REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

DIRECTOR • D. PABLO GARCÍA MEXÍA

NÚMERO MONOGRÁFICO
REGLAMENTO EUROPEO DE IA
UAM
Universidad Autónoma de Madrid

AÑO IX • MAYO-AGOSTO 2024 • NÚMERO 34

ISSN: 2444-5762

REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

AÑO IX • MAYO-AGOSTO 2024 • NÚMERO 34

Esta revista se edita en Madrid, por RDU, revistas especializadas

ISSN versión digital 2444-5762

Depósito Legal M-6283-2016

Esta revista puede verse en línea en www.rdu.es

PRECIOS PARA 2024

España: 197,60 € (IVA inc.)

Todos los precios son en euros. IVA aplicable según la disposición vigente en cada momento.

REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

Esta revista nace en el año 2015 con el propósito de convertirse en una publicación de referencia, que aborde, desde una perspectiva jurídica, todos los aspectos relacionados con dos conceptos tan fundamentales como la privacidad e internet.

Pese a tratarse de conceptos que podrían estudiarse por separado, dadas sus especiales características individuales, la realidad nos muestra constantemente que la evolución de la privacidad se encuentra estrechamente ligada a los avances en internet y ambos conceptos, a su vez, relacionados con dos importantísimos aspectos como la economía digital y la ciberseguridad. De ahí la decisión de crear una publicación que tratara conjuntamente todos estos temas.

La presente revista se presenta como una publicación científica, de espíritu jurídico, por lo que los lectores podrán encontrar en ella estudios doctrinales, trabajos de divulgación, comentarios legislativos, recensiones de libros, jurisprudencia, etc.

No obstante, teniendo en cuenta la relativa novedad de los temas objeto de estudio y, sobre todo, siendo conscientes de la constante evolución de los mismos, por su condición de materias vivas, esta revista tratará igualmente de centrar su foco en la aplicación práctica del derecho, a través de artículos de opinión, casos prácticos y derecho comparado, entre otros.

RPDD: Revista cuatrimestral de enero a diciembre.

CONSEJO ASESOR CIENTÍFICO *(cont.)*

Dr. D. Jesús Banegas Núñez.
Expresidente Ametic. Presidente Foro de la Sociedad Civil

Dª. Alicia Coloma Duato.
Asociada Senior Privacidad, IT, Entornos Digitales y Life Sciences en Broseta Abogados

Dra. Dª. Mercedes Fuertes López.
Catedrática de Derecho Administrativo de la Universidad de León

Dra. Dª. Olga Gil.
Profesora de Politología de la UCM. Experta en gobernanza Tech e IA

Dra. Dª. Estrella Gutiérrez David.
Profesora Contratada Doctora de Derecho Digital en la UCM

Dra. Dª. Andrea Isabel Lucas Garín.
Directora del Instituto de Investigación del Derecho, de la Facultad de Derecho de la Universidad Autónoma de Chile

Dra. Dª. Ana de Marcos Fernández.
Profesora Contratada Doctora de Derecho Administrativo de la UAM

Dr. D. José Javier Martínez Herraiz.
Delegado del Rector para Seguridad de la Información de la Universidad de Alcalá de Henares

Dr. D. Juan José Montero Pascual.
Catedrático de Derecho Administrativo de la UNED

Dr. D. Ricardo Rivero Ortega.
Rector y Catedrático de Derecho Administrativo de la Universidad de Salamanca

Dr. D. Francisco Ros Perán.
Ex secretario de Estado de Telecomunicaciones. Exconsejero de Qualcomm

Dª. Rosa Touris López.
Experta en Ciberseguridad. Miembro de las Fuerzas y Cuerpos de Seguridad del Estado

Dr. D. Jorge Villarino Marzo.
Letrado de las Cortes (exced.) Director de regulación Vinces Consulting

NORMAS PARA LA REMISIÓN DE TRABAJOS PARA SU PUBLICACIÓN

DE INTERÉS PARA LOS AUTORES

Muy importante: **La no observancia de alguna de las normas que a continuación se detallan motivará el rechazo del artículo enviado sin entrar en su valoración científica, siendo devuelto de inmediato para la subsanación de los incumplimientos, defectos o vacíos detectados.**

La presentación de trabajos para su publicación está abierta a cualquier interesado. Todos los trabajos que se publiquen deberán superar un **riguroso** proceso previo de revisión por parte de Evaluadores externos a la Revista (**"pares ciegos"**), así como por el Consejo **Científico** de esta última, que valorará especialmente la originalidad, rigor e interés establecidos por la RDU, entre otros.

La remisión de trabajos para su publicación supone la cesión, por parte de todos los autores del mismo, en favor de la RDU, de un derecho exclusivo para la publicación, reproducción y comunicación pública del mismo en revistas científicas, para todo el mundo y por el plazo máximo que permita la ley.

A continuación se detallan los requisitos que deben cumplir los trabajos a efectos de su revisión:

ORIGINALIDAD Y AUTORÍA

Todo trabajo remitido para su publicación deberá ser original e inédito. En caso de que algún trabajo sea publicado en alguna otra publicación con posterioridad a su remisión a esta secretaría, el autor deberá notificarlo inmediatamente, para proceder a la devolución del mismo.

La autoría de los trabajos remitidos debe pertenecer a quien o quienes lo suscriban, lo que así se hará constar en **el modelo de declaración de autoría del trabajo (Anexo I) que deberá acompañar inexcusablemente a este último.**

DOCUMENTACIÓN

Junto con el trabajo se **deberá adjuntar, debidamente cumplimentado, el modelo de resumen ejecutivo del mismo, que se detalla en el Anexo II**.

Los datos personales que los autores faciliten durante este proceso así como cualesquiera otros datos que se generen con posterioridad como consecuencia de su relación con la RDU se incluirán en ficheros de los que RDU (con domicilio social en C/ Fray Juan Gil 7, Madrid; email: **lopd@rdu.es**) es responsable (el "Responsable"). Los datos serán tratados por el Responsable para el desarrollo, mantenimiento y control de la relación del autor con la RDU y en particular, para gestionar la publicación de trabajos y la difusión de los mismos, así como para tramitar las peticiones que el autor nos dirija, conservar las comunicaciones para mejorar nuestro servicio así como para efectos estadísticos. Para ejercitar los derechos de acceso, rectificación, cancelación y oposición en los términos legalmente previstos, el usuario debe enviar una solicitud por escrito al Responsable a la dirección o al email: **rdu@rdu.es** que se señala en esta cláusula, adjuntando copia del documento que acredite su identidad.

REMISIÓN

Los trabajos se remitirán al correo electrónico **rdu@rdu.es**. El asunto del correo deberá indicar: Remisión de Artículo para su publicación, autor/es del mismo y país de procedencia. (Ejemplo: Remisión de Artículo para publicar. Don José Fariñas Ortiz de Zúñiga. CHILE).

ESTRUCTURA Y FORMATO

La estructura de los trabajos a presentar deberá seguir el siguiente esquema:

- Declaración de autoría del artículo por el/los autores, conforme al **Anexo I**.
- Información sobre los autores y el trabajo, conforme al **Anexo II**.

CUERPO DEL TEXTO

- Título del artículo en español y traducción del mismo al inglés.
- Nombre del autor o autores, titulación y nº Orcid si disponen de él.

- Resumen del artículo y palabras clave en español y **traducción del resumen (Abstract) y palabras clave (Key Words) al inglés**.

- Sumario o índice del trabajo (obligatorio).

- Texto del artículo.

- Conclusiones, numeradas, del artículo (obligatorio).

- Bibliografía final de las fuentes bibliográficas **utilizadas** (obligatorio).

Asimismo, los trabajos deberán cumplir con las siguientes características:

- **Tanto en el Sumario como en el Texto del artículo, las partes principales en que se divida el mismo se indicarán en números romanos (I.-, II.-, III.-, etc...), los epígrafes dentro de cada uno de ellos arrastrando el correspondiente número romano y con números arábigos (II.1.-, II.2.-, II.3.-, etc...), los apartados dentro de cada epígrafe, arrastrando los respectivos números romanos y arábigos y añadiendo una letra mayúscula (III.1.A.-, III.1.B.-, etc...), los subapartados dentro de cada apartado arrastrando todos los números y letras anteriores y añadiendo números arábigos (IV.1.A.1.-, IV.1.A.2.-, etc...).**

- Extensión mínima y máxima incluyendo todas las partes que deben integrar el mismo (Título, Resumen y Palabras Clave en español/portugués, Traducción de todos ellos al inglés, Sumario, Texto, Conclusiones y Bibliografía final): 20 páginas la mínima y 30 páginas la máxima. Los trabajos realizados por dos o más autores no podrán exceder en ningún caso el 50% de la extensión máxima de los trabajos realizados individualmente, por lo que aquellos no podrán superar las 45 páginas. De hacerlo, y a criterio del Consejo Científico, podrán ser objeto de división para su publicación en números consecutivos.

- Configuración de las páginas: 2,5 cm arriba, abajo a la derecha y a la izquierda.

- Tipo de letra para el texto superior o principal: Times New Roman 12.

- Tipo de letra para las notas a pie de página: Times New Roman 10.

- Alineación: Justificada.

- Interlineado del texto superior y de las notas a pie de página: sencillo.

- Citas: Deberán figurar a pie de página, numeradas correlativamente, haciendo referencia al autor del contenido, título original de la obra, editorial, lugar, fecha y página dónde encontrar la referencia original. Las citas seguirán el modelo establecido en el **Anexo III**.

- Bibliografía utilizada: Deberá figurar al final del artículo ordenada por orden alfabético en función de los apellidos de los autores.

- Otros elementos: Si se desea, pueden aportarse fotografías, ilustraciones y/o gráficos para su publicación, incluyendo la cita y autorización correspondiente, en caso de no ser propiedad del autor, con la manifestación expresa de que los mismos no se encuentran protegidos por derechos de terceros. Dichos elementos deberán poseer una calidad mínima de 300 puntos por pulgada y deberán ser entregados en uno de los siguientes formatos: JPEG, EPS, PSD o TIFF.

PROCESO DE EVALUACIÓN

Una vez recibido el trabajo, se dará acuse de recibo del mismo y se pasará a los correspondientes expertos evaluadores (**"pares ciegos"**) que el Consejo **Científico** decida para su estudio (mínimo dos). Una vez efectuado el mismo, se dará cuenta de manera confidencial al autor o autores del resultado de la evaluación y, si esta es favorable, se procederá a la publicación del trabajo en el número y fecha que determine la dirección de la RDU.

En caso de que la evaluación del trabajo resultara desfavorable, se comunicarán al autor los motivos y le será devuelto con indicación, en su caso, de la posibilidad de que el mismo pueda ser revisado para su reevaluación.

ANEXO I

MODELO DE DECLARACIÓN DE AUTORÍA DEL TRABAJO

Señor Director de la **RDU**:

D. ..... autor/res del trabajo titulado ..... que se presenta para su evaluación y, **en su caso (tras su valoración científica positiva por "pares ciegos" externos a la Revista y supervisión ulterior por el Consejo Científico de ésta última)**, posible publicación manifiesto/ manifestamos, que soy/**somos** su/**sus** autor/**autores**, que el mismo es original, **no contiendo, por tanto, ningún tipo o clase de copia o plagio ni en su totalidad ni en ninguna parte del mismo**, que **se encuentra** inédito **en el momento de su publicación** y que los gráficos, fotografías y demás elementos de apoyo están referenciadas fielmente con sus originales y fuentes, o no están sujetos a derechos de terceros.

Todo lo cual declaro/declaramos bajo mi/nuestra absoluta responsabilidad, siendo plenamente consciente/conscientes de las consecuencias jurídicas (civiles, e incluso, penales) que su vulneración e incumplimiento puede acarrear para mi/nuestras persona/personas.

En ..... a.....de.....de.....

Firmado. El autor / Los autores

ANEXO II

MODELO DE RESUMEN EJECUTIVO DEL TRABAJO

NOMBRE Y APELLIDOS:
TÍTULO DEL ARTÍCULO:
TITULACIÓN ACADÉMICA, CATEGORÍA PROFESIONAL O PUESTO DE TRABAJO: **(Indicar cómo se desea figurar al publicar el artículo)** **Nº ORCID:**
BREVE CURRICULUM VITAE (Máximo 200 palabras):
NIF O DOCUMENTO EQUIVAENTE EN SU CASO:
DIRECCIÓN POSTAL:
TELÉFONO:
CORREO ELECTRÓNICO:
FIRMA:

ANEXO III

MODO DE CITAR LAS REFERENCIAS BIBLIOGRÁFICAS, JURISPRUDENCIALES Y OTRAS FUENTES

MONOGRAFÍA: APELLIDO APELLIDO, INICIAL DE NOMBRE., *Título de la obra (en cursiva)*, Editorial, ciudad, año, pág.

- Ejemplo: RODRÍGUEZ ORCAJO, J., *El sistema de compensación en el Derecho urbanístico español,* Reus, Madrid, 1995, pág. 176.

Repetición del mismo autor y obra:

- Ejemplo: RODRÍGUEZ ORCAJO, J., *El sistema de...,* op. cit., pág. 234.

Repetición inmediata del mismo autor y obra:

- Ejemplo: IBIDEM., pág. 417.

CAPÍTULO DE LIBRO: APELLIDO APELLIDO, INICIAL NOMBRE., "Título del capítulo del libro" (entrecomillado), en *Título del Libro (en cursiva),* Editorial, ciudad, año, pág.

- Ejemplo: NAVARRO ARENALES, R., "El suelo rústico hoy en día", en *Derecho urbanístico español,* Montecorvo, Madrid, 1987, págs. 426 y ss.

Repetición del mismo autor y obra:

- Ejemplo: NAVARRO ARENALES, R., "El suelo rústico...", op. cit., pág. 444.

Repetición inmediata del mismo autor y obra:

- Ejemplo: IBIDEM., pág. 462.

ARTÍCULO DE REVISTA: APELLIDO APELLIDO, INICIAL NOMBRE., "Título del artículo" (entrecomillado), *Título de la Revista (en cursiva),* núm., año, pág.

- Ejemplo: GARCÍA-MORENO RODRÍGUEZ, F., "Problemática jurídica de las Áreas de Transformación y de los suelos contaminados liberados por las mismas", *Revista de Derecho Urbanístico y Medio Ambiente,* núm. 216, 2005, págs. 151 y ss.

Repetición del mismo autor y obra:

- Ejemplo: GARCÍA-MORENO RODRÍGUEZ, F., "Problemática jurídica...", op. cit., pág. 160.

Repetición inmediata del mismo autor y obra:

- Ejemplo: IBIDEM., pág. 192.

JURISPRUDENCIA: Todas las Sentencias a que se aluda deberán indicar en nota a pie de página la referencia de la Base de Datos utilizada para poder ser localizada, en especial con referencia a la base de datos del CENDOJespecial a lbase de adtos del CENDOJ.localizadas.Especialmente las referencias ala base da datos del CENDOJ.

RECURSOS ELECTRÓNICOS: Todos los recursos electrónicos utilizados deberán indicar, además de su más completa referencia, cuándo fueron recuperados los mismos. Ejemplo: (Recuperado el 4 de Noviembre de 2016).

TABLAS Y GRÁFICOS: Todas las Tablas y Gráficos deberán indicar la autoría o procedencia de las mismas, o en su defecto, señalar que son de elaboración propia.

BIBLIOGRAFÍA FINAL: Todas las fuentes bibliográficas utilizadas en el correspondiente artículo deberán detallarse al final del mismo por orden alfabético de los apellidos de los autores.

NOTA: Puede ver las especificaciones en extenso en **www.rdu.es**

*De conformidad con la normativa vigente, le informamos que los datos que nos ha facilitado y los que nos facilite en el futuro por cualquier medio serán incorporados a fichero/s automatizados de RDU REVISTAS ESPECIALIZADAS,SL, con la finalidad de mantener relaciones con terceros. Usted podrá ejercer sus derechos de información,acceso, rectificación, cancelación y oposición dirigiendo un escrito a **rdu@rdu.es***

*Para acceder a nuestra política de privacidad **https: //www.rdu.es/contenidos/privacidad** deberán detallarse al final del mismo por orden alfabético de los apellidos de los autores.*

REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

SUMARIO DEL NÚMERO 34

CARTA DEL DIRECTOR

Pablo García Mexía (Director) J.D., Ph.D
www.PabloGMexia.net

Bienvenidos a la nueva edición de nuestra Revista de Privacidad y Derecho Digital, que esta vez nos presenta un más que especial número monográfico acerca del flamantemente publicado Reglamento europeo de inteligencia artificial. El número es sugerencia brillante de nuestro consejo asesor y tiene además el honor de encuadrarse en un Proyecto de investigación de la Universidad Autónoma de Madrid concretamente el que lleva por título "Nuevos avances en la legislación de transparencia en España: mejoras en la definición del marco regulatorio" (PID 2021-124724NB-100)", y del que es IP la profesora Ana de Marcos Fernández.

El número comienza con la Tribuna de Honor, a cargo de Carme Artigas, anterior Secretaria de Estado de Digitalización e Inteligencia Artificial del Gobierno de España y Co-Chair del UN AI High-level Advisory Board, ofreciéndonos un "detrás de cámaras" de la negociación del Reglamento europeo de IA – del que ella misma fue protagonista absolutamente directa –, que consiguió acordarse políticamente el pasado diciembre del 2023. Su artículo comenta también la gobernanza global inminente que acompaña desde foros multilaterales esta pionera aprobación europea.

Los estudios los encabeza la propia Ana María de Marcos Fernández, Profesora contratada doctora de Derecho Administrativo en la Universidad Autónoma de Madrid, quien nos contextualiza el novedoso Reglamento de Inteligencia Artificial. La autora arranca con el lanzamiento de ChatGPT, pero da un paso atrás para comentar los orígenes de la Inteligencia Artificial que data a la primera mitad del Siglo XX. La autora divide el trabajo en dos partes. La primera resume la evolución de la inteligencia artificial desde entonces, destacando hitos como la acuñación del término en los años 50, los algoritmos de aprendizaje automático, el aprendizaje profundo y los grandes modelos de lenguaje. La segunda parte aborda la regulación de la

inteligencia artificial por las instituciones europeas desde 2018, culminando en la aprobación del Reglamento de IA en 2024.

