

CLAVES PRÁCTICAS

FRANCIS LEFEBVRE

**Reglamento UE de
inteligencia artificial**
(Incluye los actos de desarrollo y
ejecución de la AI Act)

Fecha de edición: 22 de abril de 2025

Esta monografía de la Colección
CLAVES PRÁCTICAS
es una obra editada por iniciativa y bajo
la coordinación de
Francis Lefebvre

MOISÉS BARRIO ANDRÉS

Letrado del Consejo de Estado

Doctor en Derecho

*Director del Postgrado en Legal Tech y transformación digital de la Escuela de
Práctica Jurídica de la Universidad Complutense de Madrid*

Profesor de Derecho digital

Abogado y consultor

NOTA. – Esta obra es fruto de las reflexiones personales del autor sobre la normativa analizada. Los comentarios y conclusiones que se incluyen no suponen en ningún caso un asesoramiento jurídico directo. En consecuencia, ni la editorial ni el autor aceptarán responsabilidades por las consecuencias ocasionadas a las personas o entidades que actúen o dejen de actuar como consecuencia de las opiniones, interpretaciones e informaciones contenidas en esta obra.

© Francis Lefebvre
Lefebvre-El Derecho, S. A.
Monasterios de Suso y Yuso, 34. 28049 Madrid. Teléfono: 91 210 80 00.
clientes@lefebvre.es
www.efl.es
Precio: 36,40 € (IVA incluido)
ISBN: 978-84-10431-75-1
Depósito legal: M-10631-2025
Impreso en España por Printing'94
Carretera de Canillas, 138 – 28043 Madrid

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Dirijase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra.

*A todas las personas que me han llevado hasta aquí,
personal y profesionalmente*

Plan general

	nº marginal
Capítulo 1. Disposiciones generales	10
Capítulo 2. Sistemas de IA prohibidos	200
Capítulo 3. Sistemas de IA de alto riesgo	500
Capítulo 4. Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA	1200
Capítulo 5. Modelos de uso general.....	1400
Capítulo 6. Sistemas de bajo o nulo riesgo.....	1500
Capítulo 7. Responsabilidades a lo largo de la cadena de valor de la IA	1600
Capítulo 8. Normas armonizadas y especificaciones comunes	1700
Capítulo 9. Espacios controlados de pruebas.....	1800
Capítulo 10. Gobernanza	1900
Capítulo 11. Sanciones y otras medidas de cumplimiento.....	2200
Capítulo 12. Derecho español.....	2300
	Página
Tabla Alfabética	243

Abreviaturas

AEPD	Agencia Española de Protección de Datos
AESIA	Agencia Española de Supervisión de la Inteligencia Artificial
art.	artículo/s
CCom	Código de Comercio (RD 22-8-1885)
CDFUE	Carta de Derechos Fundamentales de la Unión Europea
cdo.	considerando
CGPJ	Consejo General del Poder Judicial
CNMV	Comisión Nacional del Mercado de Valores
Const	Constitución Española
Dir	Directiva
Directiva NIS2	Directiva (UE) 2022/2555, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la UE
DMA	Reglamento (UE) 2022/1925, sobre mercados disputables y equitativos en el sector digital
DOUE	Diario Oficial de la Unión Europea
DSA	Reglamento (UE) 2022/2065, relativo a un mercado único de servicios digitales
EIDF	Evaluación del impacto sobre los derechos fundamentales
EIPD	Evaluación del impacto sobre la protección de datos
ENISA	Agencia de la Unión Europea para la Ciberseguridad
FFCCSE	Fuerzas y Cuerpos de Seguridad del Estado
GPAI	Modelo de IA de uso general.
IA	Inteligencia Artificial
L	Ley
LECr	Ley de Enjuiciamiento Criminal
LO	Ley Orgánica
LOPDGDD	Ley Orgánica de Protección de Derechos Personales y garantía de los derechos digitales (LO 3/2018)
LOREG	Ley Orgánica del Régimen Electoral General (LO 5/1985)
LCSP	Ley de Contratos del Sector Público (L 9/2017)
LSC	Texto Refundido de la Ley de Sociedades de Capital (RDLeg 1/2010)
NML	Nuevo Marco Legislativo para la seguridad y la conformidad de los productos industriales
OCDE	Organización para la Cooperación y el Desarrollo Económicos
RBI	identificación biométrica remota
RD	Real Decreto
RDL	Real Decreto-Ley
RGPD	Reglamento General de Protección de Datos (Rgto (UE) 2016/679)
Rgto	Reglamento
RIA	Reglamento Europeo de Inteligencia Artificial (Rgto (UE) 2024/1689)
RRM	Reglamento del Registro Mercantil (RD 1784/1996)
RVM	Reglamento (UE) 2019/1020, relativo a la vigilancia del mercado y la conformidad de los productos

s.	siguientes
TCo	Tribunal Constitucional
TJUE	Tribunal de Justicia de la Unión Europea
TS	Tribunal Supremo
Tratado FUE	Tratado de funcionamiento de la Unión Europea
Tratado UE	Tratado de la Unión Europea
UE	Unión Europea

Prólogo

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 y por el que se modifican los Reglamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial), que en lo sucesivo denominaremos de forma abreviada como RIA, es la **primera regulación jurídica general** de la inteligencia artificial (IA) en la UE.

