción clave determinada. La máxima, respaldada por la jurisprudencia



en muchas jurisdicciones, de "No existen los directores falsos..." es más cierta que nunca si comenzamos a observar la evolución esperada del uso de la IA por parte de los consejos de administración de las empresas.

En definitiva, el **elemento humano cara a cara** seguirá siendo e fundamental importancia. Las herramientas de IA ciertamente pueden ayudar con ciertos tipos o aspectos de la diligencia debida, por ejemplo, y seguirán evolucionando en su utilidad y capacidades, pero la base de la decisión seguirá y debe seguir informada por cadenas humanas de confianza que unen a personas y grupos corporativos. y otras partes interesadas en la red global de comercio y finanzas. La IA y otras tecnologías nuevas deberían fortalecer esas cadenas de confianza, tanto internamente (dentro del equipo ejecutivo de una organización) como externamente (con contrapartes transaccionales, como socios de empresas del grupo).

La necesidad de garantizar la identidad de los sujetos que operan en el entorno digital como un derecho fundamental: una identidad digital soberana

La identidad digital

La normativa comunitaria (Reglamento núm. 910/2014 relativo a la identificación electrónica y los servicios de confianza para las transacciones electrónicas, modificado recientemente por el Reglamento 2024/1183, denominado eIDAS -*electronic IDentification, Authentication and trust Services*-) no contiene una definición legal de identidad digital. Lo que sí se define, en su art. 3.1), es la identificación electrónica.

Se considera **identificación electrónica** "el proceso consistente en utilizar los datos de identificación de la persona en formato electrónico que representan de manera única a una persona física o jurídica o a una persona física que representa a otra persona física o a una persona jurídica".

En sentido, y desde la perspectiva de la norma europea, la identidad digital sería el resultado del proceso de identificación electrónica, en la medida en que permite **representar de manera única a una persona física o jurídica**.

¿Existe un derecho a la identidad digital? ¿qué tipo de derecho es? ¿quiénes son sus titulares?

La conciencia de identidad -la necesidad de ser nombrado- es una característica del ser humano y que le diferencia del resto de seres vivos. Uno es alguien con nombre e identidad en relación con otros, con la sociedad. Lo es en la familia, en el grupo, en su pueblo, en su nación …

La **identificación es por tanto un derecho fundamental** que el individuo posee y que puede exigir al otro y al poder público: el derecho fundamental a que le garantice y reconozca su identidad y se le permita hacerlo frente a terceros.

A lo largo de la historia se han ido sucediendo **formas distintas de garantizar** que uno es quién dice ser: el hijo o el marido de alguien, ciudadano de tal o cual pueblo o nación. El primer modo de garantizar la identidad es a través del testimonio o reconocimiento de quienes saben quién es, de quienes nos "reconocen" –la familia, los vecinos…-; pero en la medida en los individuos comienzan a desplazarse desde su lugar de origen hacia otras tierras, van y vienen, la identificación se problematiza y se hace necesario encontrar fórmulas que garanticen la identificación de quién uno dice ser.

El **advenimiento del mundo virtual o digital** ha acentuado este problema, suscitando nuevas dificultades que, cada día más, requieren que fórmulas nuevas que permitan a las personas –físicas y jurídicas, también– asegurar y asegurarse que son quienes dicen ser cuando operan en el mundo tras las pantallas. En particular, se hace imprescindible garantizar la identidad frente al riesgo de suplantación en contexto donde ese riesgo es especialmente alto.

La generalización y profundización del uso de redes sociales, del comercio digital, el trabajo *on line*, el uso los servicios financieros digitales, de los pagos virtuales, del acceso *on line* a los servicios públicos… exigen **procedimientos que garanticen identificar quién está tras la pantalla**.

IA e identidad digital: riesgos de suplantación y uso de la biometría
La aparición de la IA generativa y las posibilidades que su uso suscita hacen aún más necesario garantizar la identidad -y la no identidad- de quien ahí aparece y sí quien parece ser en el mundo virtual es chatbot o el producto de una IA.

La IA permite la aparición de nuevas formas de suplantación y ha hecho que aumenten exponencialmente los riesgos en relación con la protección de la identidad.

En particular, la generación de imágenes o vídeos falsos a partir del uso de la imagen real plantea retos jurídicos que aún están pendientes de respuesta.

Otro de los aspectos en los que la IA plantea retos que requieren respuestas desde el derecho es el uso de esta tecnología para identificar a los individuos a través del avance de las técnicas de **biometría**. Ello plantea por un lado:

a) problemas jurídico-constitucionales desde la perspectiva del derecho fundamental a la intimidad y a la libertad y, por otro lado,

b) oportunidades nuevas precisamente para garantizar el acceso seguro al mundo virtual, al permitir precisamente la identificación del usuario

La IA como fenómeno nuevo aumenta exponencialmente los riesgos de vulneración del derecho a disponer de una identidad digital.

La IA como herramienta para la mejor protección de la identidad digital en el entorno digital.

Respuesta del ordenamiento jurídico a la identidad digital
Desde el punto de vista normativo, en la UE, el punto de partida lo supuso el

Reglamento (EU) núm. 910/2014, relativo a la identificación electrónica y los servicios de confianza para las transacciones electrónicas en el mercado interior (en adelante, **eIDAS I**, acrónimo de "*electronic Identification, Authentication and trust Services*").

Tras 10 años de vigencia, teniendo en consideración los avances que se han producido en la tecnología digital, la norma había quedado desfasada y no había sido suficiente para establecer un sistema de identidad electrónica europeo verdaderamente seguro, efectivo y eficiente, necesario para adaptarse a los cambios tecnológicos habido en esta materia. Esto justificó la aprobación del Reglamento (EU) núm. 2024/1183, de Parlamento y del Consejo de 11 de abril de 2024 (en adelante, **eIDAS II**).

El Reglamento eIDASII contiene importantes reformas en el marco normativo europeo que aborda la regulación de la identidad digital, en general y, en particular, a la incidencia que tiene la IA en este ámbito.

Respuesta del ordenamiento jurídico a la identidad digital: la no presencia en el entorno digital
Un segundo aspecto conectado con el derecho a disponer de una identidad digital es el derecho de las personas físicas **a la no presencia en el entorno digital**. A diferencia de lo que ocurre en el mundo real, donde el derecho a tener una identidad es más que dudoso que incluya el derecho a no ser identificado, la presencia en el mundo virtual es resultado de la decisión voluntaria de los individuos. Es cierto, no obstante, que el registro de datos personales que llevan a cabo los poderes y servicios públicos hacen casi imposible una existencia 100% al margen del mundo virtual. En este punto, la defensa de este derecho se busca a través de las normas de **protección de datos de carácter personal**, cuyo fundamento radica en el derecho fundamental a la intimidad.

Respuesta del ordenamiento jurídico a la identidad digital: la no presencia en el entorno digital
En relación con el **uso de seudónimos**, relativamente habitual en los entornos

digitales, el Reglamento eIDAS II, al igual que ya lo hiciera anteriormente el eIDAS I se refiere expresamente al uso de pseudónimos.

Se prevé que "sin perjuicio de las normas específicas del Derecho de la Unión o nacional que exijan a los usuarios identificarse o de los efectos jurídicos que el derecho nacional contemple para los seudónimos, *no se prohibirá la utilización de pseudónimos escogidos por los usuarios*". Por el contexto, y teniendo en cuenta las referencias que se contienen en los considerandos al uso de seudónimos, se puede concluir que no se trata de transacciones económicas, sino a cualquier tipo de interactuación en el mundo virtual.