Ricardo Rivero Ortega, catedrático y ex rector de la Universidad de Salamanca, profundiza en las obligaciones informativas y técnicas que posibilitan el control de los proveedores de sistemas. Con la lectura de su trabajo, entenderemos con mayor detalle la evaluación de conformidad a la que están sujetas los proveedores, además de la actualización progresiva de las obligaciones de los proveedores y las posibles fuentes de litigiosidad. Además, antes de acabar con la aplicación de estas obligaciones a los poderes públicos, este autor hace una reflexión acerca del siempre difícil e incluso cuestionable equilibrio entre regulación e innovación.

El tercer estudio, realizado por Mercedes Fuertes López, Catedrática en la Universidad de León, versa sobre las obligaciones específicas que han de cumplir los usuarios de los sistemas de inteligencia artificial, atendiendo las diversas categorías de riesgos que presentan estos sistemas. Aquí reside de hecho una de las grandes singularidades del Reglamento europeo, en la medida en que, lejos de aplicarse exclusivamente a proveedores, también lo es a usuarios.

De la mano de Martín María Razquin Lizarraga, Catedrático en la Universidad Pública de Navarra, exploramos en profundidad el tema de sistemas de IA prohibidos, de alto riesgo, limitado riesgo, y bajo o nulo riesgo. El artículo analiza pues la regulación de la Inteligencia Artificial desde la perspectiva del riesgo, asimismo centrada en valores éticos que priorizan la protección de la salud, seguridad y derechos fundamentales.

El estudio escrito por María Jesús Jiménez Linares, Profesora Titular en la Universidad de Granada, nos anticipa los posibles riesgos que envuelven el desarrollo y auge de la Inteligencia Artificial generativa. La autora nos hace conscientes de los diferentes riesgos de la IA, como pueden ser la confusión, desinformación, amenazas de privacidad o ciberdelincuencia, entre otros muchos que explica a lo largo de su estudio. Más aún, hace especial referencia a la preocupación, a escala europea, por los “deep fakes”, y la exigencia del Reglamento de una necesaria transparencia sobre el origen artificial de este contenido.

El apartado En Prospectiva nos aleja de los riesgos tan sonados en boca de todos sobre el avance tecnológico, y nos muestra una faceta positiva del Reglamento de inteligencia artificial europeo: la innovación. Este apartado recoge los aspectos del Reglamento que fomentan la innovación, e indiscutiblemente el crecimiento económico. La norma hace especial referencia al apoyo que deben recibir las Pymes y Startups, hace una llamada para el establecimiento de "sandboxes" regulatorios, avala la investigación y desarrollo, destacando la estandarización como elemento crucial. Este trabajo ahonda a sí mismo en la innovación como objetivo pretendidamente capital del Reglamento, junto a los mucho más notorios de garantía de la salud, seguridad y derechos y libertades de las personas, analizando de hecho hasta qué punto esta norma consigue tal objetivo.

Esperamos que disfruten de este número, y también que resulte beneficioso con vistas a ayudar a afrontar los retos que trae consigo la entrada en vigor del nuevo Reglamento de IA.

Atentamente,

Pablo García Mexía

DEL REGLAMENTO EUROPEO DE LA IA HACIA LA NECESARIA GOBERNANZA GLOBAL

FROM THE EUROPEAN AI REGULATION TO THE NECESSARY GLOBAL GOVERNANCE

Por CARME ARTIGAS

ECo-Chair del UN AI High-level Advisory Board, Senior Fellow de la Harvard Kennedy School - Belfer Center, y exSecretaria de Estado de Digitalización e Inteligencia Artificial del Gobierno de España

REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

DEL REGLAMENTO EUROPEO DE LA IA HACIA LA NECESARIA GOBERNANZA GLOBAL

Ante la reciente entrada en vigor desde el pasado 1 de agosto del Reglamento Europeo de Inteligencia Artificial (EU AI ACT), posterior a su publicación en el Diario Oficial de la Unión Europea el pasasdo12 de julio de 2024 , vale la pena hacer una reflexión de cómo hemos llegado hasta aquí y del impacto de esta ley a nivel global, incluso antes de su entrada en vigor. Como sabemos, la presidencia Española del Consejo de la UE -contra todo pronóstico- consiguió el pasado diciembre del 2023 el acuerdo final sobre el primer Reglamento internacional de Inteligencia Artificial, convirtiendo a la Unión Europea en la primera región del mundo en legislar los usos de la IA, sus límites, la protección de los derechos fundamentales de los ciudadanos y la participación en su gobernanza, garantizando a la vez la competitividad de nuestras empresas.

Llegar hasta aquí no fue fruto de la improvisación ni de la oportunidad.

Ya en el 2020 la Comisión Europea crea, por primera vez, un grupo de expertos dedicado exclusivamente a la "ética de la Inteligencia Artificial", influenciado por los "Principios éticos sobre Inteligencia Artificial" que la OCDE había publicado tan solo unos meses antes.

En paralelo, en España, se creaba en enero del 2020, la primera Secretaría de Estado de Digitalización e Inteligencia Artificial, que asumí el reto de dirigir, dependiendo del Ministerio de **Asuntos Económicos** y Transformación Digital. Toda una carta de intenciones. El potencial de lo que era una tecnología emergente en ese momento nos hacía intuir su importancia en la economía y en la sociedad. No nos equivocamos al elegir el nombre, ni tampoco en conseguir que uno de sus primeros hitos fuera la **Carta de Derechos Digitales** presentada por el

propio Presidente del Gobierno en Diciembre de ese mismo año. Hicimos bandera de algo que ahora parece que todo el mundo comparte: que el desarrollo de la tecnología no se puede llevar por delante derechos y garantías democráticas que tanto nos ha costado conseguir. Acuñamos desde el gobierno de España el concepto de "Humanismo Tecnológico" y empezamos a impulsar esta visión por todo el mundo.

El grupo de expertos de la Comisión trabajó durante meses en la profundidad de esta tecnología y su convivencia con la ética tal y como la conocemos. En 2021 se presenta la primera propuesta del Reglamento Europeo de Inteligencia Artificial, ya bautizado como ***AI Act.***

Seamos honestos. Nadie nos siguió. Es más, tanto la industria como el resto de países nos miraba por encima del hombro: "ya están los europeos hiper-regulando y matando la innovación". Pero desde Europa teníamos claro que ese era el camino. Que no nos gustaba un modelo de desarrollo tecnológico con los datos e información en manos privadas y concentrado en un puñado de grandes empresas tecnológicas. Ni tampoco un modelo en el que los datos y la inteligencia artificial estuvieran bajo el poder de los gobiernos, con capacidad de imponer un modelo de hipervigilancia y control social. Europa, por tanto, se convertía en esa "tercera vía" que hacía compatible el desarrollo tecnológico con la protección de los derechos fundamentales para sus ciudadanos y empresas.

Y entonces, como en los mejores giros de guion, pasa algo inesperado que nos da la razón. El 30 de noviembre de 2022 llega Chat GPT y la Inteligencia Artificial pasa de ser un asunto exclusivo de los expertos en tecnología y se cuela en las noticias, en las redes sociales, en las tertulias de radio y televisión y hasta en las sobremesas.

La IA y todo lo que puede llegar a hacer por sí misma sin supervisión humana, sus riesgos futuros y presentes, ya es

Revista de Privacidad y Derecho Digital
ISSN digital 2444-5762, NÚM. 34, Madrid, Mayo-Agosto (2024), págs. 20-25

una preocupación que el mundo comparte. Y la "tercera vía" de Europa deja de ser una alternativa y se convierte en "**la única vía"**: regulación, responsabilidad y sostenibilidad.

Y empieza la Presidencia española del Consejo de la UE y con ello nuestro liderazgo firme e inequívoco en la negociación del Reglamento, con cuatro *trílogos* entre junio y diciembre antes del trílogo final y decenas de reuniones técnicas, buscando el consenso entre las posiciones de los Estados Miembro y del Parlamento Europeo. Conscientes de que, si no se consigue este acuerdo durante nuestra Presidencia, probablemente no se consiga nunca, perdiendo por tanto una oportunidad histórica.

Ante la presión de que la Unión Europea imponga su modelo, la comunidad internacional se movilizó poniendo el foco en códigos de conducta voluntarios, mejores prácticas y la autoregulación, centrándose en los riesgos existenciales a largo plazo, pero olvidándose de los que ya están aquí (desinformación, discriminación, manipulación, vigilancia, *deep fakes*...). El código de conducta de la IA generativa del G7 en Hiroshima, la *Executive Order* de la Administración Biden o la *Declaración de Bletchley Park*, todas ellas anunciadas en octubre del 2023, son grandes iniciativas y avances totalmente complementarios, pero insuficientes. Porque lo que hace único al Reglamento europeo es que por primera vez le estamos diciendo al mundo lo que los ciudadanos europeos no aceptamos que haga la Inteligencia Artificial, aunque sea técnicamente posible, estableciendo prohibiciones y salvaguardas, y exigiendo transparencia y control en el caso de sus usos de alto riesgo.

El reglamento europeo de IA es, por tanto, no solo un estándar legal, ni tan solo un estándar técnico. Es un **estándar moral**.

El quinto y definitivo trílogo tuvo lugar entre el 6 de diciembre y el 8 de diciembre del 2023. La negociación duró 38 horas, las primeras 27 horas de manera ininterrumpida y sin dormir, pasando sin duda a formar parte del récord Guinness como

el trílogo más largo de la historia de la Unión Europea. Este esfuerzo maratoniano nos da una idea de la dimensión de lo que estaba en juego. Cuando entré en esa sala tenía clara la responsabilidad que teníamos los co-legisladores sobre nuestras espaldas. Habíamos llegado hasta allí soportando, a partes iguales, la presión de la opinión pública, de los *lobbys* de la industria, de los que querían sobreproteger limitando la innovación y de los que no querían ser regulados para seguir desarrollando una IA sin cortapisas. Todos bien intencionados, sin ninguna duda. Pero mantuvimos la cabeza fría, convencidos de que se puede regular sin matar la innovación y sin perder derechos ni garantías. Y con un mecanismo de actualización de la propia ley de manera dinámica para que pueda superar la prueba del tiempo a medida que avance la tecnología, a través de actos de ejecución y actos delegados.

Europa ya ha cambiado el rumbo del desarrollo tecnológico para las próximas generaciones. Un modelo de referencia que allana el camino hacia una futura gobernanza global, como ya estamos abordando en el seno de Naciones Unidas y su Consejo Asesor de IA, que me honra co-presidir. En nuestras recomendaciones, que presentaremos durante la 79 Asamblea General de Naciones Unidas a mediados de septiembre, abordamos el déficit de Gobernanza global de la IA en tanto a la inclusión del sur global, la falta de transparencia consenso científico, la necesidad de capacitación y acceso a datos y computación de manera igualitaria, los necesarios mecanismos de monitorización y coordinación de incidencias a nivel global y la necesidad de tener estándares técnicos y legislativos interoperables. Todo ello regido bajo la Carta de las Naciones Unidas, el derecho internacional y los inalienables derechos humanos.

Conseguir el consenso internacional a nuestras recomendaciones tampoco está siendo fácil. La tecnología cambia el equilibrio de poder y ya no es indistinguible de la geopolítica. Y precisamente por ello no existe otra plataforma multilateral como Naciones

Unidas con la legitimidad necesaria para conseguir estos consensos, a pesar de que su autoridad solo llega hasta donde quieran sus Estados Miembros.

Europa ha marcado la dirección correcta y está solo en nuestras manos demostrar que el desarrollo tecnológico es compatible con los valores democráticos, los derechos y garantías y el progreso social. Y ese modelo me inspira para seguir trabajando hacia algo que parecería obvio: que la Inteligencia Artificial se desarrolle para el bien de la humanidad. De ello depende nuestro futuro.

UNA DOBLE HISTORIA DE LA INTELIGENCIA ARTIFICIAL: AVANCE TECNOLÓGICO Y PROCESO DE REGULACIÓN EN EUROPA[1] (*)

A DOUBLE HISTORY OF ARTIFICIAL INTELLIGENCE: TECHNOLOGICAL ADVANCE AND REGULATION PROCESS IN EUROPE

Por ANA DE MARCOS FERNÁNDEZ

Profesora de Derecho Administrativo. Universidad Autónoma de Madrid

(*) Este trabajo se recibió el 11 de julio de 2024 y fue aceptado el 5 de septiembre.

1 Este trabajo forma parte de la investigación desarrollada por el Proyecto de Investigación PID2021-124724NB-I00, titulado NUEVOS AVANCES EN LA LEGISLACION DE TRANSPARENCIA EN ESPAÑA: MEJORAS EN LA DEFINICION DEL MARCO REGULATORIO. Ana de Marcos es la Investigadora Principal de este Proyecto, que lleva a cabo el Grupo reconocido de investigación de la Universidad Autónoma de Madrid, LAS GARANTÍAS DE LOS CIUDADANOS EN LA ACTUACIÓN DEL GOBIERNO Y LA ADMINISTRACIÓN.

REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

RESUMEN

La gran convulsión que supuso el lanzamiento, en noviembre de 2022, de ChatGPT volvió a poner en el primer plano de la actualidad y del interés social tanto el concepto de inteligencia artificial como las implicaciones que el desarrollo de la misma tiene en todos los órdenes de la vida. Pero la presentación del chatbot conversacional de lenguaje natural de la compañía OpenAI, a pesar de su importancia e impacto, no fue más que un nuevo paso en la evolución de una disciplina, la inteligencia artificial, cuyos orígenes científicos y tecnológicos se encuentran en la primera mitad del siglo XX.

La primera parte de este artículo resume la evolución de la inteligencia artificial a partir de entonces y hace mención de los principales hitos de la misma, que han ido desde los primeros y sencillos modelos de robots hasta los asistentes avanzados de la actualidad, pasando por la propia acuñación del concepto *inteligencia artificial* en los años 50, los algoritmos de aprendizaje automático, el aprendizaje profundo, los transformadores y los grandes modelos de lenguaje, entre otros avances fundamentales.

La segunda parte del artículo se refiere al proceso de elaboración de la regulación de la inteligencia artificial por parte de las instituciones europeas, a partir de 2018, que culmina con la aprobación del Reglamento de la IA en la Eurocámara el 13 de marzo de 2024, y por el Consejo el 21 de mayo de 2024. Su publicación en el Diario Oficial de la Unión Europea se produce el 12 de julio de 2024, si bien su entrada en vigor tendrá lugar de forma escalonada según las distintas materias que son objeto de su regulación.

La regulación europea de la IA se centra en la excelencia y la confianza, con el objetivo de impulsar la investigación y la capacidad industrial, garantizando al mismo tiempo la seguridad y los derechos fundamentales. Pretende conseguir el liderazgo

mundial estratégico de Europa en relación con una IA fiable, y centrada en el ser humano, que permita la innovación tecnológica y garantice la vigencia del Estado de Derecho y la democracia.

PALABRAS CLAVE: *Inteligencia artificial (IA), aprendizaje automático, sistema experto, grandes modelos de lenguaje, Estrategia europea de inteligencia artificial, Ley de inteligencia artificial, Reglamento europeo de inteligencia artificial, Responsabilidad derivada de inteligencia artificial, Oficina europea de inteligencia artificial.*

ABSTRACT

The upheaval caused by the launch, in November 2022, of ChatGPT brought back to the forefront of current affairs and social interest both the concept of artificial intelligence and the implications of its development for all aspects of life. But despite its importance and impact, the introduction of OpenAI's natural language conversational chatbot was just a new step in the evolution of a discipline - artificial intelligence - whose scientific and technological roots are to be found in the first half of the 20th century.

The article contains two main parts, and the first one summarizes the evolution of artificial intelligence since then, and highlights its main milestones, ranging from the first simple robot models to current advanced assistants, with the very emergence of the artificial intelligence concept in the 1950s, machine learning algorithms, deep learning, transformers and large language models, among other fundamental steps, coming in between.

The second part deals with the process of elaboration of the regulation of artificial intelligence by the European institutions, starting in 2018 and culminating in the adoption of the AI Regulation in the European Parliament on March 13, 2024, and by the Council on May 21, 2024. The Regulation was published in the Official Journal of the European Union on July 12, 2024, although it will come into force on a step-by-step basis according to the different matters that are the subject of its regulation.

The European regulation of AI focuses on excellence and trust, with the aim of boosting research and industrial capacity, while ensuring safety and fundamental rights. It aims to achieve Europe's strategic global leadership in trusted, human-centered AI, at the same time enabling technological innovation and ensuring the rule of law and democracy.

KEY WORDS: *Artificial intelligence (AI) machine learning, expert system, large language models, European Artificial Intelligence Strategy, Artificial Intelligence Act, European Artificial Intelligence Regulation, Liability arising from Artificial Intelligence, European Office for Artificial Intelligence.*

SUMARIO

PRIMERA PARTE: LA CONSTRUCCIÓN DE LA IA. EL AVANCE TECNOLÓGICO DE LA IA

I.- UNA OLA DE INTERÉS CRECIENTE

El 30 de noviembre de 2022, la compañía estadounidense OpenAI presenta su chatbot ChatGPT y desata la actual, inmensa y creciente ola de interés por la inteligencia artificial, que se manifiesta en múltiples niveles: tecnológico, personal, social, económico-empresarial, laboral, informativo, político y legal

Las funcionalidades de uso del lenguaje natural de ChatGPT, sus habilidades conversacionales, su usabilidad y su capacidad para responder preguntas y adaptarse a sugerencias sobre el tipo de respuesta requerido cautivaron a los usuarios desde el primer momento. Ese interés se hizo viral y así, cinco días después de su lanzamiento, ChatGPT ya contaba con más de un millón de usuarios, cifra que, por ponerla en perspectiva, TikTok tardó nueve meses en superar e Instagram, dos años y medio. En enero de 2023, ChatGPT alcanzó los 100 millones de usuarios, convirtiéndose en la aplicación de software de consumo que más rápido ha llegado a esa cifra.

Pero, más allá de estos datos, de sus capacidades y de la serie de desarrollos similares por parte de otras compañías que ha provocado, el interés de ChatGPT en el gran esquema del desarrollo de la inteligencia artificial es que ha puesto esta disciplina de las ciencias informáticas en el centro del interés de la sociedad y está haciendo que en la misma crezcan, simultáneamente, la emoción por las múltiples posibilidades de desarrollo que plantea y la aprensión, e incluso el temor, por las amenazas de diversa índole que supone.