El RIA está construido sobre la base de dos paradigmas regulatorios. En primer lugar, y de forma preponderante, su regulación obedece al modelo de **seguridad de los productos**. Como resultado, es una herramienta de regulación del mercado interior. Al haber utilizado como base jurídica primaria del Reglamento el artículo 114 del Tratado de Funcionamiento de la Unión Europea (Tratado FUE), además de la protección de datos personales (Tratado FUE art.16), el legislador sigue un enfoque tradicional de la gobernanza del mercado interior de la UE, basado en la armonización de las normas de comercialización de los sistemas de inteligencia artificial (IA) en los Estados miembros. Por eso emplea una terminología desconocida hasta ahora en el Derecho digital, pero típica de la regulación de productos (autoridades notificantes, organismos de evaluación de la conformidad, organismos notificados, procedimientos de evaluación de la conformidad, marcado «CE», autoridades de vigilancia del mercado...).

En segundo lugar, también quiere situar la protección de los **derechos fundamentales** al más alto nivel, para diferenciar el modelo regulatorio europeo del norteamericano o del chino. Así, la segmentación *ex lege* de los sistemas de IA en varios niveles de riesgo sitúa los riesgos para los derechos fundamentales en pie de igualdad con los riesgos para la salud y la seguridad, que son los elementos nucleares de la legislación sobre seguridad de los productos (así, Directiva 2001/95/CE cdo.4). Varios requisitos esenciales establecidos para los sistemas de IA de alto riesgo se formulan en términos de derechos fundamentales, como la necesidad de indicar las circunstancias en las que el uso del sistema de IA puede entrañar riesgos (RIA art.13.3.b.iii), o la obligación de diseñar mecanismos adecuados para la supervisión humana del sistema de IA (RIA art.14.2). Por último, debe evaluarse la conformidad con los requisitos esenciales, teniendo en cuenta en qué medida un sistema de IA minimiza o elimina los riesgos para los derechos fundamentales (RIA art.43.6). A la postre, la protección iusfundamental sigue siendo uno de los principios estructurales de la norma (aunque no el primero, según el tenor literal del artículo 1 del RIA).

Así las cosas, el RIA positiviza **normas armonizadas** en esta materia a escala de la UE con el objeto de mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme, en particular para el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial, a fin de promover la adopción de una IA centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección

de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea (CDFUE).

Del mismo modo, la norma garantiza la **libre circulación transfronteriza de productos y servicios** basados en IA, con lo que impide que los Estados miembros impongan restricciones al desarrollo, la comercialización y la utilización de sistemas de IA, a menos que el RIA lo autorice expresamente.

Con estos objetivos, el RIA adopta un enfoque basado en el **riesgo**, estableciendo un marco de obligaciones y requisitos distintos en función del nivel de riesgo de la tecnología de IA aplicable y su uso concreto. El RIA se construyó partiendo del modelo de la legislación europea preexistente relativa a la seguridad de los productos, añadiendo las particularidades de la IA que los expertos analizamos en el Libro Blanco de la IA de la Comisión Europea, de 19 de febrero de 2020 y titulado «Un enfoque europeo orientado a la excelencia y la confianza».

El RIA establece un **marco regulatorio horizontal** –es decir, no limitado a sectores concretos–, y ofrece una respuesta proporcional al riesgo generado por cada sistema de IA. Es muy diferente la regulación que tienen un coche autónomo (sistema de alto riesgo, con las obligaciones más estrictas) que el corrector ortográfico de Word (sistema de nulo riesgo, sin requisitos obligatorios).

A tal fin, positiviza un marco unificado previo o *ex ante* aplicable a todos los sistemas de IA puestos en el mercado de la Unión Europea, y lo hace con distintos niveles de obligaciones en atención al grado de riesgo que presenta cada sistema para la **salud, la seguridad y los derechos fundamentales** de los ciudadanos. Este enfoque basado en el riesgo lleva al RIA a prohibir determinados usos considerados de riesgo inaceptable en la UE, imponer obligaciones estrictas para usos de riesgo alto, establecer obligaciones de transparencia para ciertos sistemas de riesgo limitado, y dejar sin regulación imperativa al resto de sistemas de riesgo mínimo o nulo, respecto de los que trata de favorecer la adopción voluntaria de códigos de conducta.

El RIA asimismo contiene normas armonizadas para la introducción en el mercado de **modelos de IA de uso general** (los modelos GPT), normas sobre el seguimiento del mercado, la vigilancia del mercado, la gobernanza y la garantía del cumplimiento, así como medidas en apoyo de la innovación, prestando especial atención a las PYMES, incluidas las empresas emergentes o *startups*.

A la postre, el RIA es una norma muy compleja y con diversos problemas de calidad normativa. A mi juicio, muchas de las críticas están justificadas. Sin embargo, no deben olvidarse los numerosos beneficios para la integración europea que la legislación recién aprobada pretende aportar. Como mínimo, el Reglamento es un buen punto de partida en la creciente necesidad política de regular la IA en una amplia gama de ámbitos. Tanto si se trata de establecer condiciones para una **gobernanza** eficaz de la IA en el ámbito del sector público como de un marco de seguridad jurídica en el ámbito privado, el RIA pretende un equilibrio satisfactorio entre los objetivos económicos y los derechos fundamentales de los ciudadanos.