Este régimen de no prohibición configura, por tanto y en la práctica, **un derecho al uso de seudónimos**, dentro de los límites que determinen las normas que sí obliguen a identificarse.

La presencia digital en el ámbito de los servicios públicos

Por su parte, y en relación con la actuación del ciudadano ante las Administraciones públicas, en principio, el art. 14.1 de la Ley 39/2015, de Procedimiento Administrativo Común, reconoce a las personas físicas el *derecho* a "*elegir* en todo momento si se comunican con las Administraciones Públicas para el ejercicio de sus derechos y obligaciones a través de medios electrónicos o no". Es un ámbito importante del alcance del derecho a no existir digitalmente, que está lógicamente delimitado por la obligación de actuar digitalmente.

Quedan fuera de este derecho las personas físicas que ejerzan "actividad profesional para la que se requiera colegiación obligatoria, para los trámites y actuaciones que realicen con las Administraciones Públicas en ejercicio de dicha actividad profesional", incluidos los "notarios y registradores de la propiedad y mercantiles".

Sin embargo, este derecho tiene un alcance limitado en la medida en que el art. 14.3 prevé que, reglamentariamente, se podrá *establece la obligación de relacionarse a través de medios electrónicos* para determinados procedimientos y para ciertos colectivos de *personas físicas* que, por razón de su capacidad económica, técnica, dedicación profesional u otros motivos quede acreditado que tienen acceso y disponibilidad de los medios electrónicos necesarios".

Teniendo en cuenta que la actuación por medios electrónicos requiere la previa *identificación digital*, las AAPP entran de lleno en el ámbito de la identidad digital y en la necesidad, por tanto, de disponer para los ciudadanos y resto de personas jurídicas de herramientas tecnológicas que faciliten dicha identificación

Hasta la fecha las AAPP han dispuesto para ello de herramientas como "Clave digital" o "Certificado digital" que facilita la Fabrica Nacional de Moneda y Timbre. Son instrumentos que permiten una identificación digital segura para el ciudadano ante las AAPP. Sin embargo no permiten disponer de una identidad digital fácilmente accesible, interoperable y transfronteriza. Ese es un paso, que la tecnología actual permite pero que las AAPP aún no disponen, lo que supone un reto de primera magnitud en la digitalización de los servicios públicos.

La **cartera europea de identidad digital** (EUDI Wallet, por sus siglas en inglés), que regula el Reglamento eIDASII, busca precisamente cumplir la misión de garantizar la accesibilidad y seguridad que requiere la identificación digital, tanto para actuar en el ámbito privado, como muy especialmente, en el ámbito de los servicios públicos.

En la práctica el ejercicio del derecho a elegir –y por tanto a elegir no actuar digitalmente ante las AAPP- no será fácil. En el caso del acceso a los servicios públicos las dificultades se centran, en particular, en dos ámbitos que atañen a todo ciudadano: el de los servicios sanitarios y el del cumplimiento de las obligaciones tributarias. En el ámbito privado, sus límites vienen determinados por las limitaciones a hacer de pagos en efectivo, o visto desde el otro lado, la obligación de acudir a medios de pagos distintos del efectivo, y por tanto y mayoritariamente, controlados por la operativa bancaria, fuertemente digitalizada.

En el caso de las personas jurídicas, al menos en el ámbito de sus relaciones con las Administraciones públicas, el art. 14.2 de la Ley 39/2015 obliga a relacionarse electrónicamente con las AAPP, sin hacer distinción de tipos ni dimensión. Esta idea refuerza el argumento de la vinculación de la regulación de la identidad digital con ciertos derechos fundamentales en el caso de las personas físicas y la diferente posición que en este punto tienen las personas jurídicas. Éstas disponen de un menor nivel de protección de sus datos "personales". Lo cual justifica la ausencia de un reconocimiento un *derecho a la no presencia digital* en el ámbito público.

Los riesgos de suplantación y de uso no autorizado de nuestra identidad y nuestros datos, en particular en el caso de la IA

Más allá de la responsabilidad individual de cada uno de nosotros, en tanto que usuarios de los servicios y plataformas virtuales, de **proteger nuestra propia identidad y datos**, la primera barrera normativa y ejecutiva frente al uso no autorizado de nuestra identidad digital la constituye el marco jurídico de **protección de datos de carácter personal**, con el Reglamento europeo a la cabeza. La efectividad de dicho régimen descansa en el marco de gobernanza nacional (AEPD, en caso español) y europeo (Comisión y Comité Europeo de Protección de Datos, CEPD) encargado de garantizar el cumplimiento de la norma, mediante un denso y riguroso **sistema sancionador**. Llegado el caso la sanción puede incluso tener **dimensión penal**, cuando las conductas son tipificables penalmente (como pudieran ser: descubrimiento o revelación de secretos; amenazas, coacciones o acoso; calumnias o injurias; delitos de odio; estafas y/o delitos contra la libertad sexual o de violencia de género).

No obstante **en el caso de la IA**, aparecen nuevas perspectivas en relación con los datos que se utilizan en el **entrenamiento de estas tecnologías** y del uso de los datos que las plataformas generalistas poner en manos de los proveedores de este tipo de herramientas. En la mayoría de los casos, los datos personales cedidos de forma no consciente por parte de los usuarios (que van desde imágenes que compartimos hasta sencillamente las preferencias que mostramos con cada *click* que hacemos) son utilizados por estas plataformas, que los acumulan y usan o comercializan con terceros.

Uno de los aspectos novedosos que aborda esta norma, en 2024, es precisamente el que afecta a la regulación de los sistemas de IA que gestionan **sistemas de identificación biométrica**. En este contexto, el Reglamento IA contiene límites para el uso de tecnologías que permiten "identificación biométrica" (*unacceptable risk*). Este tipo de herramientas permiten el reconocimiento automatizado de características humanas de tipo físico, fisiológico o conductual, como la cara, el movimiento ocular, la forma del cuerpo, la voz, la entonación, el modo de andar, la postura, la frecuencia cardíaca, la presión arterial, el olor o las características de las pulsaciones de tecla, … características que permiten, en definitiva, determinar *la identidad* de una persona. Ello es posible comparando los datos biométricos de una persona con los datos biométricos de personas almacenados en una base de datos de referencia, independientemente de que la persona haya dado o no su consentimiento.

Este tipo de sistemas de IA, que permiten la "identificación biométrica remota y en 'tiempo real', en espacios de acceso público con fines de garantía del cumplimiento del derecho" son considerados por el Reglamento IA como **tecnologías prohibidas** (art. 5.1.h), aunque con algunas salvedades.

Sin embargo, sí están permitidos los sistemas de IA **destinados a la verificación biométrica, diseñados para permitir la autenticación**, y cuyo propósito es confirmar que una persona física concreta es la persona que dice ser, así como la identidad de una persona física con la finalidad exclusiva de que tenga acceso a un servicio, desbloquee un dispositivo o tenga acceso de seguridad a un local. Se trata en definitiva de encontrar el equilibrio entre seguridad y libertad, sin perder el tren de los inmensos avances que la nueva tecnología IA ofrece.