La notoriedad que en el último año y medio han alcanzado aplicaciones y desarrollos como ChatGPT y otros similares, y la

enorme curiosidad que han generado en torno a la disciplina de la inteligencia artificial no son, en cualquier caso, manifestaciones de un fenómeno científico y tecnológico nuevo, sino un capítulo más, aunque seguramente fundamental y determinante en muchos sentidos, de la historia de una disciplina que viene interesando a los estudiosos desde hace casi un siglo; y que hoy forma parte de la vida cotidiana de millones de personas a través de dispositivos y aplicaciones de uso tan común como, por ejemplo, los asistentes de voz, las sugerencias de consumo de plataformas de contenido o de compras en sitios de comercio electrónico, la generación automática de direcciones y la detección de spam en las aplicaciones de e-mail, la elección del contenido que se muestra al usuario en sus perfiles de redes sociales o la determinación de las prioridades de limpieza de las aspiradoras robóticas.

II.- DEFINICIONES DE INTELIGENCIA ARTIFICIAL

La inteligencia artificial es, según la definición que de ella da el Diccionario de la Real Academia, la "disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico".

La Encyclopedia Britannica elabora algo más el concepto y la define así: "La capacidad de un ordenador o de un robot controlado por un ordenador para llevar a cabo tareas que comúnmente se asocian con los seres inteligentes. El termino se aplica con frecuencia a proyectos de desarrollo de sistemas dotados de procesos intelectuales característicos de los humanos, tales como la capacidad de razonar, descubrir significados, hacer generalizaciones y aprender de la experiencia acumulada". A grandes rasgos la inteligencia artificial puede ser de dos tipos: la

predictiva, que a partir de algoritmos y aprendizaje automático[2], identifica patrones y hace deducciones mediante el uso datos históricos y actuales; y la generativa, que, mediante el aprendizaje profundo[3] genera contenido a partir de los datos con los que está entrenada.

La UE define también qué es la inteligencia artificial. La primera definición que dio el Grupo de Expertos es la siguiente;

> *El término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción –con cierto grado de autonomía– con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas). Estamos utilizando la IA diariamente, por ejemplo, para traducir de un idioma a otro, generar subtítulos en los vídeos o bloquear el correo electrónico no solicitado)[4].*

2 El aprendizaje automático es un subconjunto de inteligencia artificial que permite que un sistema aprenda y mejore de forma autónoma mediante redes neuronales y aprendizaje profundo, sin tener que ser programado explícitamente, a través de la ingesta de grandes cantidades de datos.
Debido a que el aprendizaje automático permite que los sistemas informáticos se ajusten y mejoren a sí mismos de forma continua a medida que acumulan más "experiencias", mientras más datos se pongan en ellos, más precisos serán los resultados. (https://cloud.google.com/learn/what-is-machine-learning?hl=es-419)

3 El aprendizaje profundo es un tipo de aprendizaje automático que usa redes neuronales artificiales para permitir que los sistemas digitales aprendan y tomen decisiones basadas en datos no estructurados y sin etiquetar. En general, el aprendizaje automático entrena sistemas de inteligencia artificial para aprender de experiencias adquiridas con datos, reconocer patrones, hacer recomendaciones y adaptarse. Con el aprendizaje profundo en particular, en lugar de simplemente responder a conjuntos de reglas, los sistemas digitales generan conocimiento a partir de ejemplos y, después, usan ese conocimiento para reaccionar, comportarse y actuar como personas. (https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-deep-learning)

4 Estrategia de la UE Inteligencia artificial para Europa, 24 de abril de 2018,https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN

El mismo Grupo de Expertos perfeccionó y actualizó con posterioridad esta definición:

> *«Los sistemas de inteligencia artificial (IA) son sistemas de software (y en algunos casos también de hardware) diseñados por seres humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante la percepción de su entorno a través de la obtención de datos, la interpretación de los datos estructurados o no estructurados que recopilan, el razonamiento sobre el conocimiento o el procesamiento de la información derivados de esos datos, y decidiendo la acción o acciones óptimas que deben llevar a cabo para lograr el objetivo establecido. Los sistemas de IA pueden utilizar normas simbólicas o aprender un modelo numérico; también pueden adaptar su conducta mediante el análisis del modo en que el entorno se ve afectado por sus acciones anteriores.*
>
> *La IA es una disciplina científica que incluye varios enfoques y técnicas, como el aprendizaje automático (del que el aprendizaje profundo y el aprendizaje por refuerzo constituyen algunos ejemplos), el razonamiento automático (que incluye la planificación, programación, representación y razonamiento de conocimientos, búsqueda y optimización) y la robótica (que incluye el control, la percepción, sensores y accionadores así como la integración de todas las demás técnicas en sistemas ciberfísicos)»*,[5]
>
> *IA específica (o débil) y general (o fuerte). Un sistema de IA general es un sistema diseñado para realizar la mayoría de las actividades que pueden llevar a cabo los seres humanos. Por el contrario, los sistemas de IA específicos solamente pueden realizar una tarea concreta o un número reducido de ellas. Los sistemas de IA desplegados actualmente constituyen ejemplos*

5 Grupo de expertos de alto nivel sobre inteligencia artificial. Una definición de la inteligencia artificial: Principales capacidades y disciplinas científicas (https://ec.europa.eu/digital-single-maret/en/high-level-expert-group-artificial-intelligence). 5 de julio 2024.

de IA específica. En las etapas iniciales de la inteligencia artificial, los investigadores utilizaban una terminología diferente (IA «débil» y «fuerte»). Hoy en día continúan existiendo numerosos desafíos éticos, científicos y tecnológicos sin resolver para desarrollar las capacidades que serán necesarias para construir sistemas de IA generales, como el razonamiento basado en el sentido común, la autoconciencia o la capacidad de la máquina para definir su propio objetivo[6].

III.- APUNTES SOBRE LA HISTORIA DE LA INTELIGENCIA ARTIFICIAL

III.1- LOS ROBOTS LAS PRIMERAS APROXIMACIONES CIENTÍFICAS

La historia del concepto de inteligencia artificial y del desarrollo del mismo encuentra su precedente, o sus capítulos iniciales, en la primera mitad del siglo XX. Fue entonces cuando se empezó a concebir la idea de la creación de un cerebro artificial y se construyeron los primeros modelos de lo que ahora llamamos robots, ingenios que reproducían acciones humanas muy básicas y que estaban inspirados por obras de ciencia ficción. La palabra robot se usó por primera en la obra de teatro de ciencia ficción *Robots Universales Rossum,* del escritor checo Karel Čapek. Estrenada en 1921, su argumento gira en torno a una fábrica de seres humanos artificiales, los robots, creados para ayudar en el trabajo a las personas. El robot que suplanta al personaje de Maria en la famosa película *Metrópolis* es otra conocida creación de este tipo.

Estos ejemplos son síntomas de un interés por las máquinas inteligentes que a finales de los años 40 cristalizó en la aparición de una generación de científicos, matemáticos y filósofos que habían asimilado la idea de una inteligencia artificial y empezaron a

6 *Ibidem.*

Revista de Privacidad y Derecho Digital
ISSN digital 2444-5762, NÚM. 34, Madrid, Mayo-Agosto (2024), págs. 26-89

pensar sobre ella y a estudiarla. En 1949, el científico informático estadounidense Edmund C. Berkley publicó el libro *Cerebros gigantes o máquinas que piensan*, en el que describió por primera vez los ordenadores para el gran público y asoció con estas máquinas el término "cerebro".

La década de los 50 fue fundamental para el desarrollo de la inteligencia artificial y de hecho fue en ella en la que se acuñó este concepto. Un primer y decisivo paso fue la publicación por parte del matemático británico Alan Turing de su artículo *Maquinaria computacional e inteligencia*[7], en el que hablaba de cómo desarrollar máquinas inteligentes al tiempo que proponía un test para cuantificar esa inteligencia. La idea central de Turing era que, si los seres humanos usan la información disponible y su capacidad de raciocinio para resolver problemas y tomar decisiones, ¿por qué no podrían las máquinas hacer lo mismo?

Las ideas de Turing se encontraban con los obstáculos de la limitada capacidad de los ordenadores de aquel momento para almacenar información y de lo caro que resultaba su uso, por lo que solo las universidades más prestigiosas y las grandes compañías tecnológicas podían permitirse destinar fondos a la investigación y el desarrollo de un campo tan especulativo como las maquinas inteligentes.

III.2- LA CONFERENCIA DE DARTMOUTH Y EL DESARROLLO EN LOS AÑOS 60

En 1955, Allen Newell, Cliff Shaw y Herbert Simon presentaron Logic Theorist, un programa diseñado para imitar las capacidades humanas de resolución de problemas que contó con la financiación de la Research and Development Corporation. Considerado el primer programa de inteligencia artificial, se presentó en la Darmouth Summer Research Project on Artificial

7 A. M. Turing: "Computing Machinery and intelligence" (https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf)

Intelligence (Conferencia de Darmouth)[8], evento promovido por John McCarthy y Marvin Minsky que se celebró en 1956 en el centro universitario Darmouth College (New Hampshire). Esta conferencia, que fue donde la expresión "inteligencia artificial" se usó por vez primera, está considerada un evento seminal en el desarrollo de la disciplina y fue el catalizador de los siguientes veinte años de investigación sobre la misma.

La década de los 60 fue testigo de un gran desarrollo científico y tecnológico de la inteligencia artificial, al tiempo que el concepto se implantaba definitivamente en la cultura popular. Los ordenadores aumentaron mucho su capacidad de almacenamiento y procesamiento de datos y su coste se abarató. Los algoritmos del aprendizaje automático mejoraron, al igual que lo hizo el conocimiento sobre qué algoritmos aplicar para la resolución de un problema determinado. Algunos hitos de esta época fueron:

- la creación en 1958, por parte de Joseph McCarthy, de List Processing (LISP), el primer lenguaje de programación para la investigación de la inteligencia artificial.
- la acuñación, en 1959, de la expresión "machine learning" ("aprendizaje automático") por parte de Arthur Samuel, a partir de sus trabajos para adiestrar un ordenador en el juego del ajedrez.
- la creación del primer sistema experto[9] -una forma de inteligencia artificial que reproduce los sistemas de toma de

8 "Dartmouth Summer Research Project: The Birth of Artificial Intelligence" (https://www.historyofdatascience.com/dartmouth-summer-research-project-the-birth-of-artificial-intelligence/). "A Look Back on the Dartmouth Summer Research Project on Artificial Intelligence" (https://www.thedartmouth.com/article/2023/05/a-look-back-on-the-dartmouth-summer-research-project-on-artificial-intelligence)

9 Los sistemas expertos son programas informáticos que tienen el objetivo de solucionar un problema concreto y utilizan la Inteligencia Artificial (IA) para simular el razonamiento de un ser humano. Se denominan sistemas expertos porque estos programas imitan la toma de decisiones de un profesional en la materia. Actualmente, se consideran dentro del global de la Inteligencia Artificial. Se crearon durante la década de los 60 (aunque alcanzaron su mayor popularidad en los años posteriores) y fueron uno de los primeros sistemas de Inteligencia Artificial utilizados con éxito. (https://www.unir.net/ingenieria/revista/sistema-experto/)

decisiones de los expertos humanos- que tuvo lugar en 1965 de la mano de Edward Feigenbaum y Joshua Lederberg.

- en 1966 Joseph Weizenbaum desarrolla ELIZA, el primer chatbot que usaba procesamiento de lenguaje natural para conversar son seres humanos.
- la publicación por parte del matemático ruso Alexey Ivakhenko, en 1968, de un artículo en el que proponía un nuevo enfoque de la inteligencia artificial que posteriormente dio lugar a lo que actualmente se conoce como aprendizaje profundo.
- el gobierno estadounidense destinó financiación a varias instituciones que trabajaban en inteligencia artificial y se mostró particularmente interesado en el desarrollo de programas que pudieran transcribir y traducir lenguaje hablado, así como en la alta capacidad de procesado de datos.

En 1970, Marvin Minsky decía en la revista *Life* que "en el plazo de tres a ocho años dispondremos de una máquina con la inteligencia general de un ser humano medio". Pero este optimismo no se vio confirmado. De nuevo, las limitaciones en la capacidad de almacenaje y procesamiento de datos de los ordenadores supusieron un freno al desarrollo de la inteligencia artificial y con la falta de avances llegó también un recorte en la financiación de los proyectos. En cualquier caso, en el año 1979 se funda la American Association of Artificial Intelligence, actualmente conocida como Association for the Advancement of Artificial Intelligence (AAAI).

III.3- APRENDIZAJE PROFUNDO, SISTEMAS EXPERTOS Y DEEP BLUE

Los años 80 supusieron un cambio de tendencia y el interés por la inteligencia y la financiación de proyectos volvieron a activarse, de la mano del desarrollo y popularizacion del aprendizaje profundo y de los sistemas expertos, procesos que habían dado

sus primeros pasos años antes. Algunos hechos reseñables de estos años son:

- Expert Configurer (XCON) se convierte, en 1980, en el primer sistema experto que llega al mercado.
- El Gobierno japonés pone en marcha en 1982 el Fifth Generation Computer Project[10], un ambicioso programa de desarrollo informático en varios campos. Entre ellos se encontraba la inteligencia artificial y, en concreto, la creación de ordenadores que pudieran procesar el lenguaje humano y conversar, así como expresar razonamientos. El proyecto se mantuvo en vigor durante doce años y, aunque no alcanzó sus objetivos, inspiró a muchos ingenieros y científicos en el desarrollo de trabajos de inteligencia artificial.
- En 1984 se presenta en una conferencia de la AAAI un programa de dibujo autónomo llamado AARON.
- un equipo de la universidad Bundeswher München construye en 1986 el primer coche capaz de circular sin conductor y lleva a cabo pruebas en vías sin obstáculos[11].
- Se lanza al mercado Alacrity, primer programa de asesoría en estrategia de gestión; estaba basado en un sistema experto con más de 3.000 instrucciones.

El final de los 80 y principios de los 90 trajeron, en la naturaleza cíclica del desarrollo de la inteligencia artificial, otra época de desinterés. El lento avance de la disciplina contrastaba con los grandes presupuestos que requería y eso frenó la financiación, lo que, al igual que en otros momentos, trajo como consecuencia un menor avance.

10 "'Fifth Generation' Became Japan's Lost Generation" (https://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html). Edward Feigenbaum y Howard Shrobe: "Tje Japanese National Fifth Generation Project: Introduction, survey and evaluation" (https://stacks.stanford.edu/file/druid:kv359wz9060/kv359wz9060.pdf)

11 "The man who invented the self-driving car (in 1986)" (https://www.politico.eu/article/delf-driving-car-born-1986-ernst-dickmanns-mercedes/)

La financiación pública para proyectos de inteligencia artificial en Estados Unidos, principal núcleo de desarrollo de la disciplina, no se recuperó durante los años 90, pero ello no fue óbice para que en la segunda parte de la misma se produjeran algunos avances significativos, entre los que destaca, también por la alta repercusión pública que tuvo, el hecho de que Deep Blue, un superordenador desarrollado por la compañía IBM, derrotara en una partida al vigente campeón mundial de ajedrez, Gary Kasparov[12], lo que supuso un gran avance en el desarrollo de la inteligencia artificial creada para la toma de decisiones. También el campo del reconocimiento e interpretación del lenguaje experimentó, el mismo año, un avance significativo: Windows lanzó un software de reconocimiento de voz desarrollado por la compañía Dragon Systems; y en 2000, la científica y emprendedora estadounidense Cynthia Breazeal presentó Kismet[13], un robot que podía reconocer y reproducir emociones.

III.4- "BIG DATA", REDES SOCIALES Y ASISTENTES DOMÉSTICOS

Los primeros diez años de este siglo registraron nuevos progresos y fueron testigos del desembarco de la IA, a través de diferentes aplicaciones y dispositivos, en la vida cotidiana. Este desarrollo se debió en parte a que la capacidad de almacenamiento y procesamiento de los ordenadores se puso a la altura de nuestras necesidades y en muchos casos la ha sobrepasado. Ello hizo posible la llegada de la era del big data o tratamiento a través de aprendizaje automático de grandes cantidades de datos estructurados y semiestructurados para la destilación de información relevante que ayude en la toma de decisiones. Algunos hitos del desarrollo de la inteligencia artificial en estos años son:

12 "Superordenadores: Deep Blue, la máquina que derrotó a Gary Kasparov" (https://blog.caixabank.es/blogcaixabank/superordenadores-deep-blue-la-maquina-que-derroto-a-gary-kasparov/#)

13 "Robots: Kismet" (https://robotsguide.com/robots/kismet)

- Se lanza la primera aspiradora Roomba en 2002.
- La NASA explora la superficie de Marte con Spirit y Opportunity, dos vehículos que realizan su tarea sin intervención humana.
- La inteligencia artificial se convierte en ingrediente esencial en el funcionamiento y experiencia de usuario que proporcionan las redes sociales, un fenómeno que inicia su explosión en este periodo con la creación de Facebook (2004) y Twitter, ahora denominada X (2006).
- Microsoft lanza Xbox 360 Kinect en 2010. Se trata del primer dispositivo para videojuegos que reconoce el movimiento del cuerpo humano y lo convierte en instrucciones para jugar.
- En 2011, Apple lanza Siri, su primer asistente virtual, e IBM presenta Watson, el sistema basado en inteligencia artificial capaz de responder a preguntas formuladas en lenguaje natural que vence a dos ganadores del concurso Jeopardy! en un show televisado[14].