A partir de ahora, el Reglamento requiere que los Estados miembros desarrollen algunos aspectos, como el **régimen sancionador** o las **autoridades de supervisión**. En mi opinión, la norma supondrá un antes y un después en la materia, y probablemente sitúe a la Unión Europea a la cabeza global de la regulación jurídica de la IA, como en su momento ocurrió con la protección de datos (el llamado «efecto Bruselas»).

La obra que el lector tiene en sus manos está escrita no sólo desde mi posición académica, sino especialmente como abogado y consultor. También conozco muy intensamente las particularidades del mundo tecnológico, dado que en 1992 fundé una empresa fabricante de software. Asimismo, he sido uno de los pocos españoles que han tenido la fortuna de participar en el proceso de elaboración del RIA. Por eso, la obra va directamente al grano de las cuestiones que plantea una norma, probablemente la más compleja de todo el Derecho digital, y quiere aportar claves prácticas para su adecuada comprensión y aplicación. Concluyo agradeciendo a Juan Pujol, presidente de Lefebvre, su encargo y confianza mostrada en este proyecto editorial que abre una nueva colección editorial dedicada a las cuestiones jurídicas que suscita la IA.

Madrid, abril de 2025.

Moisés BARRIO ANDRÉS

*Letrado del Consejo de Estado, J.D., Ph.D., Árbitro y Abogado
Asesor de diversos Estados y de la Unión Europea en materia de regulación digital
Director del Posgrado en Legal Tech y transformación digital y codirector del Posgrado
en IA y Derecho de la EPJ de la Universidad Complutense de Madrid*

<https://www.moisesbarrio.es/>

@moisesbarrioa

Capítulo 1. Disposiciones generales

A. Ámbito de aplicación.....	20	10
1. Ámbito de aplicación material	25	
2. Ámbito de aplicación subjetivo	90	
3. Ámbito territorial de aplicación.....	105	
4. Excepciones al ámbito de aplicación	115	
B. Alfabetización en materia de IA.....	175	11
C. Niveles de riesgo.....	180	

El Reglamento de Inteligencia Artificial (**RIA**) es el resultado de muchos años de debate y de un amplio trabajo preparatorio a escala internacional, europea y española.

A **nivel internacional**, diferentes organizaciones han estado y están involucradas en la regulación de la IA. La **OCDE** fue la primera, y probablemente la más influyente, en empezar a abordar el tema en 2016 con el Foro de Prospectiva Tecnológica sobre IA. Sus conclusiones fueron luego desarrolladas como principios por un grupo de expertos en IA de la OCDE (AIGO), y finalmente fueron adoptados el 22-5-2019 como la «Recomendación del Consejo de la OCDE sobre inteligencia artificial», posteriormente revisada el 3-5-2024. Esta recomendación ha influido notablemente en el texto del RIA y su huella puede encontrarse, por ejemplo, en el RIA art.9 a 15.

PRECISIONES Entre otras iniciativas internacionales, cabe citar la «Recomendación sobre la ética de la inteligencia artificial» publicada por la **UNESCO** el 23-11-2021, y adoptada por 193 Estados, que se centra en cómo puede afectar la IA a los colectivos vulnerables, la sostenibilidad medioambiental y la alfabetización digital.

Del mismo modo, la «**Declaración de Betchley**» de la Cumbre de seguridad de la IA de noviembre de 2023 se enfoca en la colaboración entre los Estados participantes para detectar los riesgos comunes de seguridad de la IA, así como del desarrollo conjunto de políticas para mitigarlos.

En 2023 comenzó a estar disponible la IA generativa, y la Declaración del G-7 sobre el «**Proceso de IA de Hiroshima**», de 30-10-2023, introdujo un código de conducta con mecanismos de seguimiento y promoviendo la colaboración internacional.

El 17-5-2024, el Comité sobre Inteligencia Artificial del Consejo de Europa aprobó el **Convenio Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho**, que constituye el primer tratado internacional jurídicamente vinculante destinado a garantizar el respeto de los derechos humanos, del Estado de derecho y las normas jurídicas democráticas en el uso de los sistemas de IA.

También hay que citar los trabajos de la Organización de las Naciones Unidas. Tras su recomendación de la UNESCO de 2021, la **ONU** creó en 2023 el Órgano Consultivo de Alto Nivel sobre Inteligencia Artificial de las Naciones Unidas. En marzo de 2024, la Asamblea General aprobó una resolución liderada por los Estados Unidos que pide a los Estados que se abstengan de utilizar sistemas de IA que no cumplan las normas internacionales de derechos humanos o las pongan en riesgo. En septiembre de 2024, el Órgano Consultivo sobre IA publicó su informe «Gobernanza de la IA en beneficio de la humanidad», que propone la creación de un grupo internacional de científicos, de foros internacionales sobre normas y polí-

ticas de gobernanza IA, de una red de centros de desarrollo de capacidades en IA, de un marco mundial de datos, y de un fondo de financiación para reducir la brecha en la materia.

- 14** En **España**, la Carta de Derechos digitales, de 14-7-2021, se ocupa de establecer las orientaciones básicas regulatorias de la IA en su numeral XXV, que aparece presidido por el mandato de que dicho conjunto de tecnologías debe asegurar un enfoque centrado en la persona y su inalienable dignidad, perseguir el bien común y asegurar cumplir con el principio de no maleficencia, al tiempo que reconoce el derecho de los ciudadanos a solicitar una supervisión e intervención humana y a impugnar las decisiones automatizadas que produzcan efectos en su esfera personal y patrimonial.