La garantía de acceso a los servicios públicos y privados automatizados y la necesaria protección de la identidad digital

En nuestro mundo actual, la identificación y la autentificación de la identidad en el entorno digital cobran especial importancia cuando se trata del **acceso a los servicios públicos** que han sido objeto de procesos de automatización y digitalización. En particular es relevante en ámbitos como el sanitario, el de las prestaciones sociales y laborales, en caso de los servicios y acceso a los distintos niveles educativos y, en general, para cualquier tipo de servicios administrativos en cuya prestación y actuación medie un trámite electrónico con cualquier tipo de Administración, ya sea estatal, autonómica o local.

Este acceso a los servicios públicos automatizados suscita no pocas dificultades, entre las que destaca la **brecha digital** y la consecuente situación de **vulnerabilidad** que acarrea. Personas mayores o en situación de discapacidad encuentran en este punto un elemento de desigualdad que se hace necesario acometer desde las exigencias del Estado social y de Derecho. Igual diagnóstico se produce en el ámbito privado y en el acceso a servicios privados, de distinta naturaleza, cuya prestación y contratación cada vez más tiende a digitalizarse.

Este problema se acentúa en el caso del acceso a los **sistemas de pago** y en la necesaria garantía de dichos sistemas para evitar fraudes y estafas.

Aunque los problemas jurídicos que plantean el acceso a los servicios públicos

(garantizar el derecho al acceso a dichos servicios) y el del acceso a los servicios privados (garantizar la seguridad jurídica de las transacciones económicas y el respeto de los derechos de la persona en el entorno digital) son muy distintos, en ambos casos se hace imprescindible **garantizar la identidad** de quien quiere acceder a un servicio público, de quien contrata, de quien paga. Para ello no solo son necesarias las mejores tecnologías de autenticación, como las que se exigen normas como las relativas los servicios de pagos. Es necesario también que **desde los poderes públicos se garantice unos servicios mínimos** en el ámbito de la identidad digital.

Esta es una parte importante de las novedades introducidas por el Reglamento eIDAS II: en concreto, la regulación de una **cartera europea de identidad digital**. Este tipo de servicios, cuya prestación puede correr a cargo de operadores privados o de operadores públicos, deja de ser una actividad ajena al regulador, para ser considerado como una suerte de servicio de interés público

La perspectiva reguladora del Reglamento eIDAS II está condicionada por el origen de la propuesta que dio lugar a la norma: la comisaría de competencia y mercado interior. Así, la norma europea se centra en ordenar la libertad de prestación y acceso a este tipo de servicios y busca armonizar las normativas nacionales de los Estados miembros para conseguir un mercado único en este tipo de servicio. Se regula así la actividad de las empresas que prestan servicios de "**carteras de identidad digital**", de sistemas de identificación electrónica y de prestación servicios cualificados de confianza, exigiéndose para dichas empresas la obtención de certificaciones ad hoc así como su sometimiento a un marco regulador único en la Unión Europea.

Desde la perspectiva del usuario de los servicios digitales, la norma se propone garantizar el **acceso a un contenido mínimo y en unas determinadas condiciones**, condiciones que permitan una cierta interoperabilidad y reconocimiento a nivel europeo.

Igualmente, y desde el Derecho europeo, se prevé la creación de un **marco de gobernanza**, diseñado para proveer las herramientas adecuadas de control y supervisión de las actividades que están llamadas a prestar las empresas antes mencionadas (ver más adelante)

La garantía de una identidad digital segura: el surgimiento de nuevo servicio económico de interés general

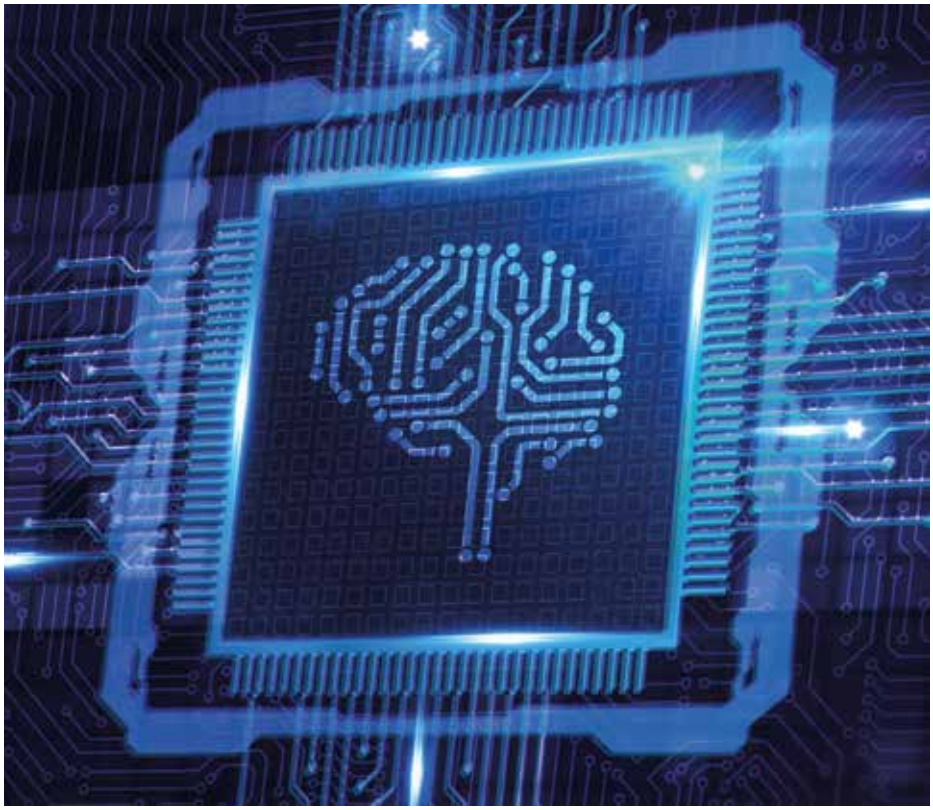
El Reglamento eIDAS II: ¿por qué un nuevo marco normativo europeo para proteger identidad digital?

La identidad digital y su protección segura funcionan como una infraestructura de la operativa digital y, por extensión, de la operativa de las aquellas herramientas de IA que implican la utilización de ciertos datos personales especialmente sensibles.

El Reglamento eIDAS II, de 2024, permite dar **un paso más en la protección la identidad** de quienes operan en el entorno virtual respecto del planteamiento -todavía algo primitivo- de su predecesor, el Reglamento eIDAS de 2014, en aquel momento centrado en garantizar el funcionamiento de los sistemas de verificación y confianza en las transacciones electrónicas, así como de la firma electrónica, como elementos relevantes de la identificación digital. El punto de vista de este marco normativo no es otro que el de la **seguridad jurídica**, su protección y garantía, todo ello al servicio de la actividad -en gran parte económica- que mueve el mundo digital.

Los **avances en materia de biometría** y la **aplicación en este ámbito de tecnología IA** entre otros, junto con el impacto del *blockchain*, han requerido la actualización y adaptación de este marco jurídico a dichas innovaciones tecnológicas.

Existe un amplio mercado de **servicios digitales que ofrecen sistemas para la verificación y autenticación** de todo tipo



de operaciones. La ausencia de un marco regulador común y las dificultades para el reconocimiento transfronterizo de estos servicios justifica algunas de las nuevas medidas acordadas.

La cartera europea de identidad digital (EU Digital Wallet): ¿un derecho?

Una "cartera europea de identidad digital" es un servicio ofrecido digitalmente que permite a los usuarios identificarse y autenticarse electrónicamente de forma transfronteriza, tanto en línea como fuera de línea, para acceder a una amplia gama de servicios públicos y privados (considerando 19 Reglamento eIDAS II). La posibilidad de que dicho servicio pueda ser **utilizado de forma transfronteriza** es esencial al concepto para el derecho europeo. Este sería el contenido básico de un EU Digital Wallet.