III.5- UNOS AÑOS DE AVANCES ESPECTACULARES: LOS GRANDES MODELOS DE LENGUAJE

El desarrollo del aprendizaje profundo y de sistemas de reconocimiento y generación de imágenes y lenguaje puede ser considerada la tendencia protagonista en la evolución de la inteligencia artificial durante los últimos años. Los avances han sido, y siguen siendo, tan espectaculares que a lo largo de los años se han alzado en ocasiones voces, algunas de ellas de grandes expertos en la materia, alertando sobre los peligros de un desarrollo incontrolado de esta tecnología. Los siguientes son algunos de los momentos destacables vividos por la inteligencia artificial en los tiempos recientes:

14 "Watson, 'Jeopardy!' champion" (https://www.ibm.com/history/watson-jeopardy)

- Geoffrey Hinton, Ilya Sutskever y Alex Krizhesky presentan en 2012 una arquitectura de red neuronal convolucional (CNN, por sus siglas en inglés)[15] que gana el reto ImageNet y da un gran impulso a la investigación y la implementación del aprendizaje profundo.
- La compañía DeepMind presenta en 2013 el aprendizaje reforzado profundo, una CNN que aprendía a partir de estímulos y a la que se enseñó a practicar diversos juegos mediante repetición, alcanzando en ellos una destreza superior a la de humanos expertos.
- El mismo año, un grupo de investigadores de Google presenta Word2vec, una técnica para el procesamiento de lenguaje natural que identifica relaciones semánticas entre las palabras.
- Igualmente en 2014, el informático Ian Goodfellow y su equipo crean las redes generativas adversarias, modelos generativos de aprendizaje automático que crean instancias nuevas de datos que se asemejan a los datos de entrenamiento y permiten que una máquina pueda crear imágenes, textos y sonidos que en apariencia no existían.
- Facebook desarrolla, también en 2014, el sistema de reconocimiento facial DeepFace, basado en aprendizaje profundo y que permite identificar rostros de personas en imágenes digitales con casi la misma precisión que los humanos.

15 Las redes neuronales convolucionales se distinguen de otras redes neuronales por su rendimiento superior con entradas de imagen, voz o señales de audio. Se componen de tres tipos principales de capas: capa convolucional, capa de agrupación y capa totalmente conectada.
La capa convolucional es la primera capa de una red convolucional. Si bien las capas convolucionales pueden ir seguidas de otras capas convolucionales o de capas de agrupación, la capa final es la capa totalmente conectada. Con cada capa, la CNN aumenta en complejidad, identificando partes cada vez más grandes de la imagen. Las primeras capas se centran en características simples, como colores y bordes. A medida que los datos de la imagen avanzan a través de las capas, la CNN comienza a reconocer elementos o formas más grandes hasta que finalmente identifica el objeto esperado. (https://www.ibm.com/es-es/topics/convolutional-neural-networks)

- En 2016, AlphaGo, programa desarrollado por DeepMind, derrota a destacados jugadores de Go, un logro que se compara al de Deep Blue con Kasparov casi veinte años antes.
- Seis investigadores de Google y otros dos expertos publican en 2017 el fundamental artículo "Atención es todo lo que necesitas"[16], en el que desarrollan una nueva arquitectura de aprendizaje profundo conocida como transformadores, los cuales inspiraron la investigación en herramientas que podían analizar texto sin etiquetar en los grandes modelos de lenguaje.
- También en 2017, el destacado científico Stephen Hawking declara: "A menos que aprendamos a prepararnos para y a evitar los riesgos potenciales, la inteligencia artificial puede ser el peor acontecimiento de la historia de nuestra civilización".
- En 2018, la compañía OpenAI presenta GPT (Generative Pre-trained Transformer), que supone un hito en el aprendizaje no supervisado del lenguaje por parte de las máquinas. Contaba con más de 117 millones de parámetros y abrió el camino a los siguientes grandes modelos de lenguaje[17].
- Google AI y el Langone Medical Center crean en 2020 un algoritmo que mejora los resultados de los radiólogos en la detección de potencial cáncer de pulmón.

16 Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser e Illia Polosukhin: "Attention is All You Need". (https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)

17 Los grandes modelos de lenguaje (LLM) son una categoría de modelos básicos entrenados sobre inmensas cantidades de datos, lo que los hace capaces de comprender y generar lenguaje natural y otros tipos de contenido para realizar una amplia variedad de tareas.Los LLM se han convertido en un nombre muy conocido gracias al papel que han desempeñado en llevar la IA generativa a la vanguardia del interés público, así como al punto en el que se están centrando las organizaciones para adoptar la inteligencia artificial en numerosas funciones empresariales y casos prácticos. (https://ibm.com/es-es/topics/large-language-models)

- Ese año, OpenAI lanza el GPT-3, con 175 millones de parámetros y capacidad para producir textos que simulan la redacción humana.
- En 2021, OpenAI presenta Dall-E, un sistema de inteligencia artificial multimodal que genera imágenes a partir de sugerencias de texto.
- Intel presenta en 2022 FakeCatcher, un detector de vídeos falsos cuya tecnología, según la empresa, puede detectar contenido deepfake con una precisión del 96%.
- En noviembre de 2022, OpenAI da a conocer ChatGPT, un chatbot de lenguaje natural que se convierte en un éxito inmediato por su capacidad de responder a preguntas y sugerencias del usuario y propicia, casi por sí solo, una nueva y gran ola de interés por las potencialidades de la inteligencia artificial.

IV.- EL IMPACTO DE CHATGPT

La aparición pública de ChatGPT y su gran éxito provocaron una enorme convulsión en el mercado. Una de las manifestaciones más evidentes de la misma fue la reacción al lanzamiento y a su impacto por parte de las grandes tecnológicas, que se apresuraron a presentar productos homólogos, como los chatbots Bing Chat de Microsoft (ahora llamado Copilot) y Bard, de Google, o los grandes modelos de lenguaje Llama, de Meta (compañía propietaria de Facebook, Instagram y WhatsApp) y Gemini, de la propia Google y desarrollado por su compañía DeepMind.

Al propio tiempo, durante el pasado año, marcas y compañías empezaron a usar la inteligencia artificial de muy diferentes formas en sus acciones de marketing; comenzaron a aparecer influencers virtuales creados con inteligencia artificial; se inició

una polémica a cuenta del uso de contenido sujeto a derechos de autor para el entrenamiento de los grandes modelos de lenguaje; algunas voces empezaron a advertir del riesgo del uso de programas de inteligencia artificial para generar contenidos falsos y difundirlos, y no faltaron ejemplos, como las famosas imágenes del papa Francisco I con singulares atuendos que se publicaron en el mes de marzo de 2023; se aceleraron o se iniciaron proyectos para dotar de regulación legal a la inteligencia artificial y, por fin, se oyeron voces de grandes personalidades del mundo tecnológico hablando al hilo de este boom de la inteligencia artificial y llamando a la responsabilidad y la prudencia respecto a su desarrollo.

Puede citarse en este sentido a Sundar Pichai, CEO de Google, que dijo: "Estamos trabajando con tecnología que será increíblemente beneficiosa, pero claramente tiene el potencial de causar daño de manera profunda. Y por eso creo que es muy importante que todos seamos responsables sobre cómo lo abordamos"; y a Bill Gates, co-fundador de Microsoft, que escribió en su blog[18]: "El desarrollo de la IA es tan fundamental como la creación del microprocesador, la computadora personal, Internet y el teléfono móvil. Cambiará la forma en que las personas trabajan, aprenden, viajan, obtienen atención médica y se comunican entre sí. Industrias enteras se reorientarán a su alrededor. Las empresas se distinguirán por lo bien que la utilizan".

"Esta nueva tecnología puede ayudar a las personas de todo el mundo a mejorar sus vidas", añadía Gates. "Al mismo tiempo, el mundo necesita establecer las reglas del camino para que los inconvenientes de la inteligencia artificial sean superados con creces por sus beneficios, y para que todos puedan disfrutar de esos beneficios sin importar dónde vivan o cuánto dinero tengan. La Era de la IA está llena de oportunidades y responsabilidades".

18 Gates Notes. The blog of Bill Gates: "The Age of AI has begun". (https://www.gatesnotes.com/The-Age-of-AI-Has-Begun)

V.- LOS ASISTENTES AVANZADOS

La cuestión de los beneficios y los inconvenientes, que Gates subrayaba en su texto, se hace aún más relevante con la nueva etapa que para la inteligencia artificial se abrió a finales del pasado mes de abril con la presentación casi simultánea por parte de OpenAI y Google de sus nuevos asistentes avanzados y que van en la dirección señalada por Sam Altman, CEO de Open AI, de que la inteligencia artificial dé lugar a la creación de un "colega supercompetente" que sea de ayuda en todo tipo de actividades y sea capaz de tomar decisiones y planificar por sí solo.

El nuevo modelo de inteligencia artificial generativa de OpenAI se denomina GPT-4o. La "o" del nombre corresponde a "omni», término con el que se alude a las capacidades para manejar y gestionar texto, voz y video en tiempo real, y generar respuestas en esos mismos formatos, con lo que la interacción humano-máquina es mucho más natural. El nuevo asistente, según ha comentado Altman, puede abordar ciertas tareas al instante y con las más complejas, hacer propuestas y plantear preguntas sobre las mismas si es necesario.

La novedad de Google es Project Astra, que la compañía define como su visión del futuro de los asistentes de inteligencia artificial y estará plenamente operativa a finales de este año. Tal y como señaló Demis Hassabis, cofundador de Google DeepMind, durante la presentación, es un prototipo de asistente de IA universal que pretende ser útil y acompañar en todos los aspectos de la vida cotidiana.

Astra es capaz de comprender el entorno y responder a las preguntas de los usuarios al respecto. Cuenta con capacidades robóticas humanizadas, como empatía, e identifica objetos, reconoce líneas de código en un ordenador y responde a instrucciones de audio con creaciones creativas; también registra y recuerda la posición de los objetos en el espacio.

VI.- OPORTUNIDADES Y AMENAZAS

El impacto de estos asistentes avanzados -que suponen un paso hacia la llamada inteligencia artificial general- ha sido analizado en un artículo elaborado por Google DeepMind con la colaboración de expertos de una decena de universidades y otras entidades[19]. En él se habla de las ventajas y riesgos que suponen y ello puede dar, a su vez, una idea de las ventajas y riesgos que presenta el desarrollo general de la inteligencia artificial.

Así, en el documento se alude a la capacidad de estos asistentes de empoderar a los usuarios, de actuar como un asesor de confianza en diversas cuestiones, de ayudar a tomar decisiones más informadas, de fomentar la creatividad y de contribuir a la resolución más rápida de ciertos problemas. En el lado negativo están la capacidad de equivocarse respecto a los intereses del individuo o la sociedad, de usar fuentes de información no fiables, de ignorar los efectos a largo plazo de sus recomendaciones y de proporcionar respuestas incorrectas o potencialmente dañinas para el usuario. Asimismo, estos asistentes podrían vulnerar la privacidad, limitar sus respuestas a los objetivos de sus compañías desarrolladoras, privilegiar determinados valores o ser usados para campañas maliciosas y ataques informáticos.

Estas consideraciones dan pie para mencionar otras ventajas de la inteligencia artificial, como su contribución a la mayor productividad y agilidad de procesos en las empresas y a la mayor seguridad de los trabajadores. En el campo de la salud, puede contribuir a la generalización de la atención médica, especialmente en países con pocos recursos, y a los descubrimientos y avances a consecuencia de su capacidad para gestionar gran cantidad de datos. La educación también podrá beneficiarse a través de

19 Jason Gabriel y otros: "The Ethics of Advanced AI Systems". (https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/ethics-of-advanced-ai-assistants/the-ethics-of-advanced-ai-assistants-2024-i.pdf)

la capacidad de la inteligencia artificial de ayudar a profesores y alumnos a identificar patrones de aprendizaje y a la difusión de conocimientos adaptados a las necesidades de cada persona o de una zona o país determinado.

En el lado negativo pueden citarse, sucintamente, el eventual impacto en el empleo por la capacidad de los sistemas de realizar, de forma más económica, determinadas tareas; la invasión de la privacidad a través de la acumulación de datos personales o la utilización de sistemas de categorización biométricas basados en características sensibles o en la captura indiscriminada de imágenes; la perpetuación de sesgos o prejuicios por un entrenamiento inadecuado de los sistemas o por el diseño algorítmico; un uso malicioso por parte de piratas informáticos; la contribución a la desigualdad económica, por cuanto que es más probable que la inteligencia artificial beneficie más a personas, empresas y países ricos; y su capacidad para la generación y difusión de información falsa y, en consecuencia, para la manipulación de las opiniones.

VII.- ALGUNOS DATOS ECONÓMICOS

El impacto de la inteligencia artificial tiene múltiples facetas, pero una que puede dar una idea de la magnitud del mismo es la económica, tanto si se atiende al volumen de mercado de esta tecnología en sí misma como al efecto que puede tener en la actividad de empresas y profesionales. En lo que respecta al volumen, Statista estimaba el pasado mes de abril que en 2024 alcanzará los 184.000 millones de dólares y que hasta 2030 experimentará una tasa de crecimiento anual compuesta (CAGR, por sus siglas en inglés) del 28,46%, con lo que en dicho año alcanzará los 826.700 millones.

Más optimista aún era el estudio que Bloomberg Intelligence dio a conocer en junio del año 2023, y que se refería exclusivamente a la inteligencia artificial generativa[20]. La división de estudios de la agencia de noticias señalaba que ese mercado puede alcanzar, en 2032, un volumen de 1,304 billones de dólares, pues experimentará, hasta entonces y desde 2022, una CAGR del 42%. En ese año, el volumen de mercado de la inteligencia artificial generativa era de 40.000 millones de dólares, según la misma fuente. Bloomberg señala asimismo que el porcentaje de la inteligencia artificial generativa en el gasto general en tecnología pasara del 3% estimado para 2024 al 12% que se calcula para 2032.

En cuanto al efecto general en la economía, PwC[21] calcula que la inteligencia puede tener un impacto de 15,7 billones de dólares en la economía mundial en el año 2030. De esa cantidad, 6,6 billones procederán de las mejoras en la productividad y 9,1 billones, del incremento del consumo.

En el ámbito europeo los programas Horizonte Europa y Europa Digital invierten 1.000 millones de euros al año en IA.

La Comisión también ha previsto movilizar inversiones adicionales del sector privado y de los Estados miembros para alcanzar un volumen de inversión anual de 20.000 millones EUR a lo largo de la década digital.

El Mecanismo de Recuperación y Resiliencia pone a disposición 134 000 millones de euros para el sector digital. Esta financiación pretende convertir a Europa en un líder mundial en el desarrollo de IA de vanguardia y confiable.

20 Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds" (https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/).

21 Sizing the prize. PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution. What's the real value of AI for your business and how can you capitalise?" (https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html)

SEGUNDA PARTE: EL PROCESO DE REGULACIÓN DE LA IA EN EUROPA

I.- CRONOLOGÍA DEL PROCESO. ANTECEDENTES

Una vez expuesta a grandes rasgos la evolución de la IA desde el punto de vista tecnológico, se va a analizar cuál ha sido el proceso hasta la reciente aprobación del Reglamento europeo sobre esta cuestión, comúnmente conocido como la Ley europea de IA, (AI Act)[22].

Los primeros pasos sobre la entrada de la IA en la normativa europea se producen en la Estrategia de la Comisión para la digitalización de la industria (COM (2016) 180 final) y en la Estrategia renovada de política industrial de la UE (COM (2017) 479 final).

Como parte de su estrategia digital, la UE decidió regular la inteligencia artificial (IA) para garantizar mejores condiciones de desarrollo y uso de esta tecnología innovadora. Consideraba que la IA podía aportar muchos beneficios, como son una mejor asistencia sanitaria, un transporte más seguro y limpio, una fabricación más eficiente y una energía más barata y sostenible.

Los principales hitos en el proceso de regulación de la IA en el Derecho Europeo son los siguientes[23]:

22 REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 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

23 https://digital-strategy.ec.europa.eu/es/policies/european-approach-artificial-intelligence

Marzo 2018	Comunicado de prensa: Grupo de expertos en IA y Alianza Europea de IA
Abril 2018	Comunicado de prensa: IA para Europa Comunicación: IA para Europa Documento de Trabajo: Responsabilidad en relación con las tecnologías digitales emergentes Declaración sobre cooperación sobre IA
Junio 2018	Lanzamiento de la Alianza sobre la IA Creación del grupo de expertos de alto nivel sobre IA
Diciembre 2018	Comisión Europea: Plan coordinado sobre IA Comisión Europea, comunicado de prensa: IA desarrollada en Europa Comisión Europea, Comunicación: IA desarrollada en Europa Consulta a los actores interesados sobre el proyecto de directrices éticas para una IA fiable
Abril 2019	Comunicación de la Comisión Europea: Construyendo la confianza en una IA centrada en el ser humano Grupo de expertos de alto nivel sobre IA:Directrices éticas para una IA fiable
Junio 2019	Primera Asamblea de la Alianza Europea sobre IA Grupo de expertos de alto nivel sobre IA: Recomendaciones sobre política e inversiones en relación con IA
Diciembre 2019	Grupo de expertos de alto nivel sobre IA: Prueba piloto de la lista de evaluación de IA confiable
Febrero 2020	Comisión Europea: Libro Blanco sobre IA; una aproximación europea a la excelencia y la confianza en la IA
Julio 2020	Evaluación de impacto inicial: requisitos éticos y legales sobre la IA Grupo de expertos de alto nivel sobre IA: lista de evaluación final sobre IA confiable Grupo de expertos de alto nivel sobre IA: recomendaciones sectoriales sobre IA confiable

(cont.)

Revista de Privacidad y Derecho Digital
ISSN digital 2444-5762, NÚM. 34, Madrid, Mayo-Agosto (2024), págs. 26-89

Octubre 2020	Segunda Asamblea de la Alianza Europea sobre IA
Abril 2021	Comisión Europea: Comunicación sobre el fomento de un enfoque europeo de la IA Comisión Europea: Propuesta de reglamento por el que se establecen normas armonizadas sobre la IA Comisión Europea: plan coordinado actualizado sobre la IA Comisión Europea: Evaluación de impacto de un reglamento sobre la IA
Junio 2021	Consulta pública sobre Responsabilidad Civil – adaptando las normas de responsabilidad a la era digital y la inteligencia artificial Comisión Europea: Propuesta de Reglamento sobre seguridad de los productos
Noviembre 2021	Consejo de la UE: texto de compromiso de la Presidencia de la IS sobre la Ley de IA . Conferencia de alto nivel sobre IA: de la ambición a la acción (3ª Asamblea de la Alianza Europea de IA). Comité Económico y Social Europeo, Dictamen sobre la Ley de IA
Diciembre 2021	Comité de las Regiones. Opinion sobre la Ley de IA Banco Central Europeo. Opinión sobre la Ley de IA.
Junio 2022	Lanzamiento del primer sandbox regulatorio de IA en España: Haciendo avanzar el Reglamento de IA
Septiembre 2022	Propuesta para una Directiva sobre responsabilidad derivada de la IA.
Diciembre 2022	Aproximación general del Consejo sobre la Ley de IA,
Junio 2023	Posición negociadora del Parlamento sobre la Ley de la IA
Diciembre 2023	Acuerdo Político alcanzado por los co-legisladores sobre la Ley de la IA
Enero 2024	Paquete de innovación en IA para apoyar a las empresas emergentes y a las pymes en Inteligencia Artificial

(cont.)