Sus previsiones tuvieron una positivización en la L 15/2022 art.23 –**Ley integral para la igualdad de trato y la no discriminación**–, que establece la primera regulación jurídica de la IA en nuestro ordenamiento en espera de la plena aplicación del RIA.

Algunas comunidades autónomas, como Extremadura o Galicia, han dictado normas autonómicas aplicables a su respectivo sector público inspiradas en estos postulados.

- 15** En **Europa**, fue inicialmente el Parlamento Europeo el que abrió sus trabajos con un informe de iniciativa legislativa propia entre 2015 y 2017, en el que se instaba a la Comisión a elaborar una directiva en el ámbito de las normas de Derecho civil para los robots. El resultado fue la importante Resolución del Parlamento Europeo, de 16-2-2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (Resol Parlamento Europeo 2015/2103 (INL)).

Esta recomendación motivó que las tres instituciones de la UE se pronunciaran a favor de regular la IA como prioridad legislativa en una declaración conjunta el 14-12-2017. Luego, la Comisión Europea comenzó su trabajo en marzo de 2018 con la creación de grupos de expertos (AI HLEG). Además de los informes sobre responsabilidad, cabe destacar las «**Directrices Éticas** del Grupo de Expertos de Alto Nivel sobre IA», de 8-4-2019, que ya intentaban proporcionar a las empresas estrategias concretas de aplicación.

- 16** Tras las elecciones europeas de 2019 y la subsiguiente declaración de la Presidenta de la Comisión de que presentaría una «Ley europea» sobre IA en un plazo de 100 días, la Comisión publicó su «**Libro Blanco de la IA: un enfoque europeo orientado a la excelencia y la confianza**» el 19-2-2020, que proponía la creación de un ecosistema para la excelencia y un ecosistema de confianza.

La propuesta legislativa de la Comisión, que incluía muchos puntos de los trabajos preparatorios a escala internacional, europea y nacional, se publicó finalmente el 21-4-2021.

Por último, las políticas de inversiones públicas y medidas de fomento se han establecido en la reciente Comunicación de la Comisión Europea titulada «**Plan de Acción para el Continente de la IA**», de 9-4-2025, con el objetivo de competir con EE.UU. y China y convertir a Europa en líder mundial en IA. Sus medidas incluyen la construcción de una infraestructura informática y de datos de IA a gran escala en la UE, aumentar el acceso a datos de gran tamaño y de alta calidad, fomento de la adopción de la IA en sectores estratégicos, reforzar las capacidades y los talentos europeos en IA, así como una simplificación normativa del

RIA para PYMES y *startups*, que incluye un Servicio de Asistencia del RIA (el AI Act Service Desk) para ayudar a las empresas a cumplir con la norma.

A. **Ámbito de aplicación**

(RIA art.2)

El RIA (art.2) define el ámbito de aplicación en términos materiales, subjetivos y territoriales. Los requisitos centrales figuran en el apartado 1 del precepto. En el apartado 2 se establece una normativa especial para los sistemas de IA de alto riesgo vinculados a un producto (RIA art.6.1 y anexo I.B). Y las excepciones al ámbito de aplicación se establecen en el apartado 3 y siguientes. **20**

El ámbito de aplicación, mucho más matizado en la versión final que en la propuesta inicial de la Comisión, se explica con más detalle en el RIA cdo.21 a 25.

I. **Ámbito de aplicación material**

Como punto de partida, es de vital importancia que el sector público y privado diferencien con precisión si el uso de IA es para un fin general o si se trata de software que no entra en la definición de sistema de IA según el RIA art.3.1. **25**

a. **Concepto de inteligencia artificial en el RIA**

La determinación del ámbito de aplicación del RIA comienza con el RIA art.2.1 con su aplicación a los sistemas de IA (RIA art.3.1) y los modelos de IA de uso general (RIA art.3.69) –los llamados modelos GPAI por sus siglas en inglés–. **30**

Sistema de IA El término sistema de IA, que se debatió intensamente en el proceso legislativo y en la doctrina, se caracteriza por un **concepto muy amplio** en línea con la definición de la OCDE, que se actualizó en 2023. Al final, este resultado representó el único compromiso posible entre las distintas posiciones en las negociaciones políticas en Bruselas. Mientras que los partidos de izquierda querían que casi todas las tecnologías quedaran incluidas en el RIA, los Estados miembros y los partidos liberal-conservadores exigían que la norma se centrara únicamente en las tecnologías de IA más recientes, como el aprendizaje automático. Este último grupo también rechazó la propuesta inicial de la Comisión, incluido su proyectado anexo I, ya que también englobó diversos métodos estadísticos como IA. **35**

La definición de la **OCDE**, leída en consonancia con el RIA cdo.12, estipula ahora en el texto legal definitivo que no se incluyen los sistemas de software y enfoques de programación convencionales más sencillos, pero al mismo tiempo deja a las autoridades competentes europeas y nacionales suficiente margen de interpretación para incluir sistemas híbridos de IA con enfoques antiguos y nuevos (por ejemplo, aprendizaje automático para la recopilación de datos, pero análisis mediante Microsoft Excel como sucedió en el escándalo holandés de las prestaciones por hijos a cargo).