El art.5 bis del Reglamento eIDAS II formula la obligación de garantizar "el acceso transfronterizo seguro, de confianza y sin incidencias a servicios públicos y privados en la Unión", para todas las personas físicas y jurídicas, manteniendo estas el pleno control sobre sus datos a través de una cartera europea de identidad digital. Dicha obligación recae sobre los Estados miembros que **deberán proporcionar *al menos* una cartera europea de identidad digital**. Dicha obligación implica el surgimiento de un correlativo derecho de las personas físicas y jurídicas a disponer de dicha cartera. Será efectiva en un futuro próximo, de fecha inicialmente incierta, pues la entrada en vigor de la norma en este punto se hace depender de la aprobación, por parte de la Comisión, de los actos de ejecución previstos en el Reglamento eIDASII. Dichos actos ejecutivos, un total de cinco, fueron publicados en el DOUE el pasado 28 de noviembre de 2024. Ello implica, por tanto, que esta obligación/derecho no será efectiva hasta mediados de enero de 2027. Hasta esa fecha, no obstante, las prestaciones de servicios de carteras europeas de identidad digital que se ofrezcan en el mercado deberán acomodar su funcionamiento a lo dispuesto en los citados Reglamentos, cuyo contenido conjunto pretende –entre otras cosas- dotar de un marco común que garantice la operativa transfronteriza segura y confiable de dichos servicios y que favorezca el reconocimiento transfronterizo de la identidad digital en el mercado europeo. Se trata pues de generar un **ecosistema de identidad digital seguro e interoperable en toda la Unión**.

Alcance subjetivo de la obligación de ofrecer carteras europeas de identidad digitales

La obligación de proporcionar EUDI Wallet a personas físicas y jurídicas recae sobre los Estados miembros.

El art. 5.bis.2 del Reglamento EIDASII contempla distintas vías para poder cumplir con dicha obligación. Así, los Estados podrán proveer directamente dichas carteras, podrán hacer un mandato a un tercero (se entiende que vía contrato, aunque no se diga expresamente), o se podrá prestar la actividad de forma independiente (parece que a través del mercado), pero con *reconocimiento* por parte de ese Estado, es decir bajo el control jurídico-público de la actividad prestada por los operadores del mercado.

En este punto, la norma europea no prejuzga el régimen bajo el que cada Estado decida garantizar la prestación de este servicio de EUDI Wallet: servicio público o mercado. La atención de la norma se concentra en los elementos que determinan la interoperabilidad de las soluciones técnicas que se establecen. Para éstas se dice, por ejemplo, que el código fuente de los componentes de programas informáticos tendrán licencia de código abierto (art. 5.bis.3) o que las carteras permitirán al usuario operar de manera intuitiva, transparente y rastreable para un buen número de tareas (firmar electrónicamente, acceder a su registro de transacciones, generar pseudónimos, descartar sus propios datos, autenticarse para acceder a servicios públicos y privados, etc.).

Y todo ello se somete a regulación y control a través del "certificado de evaluación de conformidad" y de la supervisión de los *organismos de supervisión* creados al efecto. Se prevé por tanto un marco de gobernanza de nuevo cuño, fundamental para el cumplimiento de los requerimientos técnicos y de los controles de ciberseguridad.

Y aunque la obligatoriedad para los Estados miembros de garantizar el acceso a este EUDI Wallet no sea efectiva hasta 2027, ello no implica que los Estados miembros no puedan ya a adelantarse a prestar dichos servicios y, en todo caso, a ordenar la actividad de los operadores digitales que ofrezcan la prestación de dichos servicios.

Por su parte, y desde la Unión Europea y para alcanzar el objetivo propuesto de disponibilidad de un EUDI Wallet, la propia Comisión europea ha puesto ya en marcha un proyecto piloto con cuatro propuestas ofrecidas por sendas plataformas digitales.

Acance objetivo de obligación de ofrecer carteras europeas de identidad digitales

El art. 5.bis.1 del Reglamento EIDASII se refiere a "un acceso transfronterizo seguro, de confianza y sin incidencias a servicios públicos y privados en la Unión Europea, manteniendo al mismo tiempo el pleno control sobre sus datos". La herramienta para conseguir dicho acceso (objetivo principal) es la cartera europea de identidad digital (EUDI wallet), un servicio digital que permita la identificación digital de personas físicas y jurídicas. Es importante recordar cuál es el objetivo final

de servicio EUDI wallet. Solo desde esta perspectiva se entiende qué tipo de datos personales habrán de quedar incluidos en las mencionadas carteras.

La identificación de la persona física a través del sistema digital deberá garantizar la autentificación e impedir la suplantación de forma segura, cumpliendo la función que el mundo no virtual cumplen los sistemas tradicionales de identificación, como por ejemplo el DNI. La tecnología biométrica virtual permite hoy en día sistemas de autentificación muchísimo más robustos y seguros que los físicos (desde la identificación a través de la huella digital, el iris, el rostro o la voz). A este primer elemento habrán de vincularse otros datos certificables de la persona física, como son los títulos académicos reconocidos por el sistema educativo, o los datos identificativos de su inclusión en el sistema sanitario, o la posesión de acreditaciones como el carné de conducir o un certificado de penales o el uso de firmas electrónicas. Es decir, datos personales que en muchos casos son requeridos para acceder a servicios públicos o privados y que la norma califica como "declaraciones electrónicas de atributos".

Más allá de las obligaciones que recaen en los Estados miembros, lo que realmente caracteriza el servicio que se pretende que aporte el EUDI wallet es su uso y reconocimiento transfronterizo. Esta razón ha llevado a que también la propia UE haya tomado sus propias iniciativas. Así, la Comisión Europea ha puesto en marcha 4 Proyecto pilotos que quiere aplicar la billetera EUDI a 6 casos de uso.

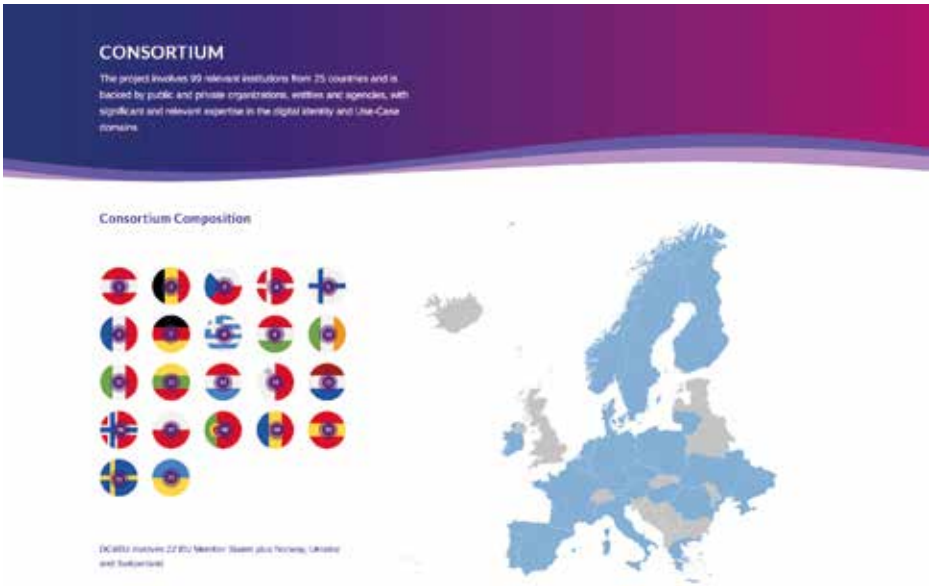
Entre estos proyectos, está, por ejemplo, el proyecto POTENCIAL, coordinado por Alemania y Francia con la participación de 17 Estados miembros y Ucrania. En él participan más de 50 administraciones públicas y más de 80 entidades privadas. El proyecto aplicará la billetera EUDI a 6 casos de uso