Febrero 2024	Oficina Europea de la IA
Marzo 2024	Aprobación por el Parlamento Europeo de la Ley de la IA
Mayo 2024	Aprobación por el Consejo Europeo de la Ley de la IA
Julio 2024	Publicación en el Diario Oficial de la Unión Europea
Agosto 2024	Entrada en vigor con carácter general
Febrero 2025	Entrada en vigor de las prohibiciones de prácticas prohibidas
Mayo 2025	Entrada en vigor de los Códigos de práctica
Agosto 2025	Entrada en vigor de las normas IA de uso general, incluida la gobernanza
Agosto 2027	Entrada en vigor de las obligaciones para sistemas de alto riesgo

Se ha identificado la irrupción de la inteligencia artificial como un cambio de era desde el punto de vista tecnológico. Al igual que la electricidad en el pasado, la inteligencia artificial (IA) está transformando nuestro mundo[24].

> *Está a nuestro alcance, cuando traducimos textos en línea o usamos una aplicación móvil para encontrar la mejor manera de ir a nuestro próximo destino. En casa, un termostato inteligente puede reducir las facturas de energía hasta en un 25% al analizar los hábitos de las personas que viven en ella y ajustar la temperatura en consecuencia. En el sector sanitario, los algoritmos pueden ayudar a los dermatólogos a realizar un mejor diagnóstico, por ejemplo, detectando el 95%*

24 Como indica la Unión *Europea:*
Al igual que hicieran la máquina de vapor o la electricidad en épocas anteriores, la IA está transformando nuestro mundo, nuestra sociedad y nuestra industria. El crecimiento de la capacidad informática y la disponibilidad de datos, así como los avances en los algoritmos, han convertido la IA en una de las tecnologías más estratégicas del siglo XXI. Es mucho lo que está en juego. Nuestra forma de abordar la cuestión de la IA definirá el mundo en el que vamos a vivir. En medio de una feroz competencia mundial, se requiere un marco europeo sólido. Inteligencia

de los cánceres de piel aprendiendo de grandes conjuntos de imágenes médicas. Al dar sentido a grandes cantidades de datos para ofrecer soluciones eficientes, la IA mejora los productos, procesos y modelos de negocio en todos los sectores económicos. Puede ayudar a las empresas a identificar qué máquinas necesitan mantenimiento antes de averiarse. La IA también transforma los servicios públicos.[25]

En paralelo, la regulación de la IA contribuye a la protección de la democracia, el Estado de derecho y la sostenibilidad medioambiental. Sin embargo, la IA potencialmente entraña un alto riesgo. El Reglamento fija una serie de obligaciones para la IA en función de sus riesgos potenciales y su nivel de impacto.

II.- LA PREPARACIÓN. 2018-2020

II.1- EL GRUPO DE EXPERTOS EN IA (AI HLEG) Y LA ALIANZA EUROPEA DE IA (AI ALLIANCE)

Para afrontar los desafíos y aprovechar las oportunidades que ofrece la IA, la Comisión publicó una Estrategia europea en abril de 2018, COM (2018) 237. La Comisión propuso en la misma un enfoque que coloca a las personas en el centro del desarrollo de la IA (IA centrada en el ser humano) y alienta el uso de esta tecnología para ayudar a resolver los mayores desafíos del mundo: desde curar enfermedades hasta combatir el cambio climático y anticipar desastres naturales, hacer que

artificial para Europa, 24 de abril de 2018,https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN

25 COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Plan coordinado sobre la inteligencia artificial, 7 de diciembre de 2019 chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://eur-lex.europa.eu/resource.html?uri=cellar:22ee84bb-fa04-11e8-a96d-01aa75ed71a1.0022.02/DOC_1&format=PDF.

el transporte sea más seguro, luchar contra la delincuencia o mejorar la ciberseguridad.

Tras la publicación de su Estrategia europea sobre la inteligencia artificial en abril de 2018, la Comisión creó el Grupo de Expertos de Alto Nivel en Inteligencia Artificial (AI HLEG). formado por cincuenta y dos expertos independientes que representaban al mundo académico, la industria y la sociedad civil.

La Comisión Europea formuló el 12 de marzo de 2018 una convocatoria para formar este grupo de expertos de alto nivel sobre Inteligencia Artificial. El plazo terminaba el 9 de abril de 2018 y los resultados se publicaron en junio de 2018.

En su comunicado de prensa, se decía que:

> *La Comisión Europea está a punto de iniciar un diálogo con todas las personas involucradas en el futuro de la Inteligencia Artificial (IA) en Europa para fomentar un debate abierto en torno a todos los aspectos del desarrollo de la IA y su impacto en la economía y la sociedad. Así, ha abierto una convocatoria para formar parte del grupo de alto nivel de expertos en IA, que funcionará como grupo director de este fórum donde intervendrán todos los diferentes agentes involucrados: empresas, universidades, políticos, organizaciones de consumidores, sindicatos y otros representantes de la sociedad civil. El fórum conformará la Alianza Europea de IA.*

La misión principal del grupo de 52 expertos era contribuir a la implementación de la Estrategia Europea de Inteligencia Artificial, así como la elaboración de recomendaciones sobre el futuro desarrollo de políticas relacionadas con la IA, y abordar las cuestiones problemáticas sobre aspectos éticos, jurídicos y sociales, incluyendo cuestiones socioeconómicas[26].

26 Entre las labores del grupo se encontraban:
- Asesorar a la Comisión sobre los próximos pasos que aborden los retos y las oportunidades de la IA a corto y medio plazo, mediante recomendaciones que se imbuirán en el proceso de redacción de políticas, así como el proceso de evaluación legislativa y el desarrollo de una estrategia digital de nueva generación.

El Grupo publicó un primer proyecto de directrices éticas en diciembre de 2018, al que siguieron una consulta con las partes interesadas y reuniones con representantes de los Estados miembros para recabar sus opiniones. Este trabajo responde al plan coordinado con los Estados miembros para fomentar el desarrollo y la utilización de la inteligencia artificial en Europa, también presentado en diciembre de 2018. El Grupo de Expertos terminó su actividad en julio de 2020.

La Alianza Europea de IA (AI Alliance) es una iniciativa de la Comisión Europea para establecer un diálogo político abierto con los ciudadanos, la sociedad civil, las organizaciones empresariales, los consumidores, los sindicatos, el mundo académico, las autoridades públicas y los expertos sobre inteligencia artificial. Desde su lanzamiento en 2018, AI Alliance ha involucrado a alrededor de 6.000 partes interesadas a través de eventos regulares, consultas públicas e intercambios de foros en línea. Este foro ha contribuido a algunas de las iniciativas políticas más importantes lanzadas en el campo de la IA.

La AI Alliance fue creada inicialmente para dirigir el trabajo del Grupo de Expertos de Alto Nivel en Inteligencia Artificial (AI HLEG). Las Directrices Éticas del Grupo, así como sus Recomendaciones de Política e Inversión fueron documentos importantes que dieron forma al concepto de IA confiable, contribuyendo al enfoque de la Comisión respecto a la IA.

- Asesorar a la Comisión sobre dinamización y mecanismos de difusión para interactuar con el grupo más amplio posible de agentes involucrados en el contexto de la Alianza AI, así como compartir información y recabar sus puntos de vista sobre el trabajo del grupo y de la Comisión.
- Proponer a la Comisión las directrices en ética de la IA, que abarcan cuestiones como la equidad, seguridad, transparencia, el futuro del mercado laboral, democracia y el impacto de la aplicación de la Carta de Derechos Fundamentales, incluyendo la privacidad y la protección de datos personales, dignidad, protección de los consumidores y no discriminación. Estas directrices se basarán en el trabajo del Grupo Europeo de Ética en la Ciencia y las Nuevas Tecnologías (un grupo asesor independiente establecido por el Presidente de la Comisión Europea) y de la Agencia Europea de Derechos Fundamentales.

Durante todo el proceso, AI Alliance[27] ha continuado promoviendo la IA confiable, compartiendo las mejores prácticas entre sus miembros y ayudando a los desarrolladores de IA y otras partes interesadas a aplicar los requisitos clave, a través de la herramienta ALTAI, una lista de evaluación práctica para la IA confiable, a la que más adelante se hace referencia.

II.2- EL PLAN COORDINADO SOBRE IA

En abril de 2018 la Comisión publicó una Estrategia Europea sobre la IA. La Estrategia coloca a la persona en el centro del desarrollo de la IA: es una IA centrada en el ser humano. Con el fin de potenciar la capacidad tecnológica e industrial de la UE, la Estrategia plantea:

- impulsar la adopción de la IA en todos los ámbitos de la economía,
- prepararse para las transformaciones socioeconómicas
- garantizar el establecimiento de un marco ético y jurídico apropiado.

27 Los principales eventos de la Alianza sobre la IA han sido los siguientes:
1.ª Asamblea Europea de la Alianza AI — junio de 2019: Celebración de un año desde el lanzamiento de la comunidad AI Alliance en la que se discutieron los principales resultados del AI HLEG.
2.ª Asamblea Europea de la Alianza de la IA — octubre de 2020: Debate sobre los resultados de la consulta pública sobre el Libro Blanco sobre la IA.
3ª Asamblea Europea de la Alianza de la IA — septiembre de 2021: Organizada en cooperación con la Presidencia eslovena del Consejo de la UE con el título "Conferencia de alto nivel sobre IA: De la ambición a la acción". Su objetivo fue discutir la excelencia y la confianza en IA a nivel global.
En marzo y junio de 2022 se celebraron dos eventos de interés : *European Excellence and Trust in the world* y *Bringing AI Regulation Forward.* En ambos casos se trataba de involucrar a la comunidad de IA en el debate de aspectos de la IA confiable, en el contexto de la divulgación internacional de la UE y la coordinación entre los Estados miembros, respectivamente.
La 4ª Asamblea de AI Alliance tuvo lugar en noviembre de 2023 en Madrid y se centró en aspectos políticos: La UE *Líder de IA confiable a nivel mundial*. La organizaron la Comisión y el Ministerio de Economía y Transformación Digital de España, en el marco de la Presidencia española del Consejo de la UE, y estuvo abierta al público.

Para desarrollar la Estrategia, la Comisión, junto con los Estados miembros, elaboró un plan coordinado sobre la inteligencia artificial, que presentó en diciembre de 2018, para *crear sinergias, reunir datos,-la materia prima de numerosas aplicaciones de IA-, e incrementar las inversiones conjuntas*. El objetivo era fomentar la cooperación transfronteriza y movilizar a todos los agentes con el fin de aumentar las inversiones públicas y privadas hasta un mínimo de 20.000 millones EUR anuales durante la década siguiente, esto es, la actual, que se denomina la década digital.

En este marco, la Comisión intensificó su diálogo con todas las partes interesadas relevantes de la industria, institutos de investigación y autoridades públicas. Se creó el programa Europa Digital, muy relevante para contribuir a que la IA esté a disposición de las pequeñas y medianas empresas en todos los Estados miembros, a través de polos de innovación digital, instalaciones de ensayo y experimentación reforzadas, espacios de datos y programas de formación.

II.3- DIRECTRICES ÉTICAS PARA UNA IA FIABLE

El 8 de abril de 2019, el Grupo de expertos de alto nivel sobre la IA presentó unas Directrices éticas para una IA fiable. Según las Directrices, la IA fiable debe ser:

(1) legal: respeto de todas las disposiciones legales y reglamentarias aplicables

(2) ética: respeto de los principios y valores éticos

(3) robusta, desde una perspectiva técnica, teniendo en cuenta al mismo tiempo su entorno social.

Las Directrices exponen un conjunto de 7 requisitos clave que los sistemas de IA deben cumplir para ser considerados fiables. Además, se establece una lista de evaluación específica que tiene por objeto facilitar la verificación del cumplimiento de cada uno de esos requisitos fundamentales:

1.- *Intervención y supervisión humanas*: los sistemas de IA deben empoderar a los seres humanos, permitiéndoles tomar decisiones con conocimiento de causa y fomentando sus derechos fundamentales. Al mismo tiempo, deben garantizar mecanismos de supervisión adecuados, lo que puede lograrse mediante enfoques humanos.

2.- *Solidez técnica y seguridad*: Los sistemas de IA deben ser resilientes y seguros. Deben garantizar un plan de retroceso en caso de que algo salga mal, así como ser exactos, fiables y reproducibles. Esta es la única forma de garantizar que también puedan minimizarse y evitarse los daños involuntarios.

3.- *Privacidad y gestión de datos*: además de garantizar el pleno respeto de la privacidad y la protección de datos, también deben establecerse mecanismos adecuados de gobernanza de datos, teniendo en cuenta la calidad y la integridad de los datos, y permitiendo un acceso legítimo a los datos.

4.- *Transparencia*: los modelos de negocio de los datos, los sistemas y la IA deben ser transparentes. Los mecanismos de trazabilidad pueden ayudar a lograrlo. Además, los sistemas de IA y sus decisiones deben explicarse de manera adaptada a las partes interesadas afectadas. Los seres humanos deben ser conscientes de que están interactuando con un sistema de IA y deben estar informados de las capacidades y limitaciones del sistema.

5.- *Diversidad, no discriminación y equidad*: debe evitarse el sesgo injusto, ya que podría tener múltiples consecuencias negativas, desde la marginación de los grupos vulnerables hasta la exacerbación de los prejuicios y la discriminación. Fomentar la diversidad: los sistemas de IA deben ser accesibles para todos, independientemente de cualquier discapacidad, e implicar a las partes interesadas pertinentes a lo largo de todo su ciclo vital.

Revista de Privacidad y Derecho Digital
ISSN digital 2444-5762, NÚM. 34, Madrid, Mayo-Agosto (2024), págs. 26-89

6.- *Bienestar social y medioambiental*: los sistemas de IA deben beneficiar a todos los seres humanos, incluidas las generaciones futuras. Por lo tanto, debe garantizarse que sean sostenibles y respetuosos con el medio ambiente. Además, deben tener en cuenta el medio ambiente, incluidos otros seres vivos, y debe estudiarse detenidamente su impacto social y social.

7.- *Rendición de cuentas*: deben implantarse mecanismos que garanticen la responsabilidad y la rendición de cuentas de los sistemas de inteligencia artificial y de sus resultados. La auditabilidad, que permite la evaluación de algoritmos, datos y procesos de diseño, desempeña un papel clave, especialmente en aplicaciones críticas. Además, debe garantizarse una reparación accesible.

El documento también proporciona una lista de evaluación que pone en práctica los requisitos clave y ofrece orientaciones para aplicarlos en la práctica.

Esta lista de evaluación se sometió a un proceso piloto, en el que se invitó a todas las partes interesadas a ponerla a prueba y a proporcionar información práctica sobre cómo podía mejorarse. El Grupo de Alto Nivel sobre IA presentó la lista de evaluación final para una IA fiable (ALTAI) en julio de 2020.

ALTAI es una herramienta práctica que traduce las directrices éticas en una lista de control accesible y dinámica (autoevaluación). La lista de control puede ser utilizada por los desarrolladores e implementadores de IA que deseen respetar los requisitos clave en la práctica.

El Grupo de Expertos inicialmente (junio de 2019) había formulado unas Recomendaciones en relación con el plano político y las inversiones para una IA fiable. Se trataba de 33 Recomendaciones para guiar una IA fiable hacia la sostenibilidad, el crecimiento, la competitividad y la inclusión. Al mismo tiempo las Recomendaciones empoderan, benefician y protegen a los ciudadanos europeos.

En julio de 2020, el Grupo de Expertos planteó la posibilidad de aplicar las Recomendaciones indicadas, que habían sido publicadas previamente por el Grupo en tres campos específicos: sector público, salud y fabricación e internet de las cosas.

II.4- COMUNICACIÓN DE LA COMISIÓN EUROPEA: GENERAR CONFIANZA EN UNA IA CENTRADA EN EL SER HUMANO

La Comisión presentó en abril de 2019 el programa para generar confianza en la inteligencia artificial continuando el trabajo del grupo de expertos de alto nivel[28].

En este programa la Comisión, en primer lugar, a partir del trabajo del grupo de expertos independientes nombrados en junio de 2018, decidió iniciar una fase piloto que pretendía garantizar que las directrices éticas para el desarrollo y el uso de la inteligencia artificial pudieran aplicarse en la práctica. La Comisión invitaba a la industria, institutos de investigación y autoridades públicas a probar la lista detallada de evaluación elaborada por el grupo de expertos de alto nivel, que complementa las directrices. Al mismo tiempo comunicaba a las empresas, las administraciones públicas y las organizaciones que ya podían adherirse a la Alianza europea de la inteligencia artificial y recibir una notificación en relación con el inicio del proyecto piloto.

Además, la Comisión confirmaba que los planes presentados formaban parte del marco de la estrategia para la inteligencia artificial de abril de 2018, cuyos objetivos eran aumentar las inversiones públicas y privadas hasta un mínimo de 20.000 millones de euros anuales en la siguiente década, facilitar el acceso a una mayor cantidad de datos, fomentar el talento y garantizar la confianza[29].

28 https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52019DC0168

29 En este sentido se pronunciaron algunos miembros de la Comisión. El vicepresidente responsable del Mercado Único Digital, Andrus Ansip, afirmó lo siguiente:

La Comisión reiteraba que la inteligencia artificial puede aportar beneficios a una amplia gama de sectores, como la asistencia sanitaria, el consumo de energía, la seguridad de los automóviles, la agricultura, el cambio climático y la gestión del riesgo financiero. La inteligencia artificial también puede ayudar a detectar el fraude y las amenazas de ciberseguridad y permite a las fuerzas y cuerpos de seguridad luchar contra la delincuencia con más eficacia. Sin embargo, también implica nuevos retos para el futuro del trabajo y plantea cuestiones jurídicas y éticas.

Al mismo tiempo la Comisión consideraba fundamental la creación de un consenso internacional sobre la inteligencia artificial centrada en el ser humano La Comisión proponía trasladar su enfoque sobre la ética de la inteligencia artificial al conjunto del planeta, ya que las tecnologías, los datos y los algoritmos no conocen fronteras. En este sentido, la Comisión decidía reforzar la cooperación con socios afines, como Japón, Canadá o Singapur, y seguir desempeñando un papel activo en las discusiones e iniciativas internacionales, incluidos el G7 y el G20.