La definición de IA del RIA se basa esencialmente en siete **elementos**: **37**

- es un sistema basado en una **máquina**;
- que está diseñado para funcionar con distintos niveles de **autonomía**;

- que puede mostrar capacidad **de adaptación** tras su despliegue;
- y para **objetivos explícitos o implícitos**;
- **infiere**, a partir de la información de entrada que recibe, la **manera de generar resultados de salida**;
- tales como **predicciones, contenidos, recomendaciones o decisiones**; y
- que pueden **influir en entornos físicos o virtuales**.

Según el RIA cod.12, este enfoque debe, por una parte, ser suficientemente compatible a nivel internacional y técnico. Por otra parte, la definición debe ser suficientemente específica y permitir diferenciar la IA de los sistemas de software convencionales y de las operaciones simples basadas en reglas automáticas de la programación informática tradicional (el *if-else*).

Asimismo, la definición no debe ser aplicada mecánicamente, sino que es preciso evaluar las **características** de cada sistema en cuestión.

- 40** Con fecha 6-2-2025, la Comisión Europea ha publicado las **Directrices sobre la definición de sistema de inteligencia artificial establecida por el RIA**, en ejecución del RIA art.96.1.f, que no tienen carácter vinculante. La interpretación final vendrá de la mano del TJUE en eventuales litigios que puedan plantearse.

Las Directrices precisan que la definición de sistema de IA adopta una perspectiva basada en el ciclo de vida que abarca dos **fases** principales:

- la fase previa al despliegue, implantación o fase de **construcción** del sistema; y
- la fase posterior al despliegue o fase de **uso** del sistema.

Los siete elementos establecidos en dicha definición no requieren estar presentes de manera continua a lo largo de ambas fases del ciclo de vida. En cambio, la definición reconoce que ciertos elementos pueden manifestarse en una fase, pero no necesariamente persistir en ambas.

A continuación, examinaremos cada uno de los elementos indicados.

- 45 Sistema basado en una máquina** El término «basado en una máquina» se refiere a que los sistemas de IA se desarrollan y operan mediante máquinas. El concepto de «máquina» incluye tanto los **componentes** de hardware como de software que permiten el funcionamiento del sistema de IA.

Los componentes de **hardware** se refieren a los elementos físicos de la máquina, tales como unidades de procesamiento, memoria, dispositivos de almacenamiento, unidades de red e interfaces de entrada y salida, que proporcionan la infraestructura física para la computación.

Los componentes de **software** abarcan el código informático, las instrucciones, los programas, los sistemas operativos y las aplicaciones que gestionan cómo el hardware procesa datos y realiza tareas.

Y comprende una amplia variedad de sistemas computacionales. Por ejemplo, los **sistemas de computación cuántica** emergentes más avanzados, que representan una notable diferencia respecto a los sistemas computacionales tradicionales, constituyen sistemas basados en máquinas, a pesar de sus principios operativos únicos y el uso de fenómenos mecánico-cuánticos, al igual que los **sistemas biológicos u orgánicos**, siempre que proporcionen capacidad computacional.

- 50 Autonomía** El segundo elemento de la definición significa que el sistema está diseñado para funcionar con distintos niveles de autonomía. El RIA cdo.12 aclara que los distintos niveles de autonomía suponen que los sistemas de IA están

diseñados para operar con cierto grado de independencia con respecto a la actuación humana y tienen ciertas capacidades para funcionar **sin intervención humana**.

El concepto de autonomía tiene como eje central la intervención humana y la interacción humano-máquina. En un extremo de la posible interacción humano-máquina se encuentran los sistemas diseñados para realizar todas las tareas mediante funciones operadas manualmente. En el otro extremo están los sistemas capaces de operar sin ninguna intervención o participación humana, es decir, de manera plenamente autónoma.

La referencia a «cierto grado de independencia» de acción en el RIA cdo.12 **51** **excluye** los sistemas diseñados para operar únicamente con **intervención humana completa y manual**. La intervención humana puede ser directa, por ejemplo, mediante controles manuales, o indirecta, por ejemplo, a través de controles basados en sistemas automatizados que permiten a los humanos delegar o supervisar las operaciones del sistema.

Por lo tanto, el nivel de autonomía es una **condición necesaria** para determinar si un sistema se define como sistema de IA. Todos los sistemas diseñados para operar con un grado razonable de independencia de acción cumplen con la condición de autonomía de la definición de sistema de IA. Se trata, así, del rasgo clave para distinguirlo de los sistemas tradicionales que son dependientes de la intervención humana y no puede funcionar sin la misma (por ejemplo, el procesador de textos Word).

Capacidad de adaptación El tercer elemento de la definición es que el sistema puede mostrar capacidad de adaptación tras el despliegue. **55**

Los conceptos de autonomía y adaptabilidad o capacidad de adaptación son distintos, pero están estrechamente relacionados. A menudo se discuten juntos, pero representan diferentes dimensiones de la funcionalidad de un sistema de IA.

El RIA cdo.12 aclara que la capacidad de adaptación se refiere a las capacidades de **autoaprendizaje**, permitiendo que el comportamiento del sistema cambie mientras está en uso. El nuevo comportamiento del sistema adaptado puede producir resultados diferentes a los del sistema anterior para las mismas entradas.