- Acceso a servicios gubernamentales
- Apertura de una cuenta bancaria
- Registro de una tarjeta SIM
- Licencia de conducir móvil
- firmas electrónicas
- Recetas electrónicas

Entre los 4 proyectos aprobados, está también el DC4EU (*Digital Credentials for Europe Consortium*), proyecto coordinado por España y en el que están involucrados 23 países miembros y, junto con Ucrania, 35 administraciones públicas y 49 entidades privadas

Este proyecto probará el uso de la billetera EUDI en el sector educativo y en el de la seguridad social. De este modo, el proyecto piloto se alineará con el Pase Europeo de Seguridad Social y el Modelo Europeo de Aprendizaje. Utilizará la Infraestructura Europea de Servicios Blockchain (EBSI) en el contexto de la billetera EUDI.

Titulares del derecho a recibir el servicio transfronterizo de carteras europeas de identidad digitales
La regulación de la billetera EUDI, en la medida en que busca proveer,



principalmente a las **personas físicas**, de un servicio para la identificación digital con un contenido mínimo e impone a los Estados Miembros un plazo para hacer real este servicio, está formulando de *facto* lo que en otros ámbitos sectoriales se califica como servicio esencial o servicio de interés económico general (SIEG). Como es bien sabido, la calificación de un servicio o actividad como esencial o de interés general no determina su publificación y es compatible con la prestación de dichas actividades o servicios a través del mercado.

La identificación es, sin duda, una función pública inicialmente destinada a las personas físicas, a los ciudadanos en general. En el Reglamento EIDAS II esta tarea se regula sin distinción de nacionalidad, referida a las personas físicas, sin más calificativos. Se entiende que se trata de un paso previo a cualquier otro. La provisión de servicios públicos –sanitarios, educativos, discapacidad o de atención social, o simplemente de accesos a trámites administrativos por vía electrónica- quizá sí plantea distinciones en función de otros elementos personales: nacionalidad, sexo, edad… No ocurre así sin embargo respecto de la identidad digital, referida solo a la condición de "persona física".

En un apartado distinto hay que mencionar a **la persona jurídica**, para la que también el Reglamento EIDASII reconoce el derecho a disponer de EUDI Wallets, aunque obviamente de contenido distinto, propio de su condición. El art. 5 bis del citado Reglamento EIDASII habla de "garantizar que todas **las personas físicas y jurídicas** dispongan de un acceso transfronterizo seguro, de confianza a servicios públicos y privados (…) manteniendo al mismo tiempo el propio control de sus datos…".

Obviamente el modo de llevar a cabo la identificación de las personas físicas es completamente distinto al de las personas jurídicas. La biometría no cabe aquí. Por otra parte, la propia regulación de las vías para acceder a la personalidad jurídica serán distintas en cada uno de los países miembros, al igual que lo serán los propios tipos de personas jurídicas. Aunque hasta la fecha la atención se haya centrado en los problemas que puede plantear esta herramienta en el caso de las personas físicas, en particular, cuando se trata de evitar la suplantación y los riesgos que de ello se pueden derivar, en el caso de las personas jurídicas la EUDI Wallet y su carácter transfronterizo se convierte en un instrumento muy potente para al servicio de las libertades económicas, de prestación y establecimiento en el contexto de la UE.

Es, por tanto, un tipo de servicio para el que razonablemente hay y habrá mercado, pues en particular las empresas encontrarán incentivos en disponer de instrumentos que les permitan gestionar y garantizar no solo su identificación también ciertos atributos (autorizaciones, licencias, certificados de estar al corriente de ciertas obligaciones, etc…) con carácter transfronterizo.

El éxito de este tipo de herramientas digitales, capaz de mantener el equilibrio entre autenticación y protección de los datos, eficacia y ciberseguridad, a la vista del impacto que han tenido otras regulaciones europeas como es el caso del régimen de la protección de datos de carácter personal pudiera llevar a pensar, quizá, en un nuevo "efecto Bruselas".

El nuevo marco de gobernanza de la ID

Estructura del nuevo marco de gobernanza del Reglamento EIDAS II

Entre las novedades que se han incorporado recientemente por el Reglamento EIDASII está la previsión de una nueva estructura institucional, de carácter mixto, público-privado, para ordenar, supervisar y coordinar la actividad de los operadores que prestan servicios de carteras europeas de identidad digital, de sistemas de identificación electrónica y de servicios de confianza cualificados.

Más allá de las **funciones estrictamente reguladoras** que, con carácter general, han quedado residenciadas en las propias instituciones europeas (Parlamento, Consejo y Comisión, en particular en esta última, mediante la aprobación de números actos de ejecución), el Reglamento EIDASII ha dispuesto la siguiente arquitectura de gobernanza:

- Los **organismos de evaluación de la conformidad**, de ámbito nacional
- Los **organismos de supervisión**, también de ámbito nacional
- El **Grupo de coordinación sobre la Identidad Digital europea**, de ámbito europeo, creado para facilitar la cooperación transfronteriza y que estará formado por los Estados miembros y la Comisión. Carece de funciones ejecutivas, aunque sin duda está llamado a tener relevancia en este ámbito de gran complejidad técnica.

Con esta nueva estructura de supervisión, el Derecho europeo trae al ámbito de los servicios digitales de identificación elementos propios de modelos regulatorios extensamente desarrollados en otros sectores económicos. Sin llegar a establecer la obligatoriedad de un sistema autorizaciones administrativas previas, sí prevé mecanismos de control a priori de la actividad, como es la **exigencia de certificaciones**, otorgadas por organismos de evaluación de la conformidad.

El cumplimiento y mantenimiento de los requisitos necesarios para obtención de las mencionadas certificaciones, se completa con un régimen de control y supervisión ex post, cuya responsabilidad recae, ahora sí, en los organismos de supervisión que los Estados miembros deberán disponer.

De este modo, el regulador europeo consolida su modelo regulador y garantiza el cumplimiento del marco normativo relativo a la identidad digital, con la

previsión de un régimen de control y sanción reforzado. En concreto, y en relación con esto último, para que su eficacia sea completa, será necesario que el legislador español ajuste las potestades que esta materia ha atribuido a los *organismos de supervisión*.

Los organismos de evaluación de la conformidad: sujetos y funciones

El acceso a la prestación de servicios de identificación digital en sus diversas modalidades no estás sometido a día de hoy a un proceso de obtención de autorización administrativa previa. Sin embargo, la necesidad de garantizar ciertas características técnicas de este tipo de operadores sí requiere de ciertos controles de evaluación o conformidad.

Este papel queda encomendado a los llamados *organismos de evaluación de la conformidad* a los que el Reglamento EIDASII hace referencia en varios momentos (p. ej. art. 5 quarter).

Este tipo de organismos, a su vez, están sometidos a un control por parte de las Autoridades Nacionales así como por la Comisión europea, que debe ser convenientemente informada por los Estados miembros de los organismos de este tipo designados por cada Estado (art. 5 quarter.7 EIDASII). Serán estos quienes podrán emitir el correspondiente informe de evaluación de la conformidad en este sector.