Además, para garantizar el desarrollo ético de la inteligencia artificial, la Comisión puso en marcha en 2019 un conjunto de redes de centros de excelencia especializados en investigación sobre inteligencia artificial y redes de polos de innovación digital. También junto con los Estados miembros y las partes interesadas, fomentó debates para desarrollar y aplicar un modelo

---

Acojo con satisfacción el trabajo realizado por nuestros expertos independientes. La dimensión ética de la inteligencia artificial no es un lujo ni un añadido. Nuestra sociedad solo puede beneficiarse plenamente de las tecnologías si existe confianza. La inteligencia artificial ética es una propuesta beneficiosa para todos, que puede convertirse en una ventaja competitiva para Europa: liderar una inteligencia artificial centrada en el ser humano en la que la gente pueda confiar.
La comisaria responsable de la Economía y Sociedad Digitales, Mariya Gabriel, expuso:
Hoy damos un paso importante hacia la inteligencia artificial ética y segura en la UE. Tras un amplio y constructivo compromiso alcanzado por muchas partes interesadas, incluidas las empresas, el mundo académico y la sociedad civil, ahora contamos con unos fundamentos sólidos basados en los valores de la UE. Vamos a poner estos requisitos en práctica y, al mismo tiempo, fomentaremos un debate internacional sobre la inteligencia artificial centrada en el ser humano.

para el intercambio de datos y para hacer el mejor uso de los espacios comunes de datos.

II.5- LIBRO BLANCO DE LA COMISIÓN SOBRE IA: UN ENFOQUE EUROPEO ORIENTADO A LA EXCELENCIA Y LA CONFIANZA[30]

En febrero de 2020, la Comisión publicó un Libro Blanco para presentar su enfoque para garantizar la excelencia y confianza en la IA. El documento, que se fundamentaba en el trabajo del Grupo de Expertos, AI HLEG, se sometió a una consulta pública abierta. Se recibieron más de 1.215 contribuciones (incluidos 400 documentos de posición) a través del cuestionario en línea y los canales de comunicación de la AI Alliance. Se organizaron reuniones y mesas redondas con expertos en IA para plantear una visión desde todos los puntos de vista sociales, económicos y científicos.

En este Libro Blanco, la Comisión respalda un enfoque basado en la regulación y en la inversión, que tiene el doble objetivo de promover la adopción de la inteligencia artificial y de abordar los riesgos[31] vinculados a determinados usos de esta nueva tecnología.

La finalidad del Libro Blanco era formular alternativas políticas para alcanzar estos objetivos. No abordaba el desarrollo ni el uso de la inteligencia artificial para fines militares. La Comisión invitaba a los Estados miembros, a otras instituciones europeas y a todas las partes interesadas, como la industria, los interlocutores sociales, las organizaciones de la sociedad civil, los investigadores, el público general y demás personas con interés

30 LIBRO BLANCO sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:-52020DC0065

31 En el LIBRO BLANCO se citan como riesgos potenciales, la opacidad en la toma de decisiones, la discriminación de género o de otro tipo, la intromisión en nuestras vidas privadas o su uso con fines delictivos.

en la materia, a que presentaran sus opiniones con respecto de las opciones que se mostraban en el Libro Blanco.

En el Libro Blanco se partía de que Europa podía aunar su potencial tecnológico e industrial con una infraestructura digital de gran calidad y un marco regulador basado en sus valores fundamentales para convertirse en líder mundial de la innovación en la economía de los datos y sus aplicaciones, tal como establecía la Estrategia Europea de Datos. Con este fundamento se podría desarrollar un ecosistema de inteligencia artificial que acercara las ventajas de la tecnología a la sociedad y la economía europeas en su conjunto:

- a los ciudadanos, para que obtuvieran nuevos beneficios, como una mejor atención sanitaria, una menor cantidad de averías de los aparatos domésticos, unos sistemas de transporte más seguros y limpios, o mejores servicios públicos;
- al desarrollo empresarial, por ejemplo, mediante una nueva generación de productos y de servicios en áreas en las que Europa es particularmente fuerte (maquinaria, transporte, ciberseguridad, agricultura, economía verde y circular, atención sanitaria y sectores de gran valor añadido, como la moda y el turismo);
- a los servicios de interés público, por ejemplo mediante una reducción de los costes de la prestación de servicios (transporte, educación, energía y gestión de los residuos), una mayor sostenibilidad de los productos, o proporcionando a los servicios y fuerzas de seguridad las herramientas adecuadas para que aseguren la protección de los ciudadanos, garantizando correctamente el respeto de sus derechos y libertades.

Los pilares fundamentales del Libro Blanco eran:

- El marco político por el que se establecían medidas para armonizar los esfuerzos a escala regional, nacional y europea. En colaboración con los sectores público y privado,

los objetivos del marco eran movilizar recursos para obtener un ecosistema de excelencia a lo largo de toda la cadena de valor, partiendo de la investigación y la innovación, así como crear los incentivos adecuados para acelerar la adopción de soluciones basadas en la inteligencia artificial, también por parte de las pequeñas y medianas empresas (pymes).

- Los elementos clave de un futuro marco normativo para la inteligencia artificial en Europa que pudieran generar un ecosistema de confianza exclusivo. Para hacerlo, este marco debía velar por el cumplimiento de las normas de la UE, especialmente las normas de protección de los derechos fundamentales y los derechos de los consumidores en relación con los sistemas de inteligencia artificial que operan en la UE y presentan un riesgo elevado.

Según la Comisión Europea, generar un ecosistema de confianza constituye un objetivo político en sí mismo, y debe ofrecer seguridad a los ciudadanos para que adopten las aplicaciones de la inteligencia artificial y seguridad jurídica a las empresas y organismos públicos para que innoven utilizando esta herramienta.

La Comisión respalda con este planteamiento un enfoque antropocéntrico que se base en la Comunicación "Generar confianza en la inteligencia artificial centrada en el ser humano", y tendrá en cuenta también los resultados obtenidos durante la fase de prueba de las directrices éticas elaboradas por el grupo de expertos de alto nivel sobre la IA.

La Estrategia Europea de Datos, que acompañaba al Libro Blanco, tenía por objeto ayudar a Europa a convertirse en la economía con agilidad en el manejo de los datos más atractiva, segura y dinámica del mundo, lo que fortalecería a Europa con información para reforzar sus decisiones y mejorar las vidas de todos sus ciudadanos. La Estrategia establecía varias medidas políticas, como la movilización de inversiones públicas y privadas, necesarias para alcanzar este objetivo.

Finalmente, en el informe de la Comisión adjunto al Libro Blanco, se analizaban las repercusiones de la inteligencia artificial, el internet de las cosas y otras tecnologías digitales en la legislación en materia de seguridad y responsabilidad civil.

Este Libro Blanco es el principal antecedente para la presentación por parte de la Comisión Europea de una propuesta de Reglamento por el que se establecen normas armonizadas sobre IA y el Plan Coordinado Revisado sobre IA, en abril de 2021.

III.- LA TRAMITACIÓN. 2021-2023

En abril de 2021, la Comisión presentó su paquete de IA, que incluía:

- su Comunicación sobre el fomento de un enfoque europeo de la IA;
- una revisión del Plan Coordinado sobre Inteligencia Artificial (con los Estados miembros de la UE),
- su propuesta de marco regulador sobre inteligencia artificial y la evaluación de impacto correspondiente.

En este marco, la Comisión consideraba que la creación de una IA fiable permitía un entorno seguro y favorable a la innovación para los usuarios, los desarrolladores y los implementadores.

La Comisión propuso tres iniciativas jurídicas interrelacionadas que contribuirían a crear una IA fiable:

1.- un marco jurídico europeo para la IA que defendiera los derechos fundamentales y abordara los riesgos de seguridad específicos de los sistemas de IA;

2.- un marco de responsabilidad civil: adaptación de las normas de responsabilidad a la era digital y a la IA;

3.- una revisión de la legislación sectorial en materia de seguridad (por ejemplo, el Reglamento sobre máquinas o la Directiva relativa a la seguridad general de los productos).

III.1- CONFERENCIA SOBRE EL FUTURO DE EUROPA

La Conferencia sobre el Futuro de Europa[32] consistió en una serie de debates y discusiones protagonizados por ciudadanos europeos que tuvieron lugar entre abril de 2021 y mayo de 2022. Estos debates permitieron a ciudadanos de toda Europa compartir sus ideas y contribuir a configurar nuestro futuro común.

La Conferencia se llevó a cabo a través de una novedosa plataforma digital multilingüe, mediante la cual todos los europeos pudieron compartir sus ideas, así como por medio de paneles nacionales y paneles europeos de ciudadanos. En la plataforma hubo más de cinco millones de visitas y en los eventos participaron más de setecientos mil ciudadanos. La Conferencia logró crear un foro público de debate abierto, inclusivo y transparente con los ciudadanos sobre diversas prioridades y retos**.**

Tras un año de debates, la Conferencia sobre el Futuro de Europa concluyó oficialmente el 9 de mayo de 2022, Día de Europa, en Estrasburgo. Ese día, los copresidentes del Comité Ejecutivo de la Conferencia presentaron un informe final con las cuarenta y nueve propuestas a los presidentes del Parlamento Europeo, del Consejo y de la Comisión**.**

Las propuestas abordaban nueve temas: cambio climático y medio ambiente; salud; una economía más fuerte, justicia social y empleo; la Unión en el mundo; valores y derechos, Estado de Derecho y seguridad; transformación digital; democracia europea; migración; educación, cultura, juventud y deporte. Las propuestas incluían objetivos generales y más de trescientas medidas concretas.

32 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/conference-future-europe_es

El Reglamento de Inteligencia artificial responde a las propuestas de los ciudadanos de la Conferencia sobre el Futuro de Europa, más concretamente a la propuesta 12(10), relativa a la mejora de la competitividad de la UE en sectores estratégicos, la propuesta 33(5), sobre una sociedad segura y fiable, incluida la lucha contra la desinformación y la garantía de que las personas tengan el control final, la propuesta 35 sobre la promoción de la innovación digital, (3) la garantía de la supervisión humana y (8) un uso fiable y responsable de la IA, el establecimiento de salvaguardias y la garantía de transparencia, y la propuesta 37(3), sobre el uso de la IA y las herramientas digitales para mejorar el acceso de los ciudadanos a la información, incluidas las personas con discapacidad.

III.2- LANZAMIENTO DEL PRIMER SANDBOX REGULATORIO DE IA EN ESPAÑA[33]

El Sandbox Regulatorio de IA[34], elaborado por el Reino de España en colaboración con la Comisión Europea, se impulsa en junio de 2022. Es un espacio digital que intenta conectar a las autoridades competentes con las compañías desarrolladoras de Inteligencia Artificial para definir de forma conjunta buenas prácticas a la hora de implementar desde ese momento la futura regulación europea de Inteligencia Artificial, y garantizar su aplicación[35].

33 Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial. Este Real Decreto entró en vigor en noviembre de 2023. Según su Disposición Final Segunda su vigencia estaba establecida en un máximo de treinta y seis meses desde su entrada en vigor o, en su caso, hasta que sea aplicable en el Reino de España el Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial.

34 https://espanadigital.gob.es/lineas-de-actuacion/sandbox-regulatorio-de-ia

35 El art.1 del Real Decreto 817/2023 dispone que el presente real decreto tiene por objeto establecer un entorno controlado de pruebas para ensayar el cumplimiento de ciertos requisitos por parte de algunos sistemas de inteligencia artificial que puedan suponer riesgos para la seguridad, la salud y los derechos fundamentales de las personas. Asimismo, se regula

El Sandbox pretende generar directrices de buenas prácticas y guías que preparen y sensibilicen a las empresas, especialmente a pymes y startups, para facilitar la puesta en marcha del entonces futuro reglamento. Su función como programa piloto ha sido comprobar la operatividad de los requisitos del futuro reglamento europeo de IA, así como las evaluaciones de conformidad o las actividades posteriores a la comercialización. Por tanto, ha permitido documentar tanto las obligaciones que deben cumplir los proveedores de sistemas de IA, y su implementación, como el método de control y seguimiento adecuado para las autoridades nacionales de supervisión.

El Sandbox ha pretendido reforzar la cooperación de todos los posibles actores a nivel europeo, abierto a los Estados miembros, que también han podido seguir o unirse al piloto[36].

el procedimiento de selección de los sistemas y entidades que participarán en el entorno controlado de prueba

36 Basándose en este concepto, los expertos en ciberseguridad pensaron que una buena forma de proteger las máquinas de los virus era aislarlos en una "caja de arena", es decir, aislar el programa infectado con un virus informático del resto del sistema para que el virus no se extienda. Esto es lo que se conoce como tecnología Sandbox.
Pero principalmente el Sandbox sirve de ayuda a las empresas a ser menos vulnerables frente al creciente número de ataques informáticos que están sufriendo.
Si el malware se activa y comienza a atacar, el entorno controlado del sandbox permite estudiar y analizar su funcionamiento y tácticas de ataque permitiendo aprender de ellas. Como si se encontrara en una placa y los científicos observaran al virus desarrollarse dentro de ella.
Después, en un entorno empresarial real, se utilizan técnicas de Machine Learning para adelantarse a ese comportamiento malicioso y neutralizarlo incluso antes de que el malware se active. Así, los sistemas de seguridad informática mejoran su eficiencia minimizando el riesgo de producirse daños o pérdida de información de las empresas por estos ataques.
Con esta tecnología, las empresas mejoran su seguridad y a la vez generan técnicas de protección que luego acaban llegando a los programas y dispositivos de consumo general protegiendo al resto de usuarios en sus casas.
https://computerhoy.com/reportajes/tecnologia/que-es-sandbox-529177

III.3- PROPUESTA PARA UNA DIRECTIVA SOBRE RESPONSABILIDAD DERIVADA DE LA IA

En una encuesta a este ámbito realizada en 2021, la cuestión de la responsabilidad civil figuraba entre los tres principales obstáculos a la utilización de la IA por parte de las empresas europeas. Se citó como el obstáculo externo más importante (43 %) en el caso de las empresas que tenían previsto recurrir a la IA, pero que aún no lo habían hecho. El problema se planteaba por la falta de garantías en relación con que las víctimas de daños causados por la IA obtuvieran una protección equivalente a la de las víctimas de daños causados por los demás productos. También por la inseguridad jurídica de las empresas que desarrollaban o utilizaban la IA en relación con su posible exposición a responsabilidad civil.

Para superar estos problemas, la Comisión ha adoptado dos propuestas con el fin de adaptar las normas de responsabilidad a la era digital, la economía circular y el impacto de las cadenas de valor mundiales al ámbito de la IA. De esta forma pretende además evitar la aparición de adaptaciones a la IA específicas, fragmentadas, de normas nacionales en materia de responsabilidad civil.

La Propuesta de Directiva, en primer lugar, propone modernizar las normas existentes sobre la responsabilidad objetiva de los fabricantes por los productos defectuosos de forma que se adapten a la nueva realidad digital; en segundo lugar, presenta una armonización específica de las normas nacionales sobre responsabilidad civil en materia de IA, con el fin de permitir que las víctimas de daños relacionados con la IA obtengan una indemnización.

En junio de 2021 se lanzó una consulta pública sobre responsabilidad civil, con el fin de preparar la adaptación de las normas sobre esta materia a la era digital y a la inteligencia artificial. La Comisión Europea formuló una Propuesta de Reglamento sobre seguridad de los productos.

En septiembre de 2022 se presentó una Propuesta de Directiva sobre responsabilidad derivada de la IA[37]

La Exposición de Motivos de la Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificia (Directiva sobre responsabilidad en materia de IA) articula la relación entre ambas cuestiones:

> *La Comisión adopta un enfoque holístico en su política de responsabilidad en materia de IA, proponiendo adaptaciones de la responsabilidad del productor por productos defectuosos en virtud de la Directiva sobre responsabilidad por los daños causados por productos defectuosos, y la armonización específica en el marco de la presente propuesta. Estas dos iniciativas políticas están estrechamente vinculadas y forman un paquete, ya que las demandas que entran en sus ámbitos de aplicación se refieren a diferentes tipos de responsabilidad. La Directiva sobre responsabilidad por los daños causados por productos defectuosos cubre la responsabilidad objetiva del productor por productos defectuosos, lo que da lugar a una indemnización por determinados tipos de daños, principalmente sufridos por particulares. La presente propuesta cubre las demandas nacionales de responsabilidad fundamentadas principalmente en la culpa de cualquier persona con el fin de indemnizar por cualquier tipo de daño y a cualquier tipo de víctima. Se complementan entre sí para formar un sistema general de responsabilidad civil eficaz. Juntas, estas normas promoverán la confianza en la IA (y otras tecnologías digitales) garantizando que las víctimas reciban una indemnización efectiva si, a pesar de los requisitos preventivos de la Ley de IA y otras normas de seguridad, se producen daños.*

La Propuesta garantiza que las víctimas se beneficien de las mismas normas de protección cuando se vean perjudicadas por

37 https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52022PC0496

productos o servicios de IA que cuando se producen daños en cualquier otro ámbito. También permite establecer un equilibrio entre la protección de los consumidores y el fomento de la innovación, eliminando los obstáculos adicionales para que las víctimas accedan a la indemnización. Además, establece garantías para el sector de la IA mediante la introducción, por ejemplo, del derecho a impugnar una reclamación de responsabilidad basada en una presunción de causalidad.

La Propuesta de Directiva sobre responsabilidad en materia de IA pretende establecer normas uniformes sobre el acceso a la información y la reducción de la carga de la prueba en relación con los daños provocados por los sistemas de IA, estableciendo una protección más amplia para las víctimas (ya sean particulares o empresas) y fomentando el sector de la IA mediante mayores garantías. Además, armoniza determinadas normas aplicables a las reclamaciones que no entran en el ámbito de aplicación de la Directiva sobre responsabilidad por los daños causados por productos defectuosos, en los casos en que los daños se deban a un comportamiento ilícito, lo cual incluye, por ejemplo, las violaciones de la privacidad o los daños causados por problemas de seguridad. Las nuevas normas facilitarán, por ejemplo, la obtención de una indemnización si alguien ha sido discriminado en un proceso de contratación que implique tecnología de IA.