El uso del término «puede» en relación con este elemento de la definición indica que un sistema puede, pero no necesariamente debe, poseer capacidades de adaptación o autoaprendizaje tras su despliegue para constituir un sistema de IA. **56**

En consecuencia, la capacidad de un sistema para aprender automáticamente, descubrir nuevos patrones o identificar relaciones en los datos más allá de lo que se le entrenó inicialmente es una **condición facultativa** y, por lo tanto, no decisiva para determinar si el sistema califica como un sistema de IA, según defienden las directrices interpretativas de la Comisión.

Objetivos explícitos o implícitos El cuarto elemento de la definición son los objetivos del sistema de IA. Los sistemas de IA están diseñados para operar conforme a uno o más objetivos. Estos objetivos pueden definirse de forma explícita o implícita. **60**

Los objetivos **explícitos** se refieren a metas claramente establecidas que son directamente programadas por el fabricante en el sistema. Por ejemplo, según las

directrices, pueden especificarse como la optimización de una función de coste, una probabilidad o una recompensa acumulada. Así, en un sistema de detección de *spam* basado en IA, el objetivo explícito podría ser clasificar correos electrónicos como «**spam**» o «**no spam**» con un 95% de precisión.

- 61** Los **objetivos implícitos** se refieren a metas que no están explícitamente declaradas, pero que pueden deducirse del comportamiento o de las suposiciones subyacentes del sistema.

Estos objetivos pueden surgir de los datos de entrenamiento o de la interacción del sistema de IA con su entorno. En el ejemplo anterior, un objetivo implícito podría ser priorizar la identificación de correos en **inglés sobre otros idiomas**, deducido del hecho de que los datos de entrenamiento contienen principalmente correos en inglés, lo que lleva al sistema a desempeñarse mejor en ese idioma sin que esto haya sido un objetivo programado.

- 62** El RIA cdo.12 aclara que los objetivos del sistema de IA pueden ser diferentes de la finalidad prevista del sistema de IA en un contexto específico. Los objetivos de un sistema de IA son **internos al sistema**, refiriéndose a las metas de las tareas a realizar y sus resultados. Por ejemplo, un sistema de asistente virtual corporativo de IA puede tener como objetivo responder preguntas de los usuarios sobre un conjunto de documentos con alta precisión y una baja tasa de errores.

En contraste, la **finalidad prevista** es de carácter externa e incluye el contexto en el que el sistema está diseñado para ser desplegado y cómo debe operarse.

- 63** De hecho, según el RIA art.3.12, la finalidad prevista de un sistema de IA se refiere al **uso** para el que un proveedor concibe un sistema de IA, incluidos el contexto y las condiciones de uso concretos, según la información facilitada por el proveedor en las instrucciones de uso, los materiales y las declaraciones de promoción y venta, y la documentación técnica. Por ejemplo, en el caso de un sistema de asistente virtual corporativo de IA, la finalidad prevista podría ser asistir a un determinado departamento de una empresa en la realización de ciertas tareas.

Esto podría requerir que los documentos que utiliza el **asistente virtual** cumplan con ciertos requisitos (p.e., longitud, formato) y que las preguntas de los usuarios se limiten al dominio en el que el sistema está destinado a operar.

Esta finalidad prevista se cumple no solo a través del funcionamiento interno del sistema para alcanzar sus objetivos, sino también mediante otros factores, como la integración del sistema en un flujo de trabajo de atención al cliente más amplio, los datos utilizados por el sistema o las instrucciones de uso.

- 65 Inferencia** El quinto elemento de un sistema de IA es que infiere de la información de entrada que recibe la manera de generar resultados de salida. El considerando 12 del RIA aclara que es una característica clave. Como se explica además en dicho considerando, los sistemas de IA deben distinguirse de «los sistemas de software o los planteamientos de programación tradicionales y más sencillos, y no debe incluir los sistemas basados en las normas definidas únicamente por personas físicas para ejecutar automáticamente operaciones».

Esta capacidad de inferencia es, por lo tanto, una **condición indispensable** que distinga a los sistemas de IA de otros tipos de sistemas de software que no son IA.

- 66** El RIA cdo.12 también explica que este elemento se refiere al proceso de obtención de **resultados de salida**, como predicciones, contenidos, recomendaciones o

decisiones, que puede influir en entornos físicos y virtuales, y a la capacidad de los sistemas de IA para deducir modelos o algoritmos, o ambos, a partir de información de entrada o datos.

Como explican las directrices de la Comisión, esta interpretación del concepto de «inferencia» no contradice la **norma ISO/IEC 22989**, que define la inferencia como razonamiento mediante el cual se derivan conclusiones a partir de premisas conocidas y esta norma técnica incluye una nota específica sobre IA que establece: en IA, una premisa es un hecho, una regla, un modelo, una característica o datos brutos.

A renglón seguido, el propio considerando aclara que las técnicas que permiten la inferencia al construir un sistema de IA **incluyen** estrategias de aprendizaje automático que aprenden de los datos cómo alcanzar determinados objetivos y estrategias basadas en la lógica y el conocimiento que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse. La capacidad de inferencia de un sistema de IA trasciende el tratamiento básico de datos, al permitir el aprendizaje, el razonamiento o la modelización.