Este modelo de regulación, también llamado de autorregulación regulada, en definitiva, traslada a la propia industria la primera barrera de control, correspondiéndole evaluar la conformidad de los operadores con las normas técnicas y actos de ejecución que la Unión Europea debe dictar en este ámbito por delegación del propio Reglamento.

Los organismos de supervisión: sujetos y funciones

El modelo de *gobernanza* ya estaba incluido en 2014, con EIDASI. En 2024, EIDAS II ha reforzado los instrumentos institucionales responsables de garantizar el buen funcionamiento del marco normativo con la previsión de los organismos de supervisión.

El art. 46 bis. 1. del Reglamento EIDASII dispone que los nuevos *organismos de supervisión* "disfrutaran de la competencias necesarias y los recursos adecuados para el ejercicio de sus funciones de forma eficaz, eficiente e independiente".

Estos organismos de supervisión, que podrán ser uno o más, desempeñarán sus funciones en relación con los proveedores de carteras europea de identidad digital (art. 46 bis) así como con los prestadores de servicios de confianza (art. 46 ter).

A pesar de que el Reglamento no lo dice expresamente, al diseñar estos *organismos de supervisión*, parece referirse a un regulador-supervisor de tipo

independiente, al modo que existe en otros sectores -protección de datos, por ejemplo-. Aunque todavía en España no se ha desarrollado, no parece descabellado pronosticar que esta previsión termine dando lugar a la creación de una entidad del tipo *administración independiente* o, bien, a ensanchar las funciones de alguna de las existentes para encomendarle las tareas que para este tipo de *organismo de supervisión* prevé el art. 46 bis.3.

En todo caso y teniendo en cuenta algunas de las tareas –potestades públicas- que según el Reglamento EIDAS II habrán de desempeñar este tipo de organismos y a la vista de las exigencias que nuestro derecho constitucional prevé, el ejercicio de dichas potestades deberá ser atribuido a funcionarios públicos incardinados en una Administración pública, quienes adoptará sus decisiones sometidas a derecho administrativo. Y serán decisiones susceptibles, en definitiva, de ser revisadas por los tribunales, en este caso, contencioso-administrativos.

Entre otras tareas mencionadas, y significativamente, se menciona la potestad prevista en el art. 46 bis.4.f) referida a la posibilidad, en caso de "uso ilegal o fraudulento", de "**suspender o cancelar** el registro y la inclusión de las parte usuarias" en el mecanismo común que los Estados miembros facilitarán para permitir la identificación y autenticación, así como la de "**suspender o cesar** la provisión de la cartera europea de identidad digital" (art. 46 bis.5). Se trata de medidas –decisiones- de claros efectos desfavorables, materialmente sancionadoras y con un eventual impacto negativo en términos económicos.

Si a estas tareas sumamos las de inspección y supervisión, incluidas las inspecciones *in situ* (art. 46.bis.4.d)) o las de requerir la adopción de medidas de corrección art. 46.bis.4.e)), que se mencionan en el Reglamento EIDASII, lo razonable, como se ha dicho será la creación una entidad independiente reguladora para desempeño de esta nueva tarea y la garantía de cumplimiento de este marco normativo ahora reformado.

En definitiva, con este *marco de gobernanza* se implanta un modelo institucional y de regulación, de carácter jurídico-público, a una actividad prestada hasta la fecha principalmente por operadores privados que, debido a su novedad, carecía de respuestas ad hoc. El interés general llamado a ser protegido es múltiple y va desde la protección del derecho fundamental a la intimidad por el uso de los datos personales, a la ciberseguridad, pasando por la protección de las libertadas de económicas –de prestación de servicios y de circulación de los usuarios- en Europa.

IA en todas partes: Magia, pero son algoritmos[1]

Dos años después de que la Inteligencia Artificial generativa se convirtiera en las palabras más de moda de todo el mundo, parece que lo que nos depara el futuro de la tecnología es simplemente... más IA.

Sin embargo, eso es solo una parte de la historia. En el futuro, esperamos que la IA se entreteja de manera tan fundamental en nuestras vidas, que acabe estando todas partes, y sea tan fundamental que como ha pasado con muchas de otras tecnologías dejemos de notarla.

Tomemos la electricidad, por ejemplo. ¿Cuándo fue la última vez que pensaste en los electrones? Ya no nos maravillamos de que las luces se enciendan, simplemente esperamos que funcionen. Lo mismo ocurre con Internet y los protocolos que la sostienen, son como hilos invisibles que mantienen unido a Internet. Los usamos todos los días, pero apuesto a que la mayoría de nosotros no hemos pensado que detrás de estas comodidades, hay tecnologías que, como la IA, fueron en su día innovadoras.

Con el tiempo, la IA seguirá un camino similar, volviéndose tan omnipresente que formará parte de la subestructura invisible de todo lo que hacemos, y finalmente ni siquiera sabremos que está ahí. Zumbará silenciosamente en el fondo, optimizando el tráfico en nuestras ciudades, personalizando nuestra atención médica y creando rutas de aprendizaje adaptativas y accesibles en la educación. No vamos a "usar" la IA. Simplemente experimentaremos un mundo en el que las cosas funcionan de manera más inteligente, rápida e intuitiva, como por arte de magia, pero basadas en algoritmos. Esperamos que proporcione una base para el crecimiento empresarial y personal, al mismo tiempo que se adapte y se mantenga a lo largo del tiempo.

En ningún lugar es más evidente este futuro infundido por la IA que en todos los informes de tendencias tecnológicas de este año, siendo casi todas ellas elevadas por la aparición de este fenómeno. No es la "única tendencia" ni "todas las tendencias", la IA es el andamiaje y el hilo conductor que apuntala casi todas las tendencias.

A medida que la IA evoluciona, el enfoque empresarial en grandes modelos de lenguaje está dando paso a modelos de lenguaje pequeños, modelos multimodales, simulaciones basadas en IA y agentes que pueden ejecutar tareas discretas, así como después de años de dominio del software, el hardware está recuperando el centro de atención, en gran parte debido al impacto de la IA en los chips informáticos, sus necesidades abrumadoras de computación y su integración en los dispositivos de los usuarios finales, el Internet de las cosas y la robótica.

La aplicabilidad de la IA a la escritura de código, las pruebas de software y el aumento del talento tecnológico está transformando la TI y provocando un alejamiento de la virtualización y los presupuestos austeros.

Debido a que esperamos que la IA se convierta en parte del núcleo fundacional del mañana, como la electricidad, Internet y tantas otras tecnologías, es emocionante pensar en cómo podría evolucionar la IA en los próximos años a medida que avanza hacia la ubicuidad, y cómo nosotros, como humanos, podemos beneficiarnos.

1 El texto de este capítulo se ha redactado con la ayuda de sistemas de Inteligencia Artificial que han generado una parte relevante del texto y de su contenido. La revisión final ha sido realizada por porfesionales en este ámbito.

¿Qué sigue para la IA?

A medida que los grandes modelos de lenguaje continúan avanzando, los nuevos enfoques basados en agentes están demostrando ser más efectivos en tareas discretas, demostrando que la IA necesita diferentes caballos para diferentes carreras y abriendo mucho campo a múltiples estrategias.