En conclusión, la Propuesta ofrece más garantías a las personas que sufren daños en el ámbito de la IA.

En primer lugar, a través de la presunción de causalidad, la Propuesta facilita a las víctimas la demostración de que una persona ha provocado los daños; en circunstancias en las que se haya probado la culpa y parezca razonablemente probable que exista un nexo causal entre el daño y el funcionamiento de la IA, la denominada "presunción de causalidad" permite abordar las dificultades experimentadas por las víctimas para tener que explicar detalladamente la manera en que se ha provocado el

daño por una culpa u omisión concretas, lo que puede ser especialmente difícil cuando se plantea en el marco de sistemas de IA complejos.

En segundo lugar, las víctimas dispondrán de más herramientas para solicitar reparación legal gracias a la introducción de un derecho de acceso a las pruebas presentadas por empresas y proveedores, en los casos en que esté implicada IA de alto riesgo.

Esta regulación es fundamental para garantizar una IA fiable, y conforme con las garantías que exige un Estado de Derecho.

III.4- ACUERDO POLÍTICO ALCANZADO POR LOS CO-LEGISLADORES SOBRE EL REGLAMENTO DE LA IA

El 8 de diciembre de 2023, tras meses de intensas negociaciones entre las instituciones europeas (Comisión, Consejo y Parlamento), se logró un acuerdo político histórico en la regulación de la inteligencia artificial (IA). El acuerdo político sitúa a la UE como líder en la carrera por regular la IA.[38]

El acuerdo pretendía garantizar la protección de los derechos fundamentales, la democracia, el Estado de Derecho y la sostenibilidad medioambiental frente a los riesgos de la IA, así como impulsar la innovación en esta tecnología en la UE.

El texto del acuerdo manifestaba la tensión entre innovación y regulación. La rapidez del desarrollo y evolución de la inteligencia artificial dificulta su regulación. Además, junto a los problemas que genera la complejidad de la materia objeto de regulación, la inteligencia artificial, el reto es que la normativa sea un marco que goce de cierta anticipación en relación a posibles evoluciones tecnológicas.

38 chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.consilium.europa.eu/es/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/pdf

La regulación que se proponía se basaba en un sistema basado en el riesgo. Las obligaciones de los distintos agentes se fijaban en función de su capacidad para causar daños: cuanto mayor fuera el riesgo, más estrictas serían sus obligaciones y mayor la supervisión. El problema que plantea una excesiva regulación es que puede frenar el avance tecnológico y hacer menos competitiva la industria europea.

Se podían distinguir tres bloques:

1.- Los sistemas de IA de riesgo inaceptable, que resultaban prohibidos porque suponían una amenaza para los derechos fundamentales[39].

2.- Los sistemas de IA de "alto riesgo" que se sometían a obligaciones que deberían cumplir antes y después de su comercialización, como la implementación de sistemas de mitigación de riesgos y registro de la actividad, requisitos de alta calidad de los conjuntos de datos, elaboración de documentación detallada, supervisión humana, requisitos de precisión y ciberseguridad de los sistemas[40].

3.- Los sistemas de IA que solo presentaban un "riesgo limitado" estarían principalmente sujetos a obligaciones de transparencia (por ejemplo, revelar que el contenido fue generado por IA para que los usuarios pudieran tomar decisiones informadas sobre su uso posterior).

Además, se añadieron nuevas disposiciones para tener en cuenta aquellas situaciones en las que los sistemas de IA pudieran utilizarse con muchos fines diferentes (IA de uso general) y

39 Ejemplos, la manipulación cognitiva del comportamiento, la extracción no selectiva de imágenes faciales de Internet o de grabaciones de CCTV para crear bases de datos de reconocimiento facial, el reconocimiento de emociones en el lugar de trabajo y en instituciones educativas, el social scoring, o la categorización biométrica para inferir datos sensibles (por ejemplo, orientación sexual, creencias religiosas).

40 Ejemplos de estos sistemas de IA de alto riesgo son determinadas infraestructuras críticas; dispositivos médicos; sistemas para determinar el acceso a instituciones educativas o para reclutar personas; o determinados sistemas utilizados en los ámbitos de la aplicación de la ley, el control de fronteras, la administración de justicia y los procesos democráticos, entre otros.

aquellas en las que la tecnología de IA de uso general se integra posteriormente en otro sistema de alto riesgo.

También se acordaron normas específicas para los modelos fundacionales, sistemas de gran magnitud capaces de realizar de manera competente una amplia gama de tareas diferenciadas, como la generación de vídeo, texto e imágenes, la conversión en lenguaje lateral, la informática o la generación de códigos informáticos. El acuerdo establecía que los modelos fundacionales debían cumplir obligaciones específicas en materia de transparencia antes de ser introducidos en el mercado.

Se introdujo un régimen más estricto para los modelos fundacionales de gran impacto. Se refería a modelos fundacionales entrenados con gran cantidad de datos y con una complejidad y capacidades avanzadas y unos resultados muy superiores a la media, que pudieran difundir riesgos sistémicos a lo largo de la cadena de valor.

Una de las cuestiones más debatidas fue la vigilancia biométrica en los espacios públicos. Los miembros del Parlamento defendían la prohibición absoluta del uso de la IA para el control biométrico "en tiempo real", frente a una posición mucho menos restrictiva de los gobiernos que promovían su autorización con fines de seguridad nacional. Finalmente, los sistemas de identificación biométrica en espacios de acceso público se sometieron a autorización judicial previa. Su uso se limitó a las búsquedas de víctimas (secuestro, trata, explotación sexual); a la prevención de una amenaza inesperada de atentado terrorista, o a la localización e identificación de sospechosos de haber cometido determinados delitos.

Un elemento muy importante del acuerdo fueron las sanciones, que se establecen con una cuantía muy importante con el fin de garantizar la efectividad del Reglamento.

El acuerdo aclaraba que la regulación no afecta a las competencias de los Estados miembros en materia de seguridad nacional.

Además, no se iba a aplicar a los sistemas utilizados exclusivamente con fines militares o de defensa, investigación e innovación, ni a las personas que utilicen la IA con fines no profesionales.

IV.- LA APROBACIÓN DEL REGLAMENTO DE IA Y SU GOBERNANZA. 2024

IV.1- PAQUETE DE INNOVACIÓN EN IA PARA APOYAR A LAS EMPRESAS EMERGENTES Y A LAS PYMES EN INTELIGENCIA ARTIFICIAL

En su discurso sobre el estado de la Unión de 2023, la presidenta Von der Leyen anunció una nueva iniciativa para poner los superordenadores de Europa a disposición de empresas emergentes europeas innovadoras en IA para formar sus modelos de IA fiables.

Como primer paso, la Comisión creó en noviembre de 2023 un premio que ofrece apoyo financiero a las empresas emergentes de IA y acceso a la supercomputación.

En enero de 2024, la Comisión puso en marcha el paquete de innovación en materia de IA para apoyar a las empresas emergentes y las pymes en el ámbito de la inteligencia artificial. El paquete incluye varias medidas para apoyar a las empresas emergentes y las pymes europeas en el desarrollo de una IA fiable que respete los valores y las normas de la UE.

Un elemento clave de este paquete es la Comunicación sobre el impulso de las empresas emergentes y la innovación en inteligencia artificial fiable, que establece un marco estratégico de inversión en IA fiable para que la Unión capitalice sus activos, en particular su infraestructura de supercomputación, líder en el mundo, y fomente un ecosistema europeo innovador de IA.

La principal iniciativa histórica de la Comunicación es GenAI4EU cuyo fin es estimular la adopción de la IA generativa en todos los ecosistemas industriales estratégicos clave de la Unión y fomentar el desarrollo de grandes ecosistemas de innovación abierta que permitirán impulsar la colaboración entre las empresas emergentes de IA y los implementadores de IA en la industria y el sector público.

La iniciativa GenAI4EU tiene como objeto apoyar el desarrollo de casos de uso novedosos y aplicaciones emergentes en los 14 ecosistemas industriales de Europa, así como en el sector público. Los ámbitos de aplicación incluyen la robótica, la salud, la biotecnología, la fabricación, la movilidad, el clima y los mundos virtuales.

La Comisión también está creando, con una serie de Estados miembros, dos Consorcios Europeos de Infraestructuras Digitales (EDIC):

1.- La Alianza para las Tecnologías Lingüísticas (ALT-EDIC) tiene por objeto desarrollar una infraestructura europea común en tecnologías lingüísticas para hacer frente a la escasez de datos lingüísticos europeos para el entrenamiento de soluciones de IA, así como defender la diversidad lingüística y la riqueza cultural de Europa. Esta iniciativa apoyará el desarrollo de grandes modelos lingüísticos europeos.

2.- El EDIC «CitiVERSE» aplicará herramientas de IA de última generación para desarrollar y mejorar los gemelos digitales locales para comunidades inteligentes, ayudando a las ciudades a simular y optimizar los procesos, desde la gestión del tráfico hasta la gestión de residuos.

En este marco se plantea una modificación del Reglamento EuroHPC para crear fábricas de IA, un nuevo pilar para las actividades de la Empresa Común de superordenadores de la UE, que incluye:

1.- Adquisición, mejora y funcionamiento de superordenadores dedicados a la IA para permitir el aprendizaje automático rápido y la formación de grandes modelos de IA de uso general (GPAI);

2.- Facilitar el acceso a los superordenadores dedicados a la IA, contribuyendo a ampliar el uso de la IA a un gran número de usuarios públicos y privados, incluidas las empresas emergentes y las pymes;

3.- Ofrecer una ventanilla única para las empresas emergentes y los innovadores, apoyar el ecosistema de empresas emergentes e investigadoras de IA en el desarrollo algorítmico, probar la evaluación y la validación de modelos de IA a gran escala y proporcionar instalaciones de programación adaptadas a los superordenadores y otros servicios facilitadores de la IA;

4.- Permitir el desarrollo de una variedad de aplicaciones emergentes de IA basadas en modelos de IA de finalidad general.

Por último, es necesario tener en cuenta que, según la Comisión, la innovación en IA exige actividades adicionales:

1.- Apoyo financiero de la Comisión a través de Horizonte Europa y el programa Europa Digital dedicado a la IA generativa. Este paquete generará una inversión pública y privada adicional global de alrededor de 4 000 millones EUR hasta 2027.

2.- Iniciativas de acompañamiento para reforzar la reserva generativa de talento en IA de la UE a través de actividades de educación, formación, capacitación y reciclaje profesional.

3.- Seguir fomentando las inversiones públicas y privadas en empresas emergentes y en expansión de IA, en particular mediante el capital riesgo o el apoyo al capital (también a través de nuevas iniciativas del programa acelerador del Consejo Europeo de Innovación e InvestEU.

4.- La aceleración del desarrollo y la implantación de espacios comunes europeos de datos, puestos a disposición de la comunidad de la IA, para quien los datos son un recurso clave para entrenar y mejorar sus modelos.

IV.2- OFICINA EUROPEA DE LA IA

La Comisión Europea ha creado la Oficina Europea de IA, con sede en Bruselas, como parte del paquete de medidas con el que se pretende fomentar el desarrollo y uso de una inteligencia artificial compatible y respetuosa con los valores de la UE[41].

Como se ha indicado en el apartado anterior, en enero de 2024, la Comisión puso en marcha un paquete de innovación en materia de IA para apoyar a las empresas emergentes y las pymes en el desarrollo de una IA fiable que cumpla los valores y normas de la UE. Tanto la iniciativa GenAI4EU, a la que ya se ha hecho referencia, como la oficina de IA formaban parte de este paquete.

El apoyo de la Oficina de IA en relación con la aplicación del Reglamento de IA se manifiesta mediante:

1.- Su contribución a la aplicación coherente del Reglamento de IA en todos los Estados miembros, incluida la creación de órganos consultivos a escala de la UE, facilitando el apoyo y el intercambio de información.

2.- El desarrollo de herramientas, metodologías y puntos de referencia para evaluar las capacidades y el alcance de los modelos de IA de propósito general, y clasificar modelos con riesgos sistémicos**.**

3.- La elaboración de códigos de práctica de última generación para detallar las normas, en cooperación con los principales desarrolladores de IA, la comunidad científica y otros expertos.

41 https://digital-strategy.ec.europa.eu/es/policies/ai-office#ecl-inpage-task

4.- La investigación de posibles infracciones de las normas, incluidas las evaluaciones para verificar las capacidades del modelo, y solicitar a los proveedores que tomen medidas correctoras.

5.- La preparación de orientaciones y directrices, actos delegados y de ejecución, y otras herramientas para apoyar la aplicación efectiva y la supervisión del Reglamento de la IA.

La Oficina Europea de IA constituye el centro del sistema europeo de gobernanza de la IA mediante el desarrollo y coordinación de las políticas europeas en la materia, así como a través de la aplicación y supervisión del cumplimiento del Reglamento de IA. La Oficina Europea de IA se integra en la Dirección General de Redes de Comunicación, Contenido y Tecnologías de la Comisión. Entre sus funciones se incluyen:

- Apoyo al Reglamento de IA y aplicación de normas de IA para fines energéticos.
- Reforzar el desarrollo y el uso de una IA fiable en todo el mercado interior.
- Fomento de la cooperación internacional.
- Cooperación con instituciones, expertos y partes interesadas.

La Oficina Europea de Inteligencia Artificial debe desempeñar sus funciones, en particular para emitir orientaciones, de manera que no duplique las actividades de los órganos y organismos pertinentes de la Unión en virtud de la legislación sectorial específica.

En consecuencia, esta Oficina Europea de Inteligencia Artificial debe operar de conformidad con los procesos internos de la Comisión y su creación no debe afectar a las facultades y competencias de las autoridades y órganos, oficinas y agencias nacionales competentes de la Unión en la supervisión de los

sistemas de IA, tal como se prevé en el Reglamento por el que se establecen normas armonizadas en materia de inteligencia artificial y otra legislación sectorial de la Unión. Por tanto, se le atribuyen competencias sin perjuicio de las funciones de otros servicios de la Comisión en sus respectivos ámbitos de responsabilidad, y del Servicio Europeo de Acción Exterior en el ámbito de la política común, exterior y de seguridad.

IV.3- APROBACIÓN POR EL PARLAMENTO EUROPEO Y POR EL CONSEJO DEL REGLAMENTO DE LA IA. PUBLICACIÓN Y PROCESO DE APLICACIÓN. 2024 –2027

El Reglamento de inteligencia artificial, (conocido como Ley europea de Inteligencia Artificial)[42] fue inicialmente acordado en las negociaciones con los Estados miembros en diciembre de 2023. Fue respaldado por la Eurocámara con 523 votos a favor, 46 en contra y 49 abstenciones. el 13 de marzo de 2024. El Consejo lo aprobó el 21 de mayo de 2024.

Su publicación en el Diario Oficial de la Unión Europea se produjo el 12 de julio de 2024 y entró en vigor el 1 de agosto de 2024. A partir de esa fecha, será plenamente aplicable 24 meses después, excepto para ciertas disposiciones específicas:

- La prohibición de sistemas de IA que planteen riesgos inaceptables deberá ser aplicada 6 meses después de la entrada en vigor del Reglamento.
- Los Códigos de buenas prácticas deberán ser aplicadas 9 meses después de la entrada en vigor del Reglamento.

42 REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 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

- Las Normas de IA de uso general, incluida la gobernanza, que deban cumplir requisitos de transparencia, deberán ser aplicadas 12 meses después de la entrada en vigor del Reglamento.
- Las obligaciones para los sistemas de alto riesgo deberán ser aplicados 36 meses después de la entrada en vigor del Reglamento.

El nuevo Reglamento persigue garantizar la seguridad y el cumplimiento de los derechos fundamentales, al tiempo que impulsa la innovación en el ámbito de la inteligencia artificial. Este marco legal representa un paso significativo para la UE en su objetivo de convertirse en un líder global en IA confiable.

Gracias al Reglamento, las organizaciones ahora tienen directrices claras para la automatización en áreas de bajo riesgo, como chatbots simples, facilitando su implementación escalable con certeza y transparencia. Asimismo, se categorizan los riesgos para guiar el desarrollo de modelos más complejos, conocidos como "caja negra".

El impacto global del Reglamento dependerá de su aplicación en los distintos países, que ya han mostrado divergencias iniciales en su interpretación. Esta división de criterios podría suponer un desafío para las organizaciones que operan a nivel internacional, por lo que es esencial desarrollar estrategias de IA flexibles que puedan adaptarse a medida que se clarifiquen los detalles durante la aplicación. En este sentido es fundamental una preparación anticipada para cumplir con la nueva regulación a la hora de aprovechar oportunidades emergentes en el campo de la IA.

CONCLUSIONES

PRIMERO.- El enfoque de la UE con respecto a la inteligencia artificial se centra en la excelencia y la confianza, con el objetivo de impulsar la investigación y la capacidad industrial, garantizando al mismo tiempo la seguridad y los derechos fundamentales.

SEGUNDO.- La Estrategia Europea de IA tiene por objeto convertir a la UE en un centro de categoría mundial para la IA y garantizar que la IA esté centrada en el ser humano y sea fiable. Este objetivo se traduce en el enfoque europeo de la excelencia y la confianza a través de normas y acciones concretas. En este sentido, fomentar la excelencia en la IA reforzará el potencial de Europa para competir a escala mundial.

TERCERO.- La consecución de estos objetivos exige:

- permitir el desarrollo y la adopción de la IA en la UE;
- que la UE se convierta en el lugar donde la IA prospera desde el laboratorio hasta el mercado;
- la garantía de que la IA funcione para las personas y sea una fuerza para el bien en la sociedad;
- construir un liderazgo estratégico en sectores de alto impacto.