Las directrices de la Comisión establecen distintos ejemplos de técnicas de IA que habilitan la inferencia, que incluyen enfoques de aprendizaje automático (el popular **machine learning** y sus subdivisiones como el aprendizaje supervisado, el aprendizaje no supervisado, el aprendizaje autosupervisado, el aprendizaje por refuerzo y el aprendizaje profundo) que aprenden de los datos cómo alcanzar determinados objetivos, y enfoques basados en la **lógica y el conocimiento** que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse (aquí se incluyen los sistemas de representación del conocimiento, la programación lógica inductiva, las bases de conocimiento, los motores de inferencia y deducción, el razonamiento (simbólico), los sistemas expertos y los métodos de búsqueda y optimización).

Según las directrices de la Comisión, se **excluyen**:

- los sistemas para mejorar la optimización matemática (como los métodos de regresión lineal o logística);
- el procesamiento básico de datos (como las hojas de cálculo sin IA o sistemas de gestión de bases de datos utilizados para ordenar o filtrar datos según criterios específicos);
- los sistemas basados en heurísticas clásicas (p.e., un programa de ajedrez que utiliza un algoritmo *minimax*); y
- los sistemas de predicción simples (que utilizan reglas estadísticas básicas, como el promedio de temperaturas de la última semana).

Por tanto, bajo el RIA la inferencia debe entenderse como referida a la **fase de construcción**, mediante la cual un sistema deriva resultados a través de técnicas de IA que habilitan la inferencia. Y es el proceso mediante el cual el sistema utiliza la información disponible para deducir o calcular un resultado, basándose en modelos, algoritmos o reglas previamente establecidos durante su fase de entrenamiento o construcción.

67

68

70 Resultados de salida como predicciones, contenidos, recomendaciones o decisiones El sexto elemento de la definición es que el sistema infiere la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones.

En efecto, la capacidad de un sistema para generar resultados, como **predicciones, contenidos, recomendaciones o decisiones**, basándose en los datos de entrada que recibe y utilizando enfoques de aprendizaje automático o enfoques basados en lógica y conocimiento, es fundamental para distinguir lo que hacen los sistemas de IA y lo que los diferencia de otras formas de software.

La capacidad para generar resultados y el tipo de resultado que el sistema puede generar son **condiciones indispensables** para comprender la funcionalidad y el impacto de un sistema de IA. Cada categoría de salida difiere en su nivel de intervención humana.

71 a) Las **predicciones** son uno de los resultados de salida más comunes que producen los sistemas de IA y que requieren la menor intervención humana. Una predicción es una **estimación** sobre un valor desconocido (el resultado) a partir de valores conocidos suministrados al sistema (las entradas). Los sistemas de software se han utilizado durante décadas para generar predicciones. Los sistemas de IA que utilizan aprendizaje automático son capaces de generar predicciones que descubren patrones complejos en los datos y hacen predicciones precisas en entornos altamente dinámicos y complejos. Por ejemplo, los sistemas de IA desplegados en automóviles autónomos están diseñados para hacer predicciones en tiempo real en un entorno extremadamente complejo y dinámico, con múltiples tipos de agentes e interacciones, y un número prácticamente infinito de situaciones posibles, y toman decisiones para ajustar su comportamiento en consecuencia.

72 b) Los **contenidos** consisten en la generación de nuevo material por parte de un sistema de IA. Pueden ser **texto, imágenes, videos, música** u otros tipos de resultados de salida (o una combinación de los mismos). Hay un número creciente de sistemas de IA que utilizan modelos de aprendizaje automático (p.e., basados en tecnologías GPT, Generative Pre-trained Transformer, para generar contenido).

c) Las **recomendaciones** se refieren a **sugerencias** de acciones, productos o servicios específicos para los usuarios basadas en sus preferencias, comportamientos u otros datos de entrada. Los sistemas de recomendación basados en IA, por ejemplo, pueden aprovechar datos a gran escala, adaptarse al comportamiento del usuario en tiempo real, proporcionar recomendaciones altamente personalizadas y escalar eficientemente a medida que crece el conjunto de datos. En cambio, estas funcionalidades no están presentes en los sistemas que no utilizan IA, y por ello dependen de mecanismos estáticos basados en reglas y datos limitados. En otros casos, las recomendaciones se refieren a decisiones potenciales, como un candidato a contratar en un sistema de selección de personal, que serán evaluadas por humanos. Si estas recomendaciones se aplican automáticamente, se convierten en decisiones.

73 d) Las **decisiones** se refieren a **conclusiones o elecciones** tomadas por un sistema. Un sistema de IA produce una decisión automatizada en procesos que tradicionalmente son manejados por el decisor humano. Tal sistema implica un proceso completamente automatizado mediante el cual se produce un determinado

resultado en el entorno que rodea al sistema sin ninguna intervención humana (p.e., un *roboadvisor*, que gestiona automáticamente una cartera de valores).

En resumen, como observan las directrices de la Comisión, los sistemas de IA, incluidos los sistemas basados en enfoques de aprendizaje automático y sistemas basados en lógica o conocimiento, difieren de los sistemas no basados en IA en su capacidad para generar resultados como predicciones, contenidos, recomendaciones y decisiones en los que pueden manejar relaciones y patrones complejos en los datos. Los sistemas de IA pueden producir generalmente resultados más matizados que otros sistemas, por ejemplo, aprovechando patrones aprendidos durante el entrenamiento o utilizando reglas definidas por expertos para tomar decisiones, ofreciendo un razonamiento más sofisticado en entornos estructurados.