No es un secreto que la velocidad del avance de la Inteligencia Artificial está superando las expectativas. El año pasado, las organizaciones se esforzaban por comprender cómo adoptar la IA generativa, intentaban diferenciarse de los competidores y adoptaban un enfoque estratégico para escalar el uso de grandes modelos de lenguaje (LLM). Hoy en día, los LLM se han arraigado, y hasta el 70% de las organizaciones, según algunas estimaciones, exploran o implementan activamente casos de uso sobre estos LLM.

Pero las organizaciones líderes ya están considerando el próximo capítulo de la IA. En lugar de confiar en modelos básicos construidos por grandes actores de la IA, que pueden ser más potentes y basarse en más datos de los necesarios, las empresas ahora están pensando en implementar modelos múltiples y más pequeños que puedan ser más eficientes para los requisitos comerciales. Los LLM continuarán avanzando y serán la mejor opción para ciertos casos de uso, activando sobre ellos aplicaciones como asistentes de propósito general o simulaciones para investigación científica. Pero el asistente que examina sus datos financieros para pensar en las oportunidades de ingresos perdidas no tiene por qué ser el mismo modelo que responde a las consultas de los clientes. En pocas palabras, vamos hacia modelos más personalizados, más pequeños y de uso más focalizado en ámbitos o incluso tareas concretas.

Una serie de modelos más pequeños que trabajan en conjunto pueden terminar sirviendo para casos de uso diferentes a los enfoques actuales de LLM. Las nuevas opciones de código abierto y las entradas/salidas multimodales están permitiendo a las organizaciones desbloquear enfoques completamente nuevos.

Las organizaciones están embarcadas en un cambio fundamental en la IA, que pasa de aumentar el conocimiento a aumentar la producción. Las inversiones que se están realizando hoy en día en Agentic AI, como se denomina a este próximo enfoque más transaccional, cambiará la forma en que trabajamos y vivimos, al permitir a los consumidores y las empresas contar con “ejércitos” de asistentes que no sólo puedan llevar a cabo tareas discretas, más de consulta o generación pura de información, si no a realizar tareas complejas con transaccionalidad, como entregar un informe financiero en una reunión de la junta directiva o solicitar una subvención. Podemos pasar de "Hay una aplicación para eso" a "Hay un agente para eso".

Los datos y la IA, una relación intrínseca

Según el informe State of Generative AI in the Enterprise Q3 2024 de Deloitte, el 75% de las organizaciones encuestadas han aumentado sus inversiones en la gestión del ciclo de vida de los datos gracias a la IA generativa. Los datos son fundamentales para los LLM, porque las malas entradas conducen a peores salidas (en otras palabras, basura que entra, basura al cuadrado).

Mientras que algunas empresas de IA usan extensivamente lo que está disponible en Internet para construir los modelos más grandes posibles, otras apuestan por una "educación" específica del dominio para sus LLM cuidando muy bien la fuente de datos usada en su entramiento.

Las organizaciones encuestadas por Deloitte han expresado que los nuevos problemas podrían quedar expuestos por la ampliación de los pilotos de IA, las regulaciones poco claras en torno a los datos confidenciales y las preguntas sobre el uso de datos externos (por ejemplo, datos de terceros con licencia). Es por eso que el 55% de las organizaciones encuestadas evitaron ciertos casos de uso de IA debido a problemas relacionados con los datos, y una proporción igual está trabajando para mejorar la seguridad de sus datos. Las organizaciones podrían solucionar estos problemas mediante el uso de modelos listos para usar ofrecidos por los proveedores, pero el impacto diferenciado de la IA probablemente requerirá datos empresariales diferenciados.

Afortunadamente, una vez que se sientan las bases, los beneficios son claros: dos tercios de las organizaciones encuestadas dicen que están aumentando las inversiones en IA generativa porque han visto un gran valor hasta la fecha. También están apareciendo ejemplos iniciales de valor en el mundo real en todas las industrias, desde la revisión de reclamaciones de seguros hasta la resolución de problemas de telecomunicaciones y herramientas de segmentación de consumidores. La IA generativa y los LLM también están causando sensación en casos de uso más especializados, como las reparaciones en el espacio, el modelado nuclear y el diseño de materiales, por no hablar de la biomedicina o la genética.

A medida que las entradas de datos subyacentes mejoran y se vuelven más sostenibles, los LLM y otros modelos avanzados (como las simulaciones) pueden ser más fáciles de implementar y escalar. Pero el tamaño no lo es todo. Con el tiempo, a medida que proliferan los métodos para el entrenamiento y la implementación de la IA, es probable que las organizaciones pongan a prueba modelos más pequeños. Muchos pueden tener datos que pueden ser más valiosos de lo que se imaginaba anteriormente, y ponerlos en acción a través de modelos más pequeños y orientados a tareas puede reducir el tiempo, el esfuerzo y las molestias. Estamos preparados para pasar de los proyectos de IA a gran escala a la IA en todas partes, como ya hemos mencionado en la introducción.

Diferentes monturas para diferentes recorridos

Si bien los LLM tienen una amplia gama de casos de uso, la biblioteca no es infinita (todavía). Los LLM requieren recursos masivos, se ocupan principalmente de texto y están destinados a aumentar la inteligencia humana en lugar de asumir y ejecutar tareas discretas. Como resultado, dice Vivek Mohindra, vicepresidente senior de estrategia corporativa de Dell Technologies, "no existe un enfoque único para la IA. Va a haber modelos de todos los tamaños y opciones especialmente diseñadas, esa es una de nuestras creencias clave en la estrategia de IA".

En los próximos 18 a 24 meses, es probable que los principales proveedores de IA y usuarios empresariales tengan un conjunto de herramientas de modelos que comprenden LLM cada vez más sofisticados y robustos, junto con otros modelos más aplicables a los casos de uso cotidianos. De hecho, cuando los LLM no son la opción óptima, tres pilares de la IA están abriendo nuevas vías de valor: los

modelos de lenguaje pequeño, los modelos multimodales y la IA agentica (figura 1).

Modelos lingüísticos pequeños
Los proveedores de LLM están compitiendo para hacer que los modelos de IA sean lo más eficientes posible. En lugar de permitir nuevos casos de uso, estos esfuerzos tienen como objetivo ajustar u optimizar los modelos para los casos de uso existentes. Por ejemplo, los modelos masivos no son necesarios para tareas mundanas como resumir un informe de inspección: un modelo más pequeño entrenado con documentos similares sería suficiente y más rentable.

Las empresas pueden entrenar modelos de lenguaje pequeños (SLM) con conjuntos de datos más pequeños y altamente seleccionados para resolver problemas más específicos, en lugar de consultas generales. Por ejemplo, una empresa podría entrenar a un SLM con su información de inventario, lo que permitiría a los empleados recuperar rápidamente información en lugar de analizar manualmente grandes conjuntos de datos, un proceso que a veces puede llevar semanas.

Naveen Rao, vicepresidente de IA de Databricks, cree que más organizaciones adoptarán este enfoque de sistemas con IA: "Una computadora mágica que entiende todo es una fantasía de ciencia ficción. Más bien, de la misma manera que organizamos a los seres humanos en el lugar de trabajo, debemos separar nuestros problemas. Los modelos personalizados y específicos del dominio pueden abordar tareas específicas, las herramientas pueden ejecutar cálculos deterministas y las bases de datos pueden extraer datos relevantes. Estos sistemas de IA ofrecen la solución mejor de lo que cualquier componente podría hacerlo por sí solo".