Influencia en entornos físicos o virtuales El séptimo y último elemento de la definición de un sistema de IA es que los resultados del sistema pueden influir en entornos físicos o virtuales. Este elemento debe entenderse como un énfasis en el hecho de que los sistemas de IA no son pasivos, sino que impactan **activamente** en los entornos en los que se despliegan. **75**

La referencia a entornos físicos o virtuales indica que la influencia de un sistema de IA puede ser tanto sobre **objetos físicos** tangibles (p.e., un brazo robótico) como sobre **entornos virtuales**, incluyendo espacios digitales, flujos de datos y ecosistemas de software (plataformas, redes sociales, metaversos, etc.).

En todo caso, es poco probable que este último criterio aporte mucha claridad adicional, ya que, en última instancia, todo el software (incluido el ordinario) tiene un impacto potencial en el entorno (al menos en el virtual).

Modelos de IA de uso general (GPAI) (RIA art.3.63) Un modelo de IA de uso general, un modelo GPAI por sus siglas inglesas de general-purpose AI model, es un modelo que puede desempeñar una amplia gama de **tareas** distintas. Se define como un modelo de IA, también uno entrenado con un gran volumen de datos utilizando autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas, independientemente de la manera en que el modelo se introduzca en el mercado, y que puede integrarse en diversos sistemas o aplicaciones posteriores, excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su introducción en el mercado. **80**

Puede **integrarse** en diversos sistemas o aplicaciones posteriores. Los modelos GPAI suelen entrenarse con grandes cantidades de datos utilizando diferentes métodos, como el aprendizaje supervisado, el aprendizaje no supervisado y el aprendizaje por refuerzo.

Los elementos de esta definición se remontan a la definición de **modelo fundamental o modelo de base** (*foundation model*) de la Universidad de Stanford, que es probablemente el intento más reconocido a nivel mundial de describir con más detalle este nuevo campo de la inteligencia artificial. Al final, en el RIA se decidió no utilizar el término modelos de base, a pesar de ser el dominante en la industria, y utilizar en su lugar modelo de IA de uso general, ya que se esperaba que este término estuviera más abierto a futuros desarrollos y no se comprometiera demasiado con una dirección técnica de desarrollo. **81**

Además, la referencia al uso o propósito general en el término elegido encaja mejor con el enfoque de seguridad de los productos del RIA. No obstante, en su definición jurídica se han adoptado la mayoría de los elementos básicos de la definición de Stanford. La **finalidad general** viene determinada por la capacidad de realizar de forma competente una amplia gama de tareas distintas (RIA cdo.97).

- 82** Los grandes modelos de lenguaje de IA (o LLMs, Large Language Models), como **GPT-4, Claude 3 o LLaMA 2**, son un ejemplo típico de modelos GPAL, ya que permiten la generación flexible de contenidos, por ejemplo, en forma de texto, audio, imagen o vídeo, que pueden asignar fácilmente una amplia gama de tareas distintas (RIA cdo.99). Los modelos GPAL pueden **comercializarse** de varias maneras, por ejemplo, mediante bibliotecas, interfaces de programación de aplicaciones (API), descarga directa o como copia física (RIA cdo.97).

Aunque los modelos GPAL (p.e., GPT-4 de OpenAI) son componentes esenciales de los sistemas de IA, **no constituyen por sí mismos sistemas de IA**. Sólo pueden convertirse en sistemas de IA añadiendo necesariamente otros componentes, como una interfaz a través de la cual los usuarios puedan interactuar con el sistema (p.e., la aplicación ChatGPT de OpenAI).

El RIA regula los modelos GPAL y aquellos que también albergan un riesgo sistémico si se utilizan para el mismo fin general.

b. Bienes jurídicos protegidos

- 85** Según el RIA art. I.1, el **objetivo** de la **norma** es mejorar el funcionamiento del mercado interior y promover la adopción de una IA centrada en el ser humano y fiable. Al mismo tiempo, debe garantizarse un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la UE, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, así como prestar apoyo a la innovación.

PRECISIONES Desde una perspectiva de derechos fundamentales, un sector doctrinal hemos señalado que el RIA es el resultado de las restricciones constitucionales de la Unión Europea. Dado que la UE no tiene una competencia general para armonizar los derechos fundamentales en los Estados miembros, el Reglamento se basa en cambio en su competencia para promover el mercado interior (TFUE art. I 14). Como resultado, el RIA no sigue un enfoque de derechos fundamentales, sino que utiliza la legislación sobre seguridad de los productos y la regulación de los riesgos. Sin embargo, este enfoque no es suficientemente adecuado para proteger los derechos fundamentales.

- 86** Ahora bien, los intereses protegidos de la democracia, el Estado de Derecho y la protección del medio ambiente, que también se mencionan, sólo se protegen mediante la regulación de la IA si entran dentro del ámbito de aplicación de la **Carta de los Derechos Fundamentales de la UE** (CDFUE), debido a la palabra «incluidos» en el apartado I.

Un caso concreto es el **medio ambiente**, que queda así estrechamente limitado por la CDFUE art.37.

- 87** Este variopinto ramillete de intereses jurídicos protegidos por un derecho subjetivo, por un lado, y los derechos colectivos, por otro, queda poco claro en el RIA