Un beneficio adicional de los modelos más pequeños es que las empresas pueden ejecutarse en el dispositivo y entrenarlos en conjuntos de datos más pequeños y altamente seleccionados para resolver problemas más específicos, en lugar de consultas generales

Empresas como Microsoft y Mistral están trabajando actualmente para separar estos modelos SLM, basados en menos parámetros y menos costosos de computar, de sus ofertas de IA más grandes. Un claro ejemplo es Meta que ofrece múltiples opciones en modelos más pequeños y ajustados a cada necesidad.

Por último, gran parte del progreso que se está produciendo en los SLM se produce a través de modelos de código abierto ofrecidos por empresas como Hugging Face. Estos modelos están listos para el uso empresarial, ya que pueden personalizarse para cualquier número de necesidades, siempre que los equipos de TI tengan el talento interno de IA para ajustarlos. De hecho, un informe reciente de Databricks indica que más del 75% de las organizaciones están eligiendo modelos de código abierto más pequeños y personalizándolos para casos de uso específicos. Dado que los modelos de código abierto mejoran constantemente gracias a las contribuciones de una comunidad de programación diversa, es probable que el tamaño y la eficiencia de estos modelos mejoren a un ritmo rápido.

Modelos multimodales
Los seres humanos interactúan a través de una variedad de medios: texto, lenguaje corporal, voz, videos, entre otros. La IA ahora espera ponerse al día y equilibrar la balanza. Dado que las necesidades comerciales no se limitan al texto, no es de extrañar que las empresas estén esperando una IA que pueda absorber y producir múltiples medios. De alguna manera, ya estamos acostumbrados a la IA multimodal, como cuando hablamos con asistentes digitales y recibimos texto o imágenes a cambio, o cuando viajamos en automóviles que usan una combinación de visión por computadora y señales de audio para brindar asistencia al conductor.

La IA generativa multimodal, por otro lado, se encuentra en sus primeras etapas. Los primeros modelos importantes como GPT-4 Omni de OpenAI, se exhibieron en mayo de 2024. El progreso en la IA generativa multimodal puede ser lento porque requiere cantidades significativamente mayores de datos, recursos y hardware. Además, los problemas existentes de alucinación y sesgo que afectan a los modelos basados en texto pueden verse exacerbados por la generación multimodal.

Aun así, los casos de uso empresarial son prometedores. La noción de "entrenar una vez, correr en cualquier lugar (o de cualquier manera)" promete un modelo que podría entrenarse con texto, pero que podría ofrecer respuestas en imágenes, video o sonido, según el caso de uso y la preferencia del usuario, lo que mejora la inclusión digital. Em presas como AMD tienen como objetivo utilizar la incipiente tecnología para traducir rápidamente materiales de marketing del inglés a otros idiomas o para generar contenido. Para la optimización de la cadena de suministro, la IA generativa multimodal se puede entrenar con datos de sensores, registros de mantenimiento e imágenes de almacén para recomendar cantidades ideales de existencias. A medida que la tecnología avanza y la arquitectura del modelo se vuelve más eficiente, podemos esperar ver aún más casos de uso en los próximos 18 a 24 meses.

AGENTIC AI
Este nuevo pilar de la IA, del que ya hemos comentado algo anteriormente, puede allanar el camino para cambios en nuestra forma de trabajar durante la próxima década. Los modelos de acción grandes (o pequeños) van más allá de las capacidades de preguntas y respuestas de los LLM y completan tareas discretas en el mundo real.

Los ejemplos van desde reservar un vuelo en función de sus preferencias de viaje hasta proporcionar un servicio de atención al cliente automatizado que pueda acceder a las bases de datos y ejecutar las tareas necesarias, probablemente sin la necesidad de indicaciones altamente especializadas.

La proliferación de estos modelos de acción, que funcionan como agentes digitales autónomos, anuncia los inicios de los agentes de IA, y proveedores de software empresarial como Salesforce y ServiceNow ya están promocionando estas posibilidades e incorporándolas en sus herramientas.

Chris Bedi, director de atención al cliente de ServiceNow, dice que "la IA agentica no puede reemplazar completamente a un humano, pero lo que puede hacer es trabajar junto a sus equipos, manejando tareas repetitivas, buscando información y recursos, haciendo trabajo en segundo plano las 24 horas del día, los 7 días de la semana, los 365 días del año".

Aparte de las diferentes categorías de modelos de IA mencionadas anteriormente, los avances en el diseño y la ejecución de la IA también pueden afectar a la adopción empresarial, es decir, a la llegada de las redes neuronales líquidas. "Líquido" se refiere a la flexibilidad de esta nueva forma de entrenar a la IA a través de una red neuronal, un algoritmo de aprendizaje automático que imita la estructura del cerebro humano. De manera similar a cómo las computadoras cuánticas se liberan de la naturaleza binaria de la computación clásica, las redes neuronales líquidas pueden hacer más con menos: un par de docenas de nodos en la red podrían ser suficientes, frente a 100,000 nodos en una red más tradicional.

Figura 1
Different AI for different needs

	Small language models	Multimodal	Agentic
Focus	**Text, customizable, applied to different use cases (trainable)**	Can't train on smaller data sets; needs greather input and has widervariety of output	Can take concrete actions
Input	Text	More than text	Text
Output	Some	More	Most
Data	Less	Significant	To be determined
Customization	Need to be customized and trained on data they would work with	Less cutomization possible due to the volume of data required	Vendor provide out-of-the-box capabilities, but works best when tailored

Fuente: Deloitte research

La tecnología de vanguardia tiene como objetivo funcionar con menos potencia de cómputo, con más transparencia, lo que abre posibilidades para incorporar la IA en dispositivos periféricos, robótica y sistemas críticos para la seguridad.

En otras palabras, no son solo las aplicaciones de la IA, sino también sus mecanismos subyacentes los que están listos para mejorar y ser disruptivos en los próximos años.

Figura 2
Compound AI journey

Fuente: Deloitte research

Conclusión

La IA tiene un potencial transformador, como todo el mundo ha escuchado durante el último año, pero solo en la medida en que el liderazgo lo permita.

La aplicación de la IA como una forma más rápida de hacer las cosas de la forma en que siempre se han hecho resultará, en el mejor de los casos, en un potencial perdido y, en el peor, en sesgos amplificados.

El liderazgo imaginativo y valiente llevará esta tecnología a otro nivel estableciendo las mejores prácticas y creando las "próximas prácticas", en las que encontremos nuevas formas de organizarnos a nosotros mismos y a nuestros datos hacia un mundo habilitado por la IA.

Cuando se trata de IA, es probable que las empresas tengan en el futuro las mismas consideraciones que hoy: datos, datos y datos. Hasta que los sistemas de IA puedan alcanzar la Inteligencia Artificial general o aprender tan eficientemente como el cerebro humano, estarán hambrientos de más datos que los ayuden a ser más poderosos y precisos. Las medidas adoptadas hoy para organizar, racionalizar y proteger los datos empresariales podrían dar sus frutos en los próximos años, ya que la deuda de datos podría convertirse algún día en la mayor parte de la deuda técnica.

Este trabajo preliminar y un correcto liderazgo también deberían ayudar a las empresas a prepararse para la letanía de desafíos regulatorios e incertidumbres éticas (como las limitaciones de recopilación y uso de datos, preocupaciones sobre la equidad, la falta de transparencia) que conlleva guiar esta nueva y poderosa tecnología hacia el futuro.

Lo que está en juego no es baladí y marcará la diferencia a las futuras generaciones.

Deloitte.