BIBLIOGRAFÍA

- ANYOHA, R., "The History of Artificial Intelligence", *Science in the News, Special Edition on Artificial Intelligence*, Harvard Kenneth C. Griffin Graduate School of Arts and Sciences, 2017 (https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/) (Recuperado el 7 de julio de 2024)
- DELCKER, J., "The man who invented the self-driving car (in 1986)", *Político,* 2018 (https://www.politico.eu/article/delf-driving-car-born-1986-ernst-dickmanns-mercedes/) (Recuperado el 7 de julio de 2024)
- FEIGENBAUM, E. y SHROBE, H., "The Japanese National Fifth Generation Project: Introduction, survey and evaluation", *Future Generation Computer Systems*, 9, 1993, págs. 105-117 (https://stacks.stanford.edu/file/druid:kv359wz9060/kv359wz9060.pdf) (Recuperado el 7 de julio de 2024)
- FRIEL, K., "A Look Back on the Dartmouth Summer Research Project on Artificial Intelligence", *The Darmouth*, 2024 (https://www.thedartmouth.com/article/2023/05/a-look-back-on-the-dartmouth-summer-research-project-on-artificial-intelligence) (Recuperado el 7 de julio de 2024)
- GABRIEL, J. et al, "The Ethics of Advanced AI Assistants". Google DeepMind, 2024 (https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/ethics-of-advanced-ai-assistants/the-ethics-of-advanced-ai-assistants-2024-i.pdf) (Recuperado el 7 de julio de 2024)
- GATES, B., "The Age of AI has begun", *Gates Notes. The blog of Bill Gates,* 2023 (https://www.gatesnotes.com/The-Age-of-AI-Has-Begun) (Recuperado el 7 de julio de 2024)

- KARJAN, R., "The History of Artificial Intelligence: Complete AI timeline", Tech Target. Enterprise AI, 2023 (https://www.techtarget.com/searchenterpriseai/tip/The-history-of-artificial-intelligence-Complete-AI-timeline) (Recuperado el 7 de julio de 2024)
- POLLACK, A., "'Fifth Generation' Became Japan's Lost Generation", *The New York Times*, 1992 (https://www.nytimes.com/1992/06/05/business/fifth-generation-became-japan-s-lost-generation.html) (Recuperado el 7 de julio de 2024)
- RAO, A.N, VERWEIJ, G., "Sizing the prize. PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution", PwC, 2017 (https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf) (Recuperado el 7 de julio de 2024)
- SANZ, M., "¿Qué es Sandbox y en qué consiste?", *Computer Hoy*, 2019 (https://computerhoy.com/reportajes/tecnologia/que-es-sandbox-529177) (Recuperado el 8 de julio de 2024)
- TURING, A. M., "Computing Machinery and intelligence", *Mind* 49, 1950,, págs.. 433-460 (https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf) (Recuperado el 7 de julio de 2024)
- VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, L., POLOSHUKIN, I., "Attention is All You Need", *Neural Information Processing Systems*, 2017 (https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf) (Recuperado el 7 de julio de 2024)
- "Configurar el futuro digital de Europa. Oficina Europea de la IA". Comisión Europea. 2024 (https://digital-strategy.ec.europa.eu/es/policies/ai-office#ecl-inpage-task (Recuperado el 10 de julio de 2024)

- ”Dartmouth Summer Research Project: The Birth of Artificial Intelligence”, History of Data Science, 2021. (https://www.historyofdatascience.com/dartmouth-summer-research-project-the-birth-of-artificial-intelligence/) (Recuperado el 7 de julio de 2024)
- “Generative AI to Become a $1.3 Trillion Market by 2032, Research Finds”, Bloomberg Intelligence, 2023 (https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/) (Recuperado el 7 de julio de 2024)
- “Kismet” *Robots* (https://robotsguide.com/robots/kismet) (Recuperado el 7 de julio de 2024)
- “¿Qué es el aprendizaje automático?” Google Cloud (https://cloud.google.com/learn/what-is-machine-learning?hl=es-419) (Recuperado el 7 de julio de 2024)
- “¿Qué es el aprendizaje profundo?” Microsoft (https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-deep-learning) (Recuperado el 7 de julio de 2024)
- “¿Qué es un sistema experto? Usos y aplicaciones en Inteligencia Artificial?”, *UNIR Revista*. (https://www.unir.net/ingenieria/revista/sistema-experto/) (Recuperado el 7 de julio de 2024)
- “¿Qué son las redes neuronales convolucionales?” IBM (https://www.ibm.com/es-es/topics/convolutional-neural-networks) (Recuperado el 7 de julio de 2024)
- “¿Qué son los grandes modelos de lenguaje (LLM)?”, IBM (https://ibm.com/es-es/topics/large-language-models) (Recuperado el 7 de julio de 2024)

- "Superordenadores: Deep Blue, la máquina que derrotó a Gary Kasparov", *El blog de Caixa Bank*, 2018 (https://blog.caixabank.es/blogcaixabank/superordenadores-deep-blue-la-maquina-que-derroto-a-gary-kasparov/#) (Recuperado el 7 de julio de 2024)
- "Watson, 'Jeopardy!' champion", IBM, *IBM Heritage.* (https://www.ibm.com/history/watson-jeopardy) (Recuperado el 7 de julio de 2024)
- "WhatisthehistoryofArtificialIntelligence",TableaufromSalesforce (https://www.tableau.com/data-insights/ai/history#:~:text=The%20idea%20of%20%E2%80%9Cartificial%20intelligence,moved%20independently%20of%20human%20intervention) (Recuperado el 7 de julio de 2024)
- COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Inteligencia artificial para Europa, 25 de abril de 2018 https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:52018DC0237
- DIRECTIVA DEL PARLAMENTO EUROPEO Y DEL CONSEJO relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA). 28 de septiembre de 2022. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52022PC0496 (Recuperado el 10 de julio de 2024)
- COMUNICACIÓN DE LA COMISIÓN AL PARLAMENTO EUROPEO, AL CONSEJO EUROPEO, AL CONSEJO, AL COMITÉ ECONÓMICO Y SOCIAL EUROPEO Y AL COMITÉ DE LAS REGIONES Plan coordinado sobre la inteligencia artificial, 7 de diciembre de 2019 chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj

/https://eur-lex.europa.eu/resource.html?uri=cellar:22ee84bb-fa04-11e8-a96d-01aa75ed71a1.0022.02/DOC_1&format=PDF.

- REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 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.

OBLIGACIONES DE LOS PROVEEDORES (*)

OBLIGATIONS OF THE PROVIDERS

RICARDO RIVERO ORTEGA
Universidad de Salamanca

(*) Este trabajo se recibió el 3 de junio de 2024 y fue aceptado el 1 de agosto.

REVISTA DE
PRIVACIDAD Y DERECHO DIGITAL

RESUMEN

La regulación europea de la Inteligencia Artificial establece una serie de obligaciones informativas y técnicas para posibilitar el control de los proveedores de sistemas. Ante una normativa tan prolija, se hace necesario definir correctamente su alcance, límites y garantías favorables a la iniciativa e innovación de las empresas, así como precisar su aplicabilidad a los poderes públicos.

PALABRAS CLAVE: *Regulación de la inteligencia artificial, obligaciones de los proveedores de IA, evaluación de conformidad.*

ABSTRACT

The european regulation of AI establish information obligations of providers and techniques to enable its control. Given so much complex regulations, a correct definition of its scope, limits and favorable guarantees for the initiative and innovation of companies is necessary, as well as specifying its applicability to public powers.

KEYWORDS: *AI Regulation; AI providers obligations; conformity assessment procedure.*

SUMARIO

I.- INTRODUCCIÓN

La Inteligencia Artificial ha pasado de ser una quimera de ciencia ficción a convertirse en un producto/servicio cotidiano, fabricado y ofrecido por empresas sujetas a una regulación similar en muchos aspectos a la que se proyecta sobre otros sectores de la industria. Al igual que ocurre con las demás tecnologías, antes de someterlas a leyes, deben ponderarse sus riesgos, posibles beneficios e incluso los efectos trascendentes sobre el futuro del ser humano, incorporando consideraciones éticas y de filosofía jurídica a su tratamiento por el Derecho. De esto se ocupan expertos cuyas obras conviene consultar, pues nos alertan de las implicaciones del enfoque que adoptemos sobre nuestro concepto de persona y el devenir de las sociedades contemporáneas[1].

Europa quiere destacar por su defensa del humanismo y los derechos de las gentes, contraponer su modelo al de China, por ejemplo, poco escrupuloso en su limitación de las herramientas más amenazadoras. Así, todos los documentos previos al reglamento aprobado han puesto énfasis en las garantías. En lo más práctico, la regulación europea de la inteligencia artificial incorpora un "nuevo enfoque" en su pionera legislación de estas tecnologías, en el ánimo de evitar retrocesos o inhibiciones de las empresas que desarrollan aplicaciones llamadas a cambiar nuestras vidas[2].

Entre estas dos líneas -filosofía de protección de los derechos y regulación técnica de nuevo enfoque, con instrumentos de control de riesgos y de calidad- se sitúa el tratamiento de las

1 LLANO ALONSO, Fernando Higinio, *Homo ex machina. Ética de la inteligencia artificial ante el horizonte de la singularidad tecnológica*, Tirant lo Blanch, Valencia, 2024.

2 ÁLVAREZ GARCÍA, Vicente, "La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque", *Revista General de Derecho administrativo*, 63, 2023. FERNÁNDEZ HERNÁNDEZ, Carlos, "El Reglamento de Inteligencia Artificial. Un nuevo marco regulador para una tecnología en continua evolución", *Derecho Digital e Innovación*, 19, 2024.

obligaciones de los proveedores, planteado desde los primeros documentos preparatorios con el fin de proteger a los usuarios y evitar los mayores riesgos de abuso en su utilización. Por ello se ha puesto particular énfasis en la transparencia de los sistemas, su fiabilidad y su explicabilidad, la protección de datos personales y la prevención de los riesgos[3].

Todas estas exigencias se proyectan sobre los sujetos y organizaciones que se lucran o logran sus objetivos creando, comercializando o poniendo a disposición de terceros sistemas de IA; empresas y también organismos públicos afectados por la nueva norma.

A los efectos del Reglamento, se entiende que es proveedor "una persona física o jurídica o autoridad, órgano u organismo de otra índole públicos que desarrolle un sistema de IA o un modelo de IA de uso general o para el que se desarrolle un sistema de IA o un modelo de IA y lo introduzca en el mercado o pongan en servicio el sistema de IA con su propio nombre o marca comercial, previo pago o gratuitamente". La norma nos ofrece un concepto muy amplio que incluye sujetos públicos y privados, en distintos eslabones de una cadena compleja de creación y distribución de los productos/servicios de IA.

Por su amplitud y complejidad, esta definición de proveedor plantea los primeros problemas serios relacionados con el estatuto y las obligaciones, toda vez que al no diferenciar sujetos públicos y privados –particulares y autoridades- la incidencia de los controles, las exigencias de aportación de información o las posibles sanciones podrían, de no excepcionarse, no ser distintas. Algunos procedimientos no parecen fácilmente aplicables a determinados organismos públicos, que sin embargo en muchos casos entrarán en este concepto de los proveedores.

3 MARTÍNEZ ESPÍN, Pascual, "La propuesta de marco regulador de los sistemas de inteligencia artificial en el mercado de la UE", *Revista CESCO de Derecho del consumo*, 46, 2023.

Cuando no puedan ser considerados proveedores, las autoridades públicas serán "responsables del despliegue", definidos como "persona física o jurídica o autoridad, órgano u organismo de otra índole públicos que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional". Y estos sujetos serán afectados también por algunas de las obligaciones previstas en el reglamento, de modo que un círculo de organizaciones mucho más amplio estará sujeto a las previsiones que analizaremos a continuación.

Las siguientes páginas se dedican al análisis del estatuto del proveedor, a sus obligaciones concretas, predicables de sujetos públicos y privados. Comenzaré no obstante recordando algunas cuestiones básicas relativas a la posición jurídica de quien se sujeta a un marco regulatorio, la imposición de obligaciones mediante normas legales, el rango requerido y los límites a su establecimiento derivado de los principios del Estado de Derecho.

La regulación detallada del reglamento nos muestra el cumplimiento parcial de algunas de estas condiciones de imposición de deberes u obligaciones, pero su ubicación en anexos o en códigos de conducta plantea dudas de rango normativo, efectividad y aplicabilidad práctica en algunos casos. Por otro lado, buena parte de las obligaciones de los proveedores se encuentran en otra normativa sectorial de considerable importancia, que es la orientada hacia la protección de datos de carácter personal, muy avanzada en Europa y objeto de reformas recientes que afectan por supuesto a los proveedores de IA.

II.- LA NATURALEZA JURÍDICA DE LAS OBLIGACIONES DERIVADAS DE UN MARCO REGULATORIO

Todas las empresas están sujetas a condicionantes normativos pensados para la protección de los consumidores y la prevención de riesgos. En esto consiste en gran medida el objeto del Derecho administrativo económico[4]. Aunque el principio general de partida en la regulación debe ser la libertad, el derecho a emprender iniciativas económicas se modula en atención a otros intereses generales. Tales limitaciones de la libertad no han captado suficiente atención en los análisis dogmáticos, quizás porque se da por supuesto que existan regulaciones limitadoras y controles[5].

La doctrina administrativista se ha referido a la "imposición de deberes de comportamiento", empleando el concepto "deber", para separar la connotación de reciprocidad u onerosidad propia de la obligación (de signo contractual), cuando es el poder público quien establece exigencias de hacer o no hacer. Aquí se diferencian los deberes impuestos por la Administración y los "deberes normativos fiscalizados por la Administración", cuyo principal problema sería "la extensión concreta de cómo el deber legal tiene que ser cumplido por el administrado"[6].

Es en el ámbito tributario donde más se han desarrollado y analizado estos deberes, traducidos además en exigencias de suministro de datos. Así, el deber de información ha ocupado a la doctrina fiscalista, cuyas aportaciones sobre la extensión y

4 RIVERO ORTEGA, Ricardo, *Derecho administrativo económico*, Marcial Pons, 2022.

5 Los estudiosos del Derecho administrativo en Estados Unidos, en cambio, se dividen entre quienes defienden la necesidad de la regulación (SUNSTEIN y BREYER, por ejemplo) y quienes consideran que esta ha fracasado y llegan a discutir la propia legitimidad de este tipo de intervenciones (COGLIANESE, HAMBURGER). Efectivamente, los excesos regulatorios pueden asfixiar la innovación y producir graves perjuicios económicos.

6 GARCÍA DE ENTERRÍA, Eduardo/FERNÁNDEZ RODRÍGUEZ, Tomás-Ramón, *Curso de Derecho administrativo*, Cívitas, Madrid, 2022.

límites del deber de información nos pueden ayudar al definir las obligaciones de los proveedores de servicios de IA, pues casi todas ellas tienen algún carácter informativo, de transparencia o aportación de datos relevantes. Hasta qué punto una empresa tiene que "desnudarse" delante de la Administración en sus prácticas o técnicas es una cuestión que se puede plantear tanto en el ámbito financiero como en la regulación de riesgos[7].

Por supuesto, un criterio relevante a la hora de ponderar la corrección de este alcance es el principio de proporcionalidad, muy útil como piedra de toque de cualquier intervención limitadora o condicionante de los derechos y libertades, en su triple test de necesidad, adecuación y menor restricción posible.

En todo caso, el establecimiento de obligaciones tiene una serie de implicaciones jurídicas y requiere el cumplimiento de presupuestos propios del Estado constitucional. La exigencia de una norma con rango de Ley o equivalente, que se satisface en este caso de los proveedores de sistemas de IA al incluirse las obligaciones un reglamento europeo. También la necesidad de precisar conforme a un principio de tipicidad la obligación, por razones de seguridad jurídica.

Esto es lo que veremos hacen los anexos del Reglamento europeo de inteligencia artificial y otras herramientas (códigos de conducta y especificaciones técnicas). La tipificación consiste en la descripción pormenorizada del contenido, sobre todo cuando se trata de obligaciones de colaboración o de carácter informativo. En muchos sectores regulados, el punto más difícil de resolver y también el que da lugar a un mayor número de controversias precisamente consiste en la determinación del alcance de los deberes de aportar datos, cuáles, cuándo y por qué motivos[8].

7 PEÑA AMORÓS, Mª del Mar, *El deber de información*, Dykinson, 2020.

8 RIVERO ORTEGA, Ricardo, *El Estado vigilante. Consideraciones jurídicas sobre la función inspectora de la Administración*, Tecnos, 1999.

El principal problema que nos vamos a encontrar en el tratamiento jurídico de las obligaciones impuestas a las empresas en contextos regulatorios es su falta de precisión, el recurso a cláusulas generales y conceptos indeterminados para trasladar deberes a los proveedores de servicios, sin concretar muchas veces en qué consista exactamente la obligación. Así sucede en parte al determinar el alcance de las obligaciones de comunicación y transparencia, asociadas también a la explicabilidad de la IA[9].

Especialmente, la "transparencia interna" para evaluadores y usuarios, nos interesa en este análisis porque forma parte del estatuto de los proveedores en su capítulo de obligaciones. Estas obligaciones afectan a derechos fundamentales, pero será necesario valorar hasta qué punto permiten atenuar los derechos de propiedad y libertad de empresa que también reconoce nuestro Ordenamiento[10]. Sobre la mayoría de los productos y servicios que consumimos no tenemos ni necesitamos información muy detallada, ni siquiera cuando se trata de alimentos que ingerimos o medicinas que se introducen en nuestro organismo, así que en la práctica más que reforzar las obligaciones informativas sobre los usuarios, parecen tener pleno sentido las obligaciones de aportación de datos precisos a los reguladores.

El despliegue del reglamento europeo de IA planteará estas cuestiones y otras, así por ejemplo la pretensión de concretar las obligaciones mediante códigos de buenas prácticas. Inmediatamente surge la pregunta de si pueden establecerse obligaciones jurídicas mediante códigos de buenas prácticas, porque el Reglamento contempla esta herramienta de autorregulación, aunque los códigos de buenas prácticas no pueden considerarse

9 COTINO, Lorenzo, "Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida", Revista Española de Transparencia, 16, 2023.

10 PEÑA AMORÓS, Mª del Mar, "Derechos fundamentales y deber de información", Gaceta Fiscal, 2024.

auténticas normas[11]. Y, sin embargo, el Considerando 116 de la Directiva así lo apunta: "Los códigos de buenas prácticas deben abarcar las obligaciones de los proveedores de modelos de IA de uso general y de modelos de uso general que presenten riesgos sistémicos".

Los procesos de certificación y las evaluaciones de conformidad devienen también claves en el cumplimiento de estas obligaciones. De esta forma, el instrumental de seguimiento de las obligaciones es distinto al más convencional del Derecho administrativo clásico, aunque no sea tan innovador si nos fijamos en la regulación de la seguridad de los productos industriales o en otros ámbitos donde impera la regulación por sujetos privados y la autoregulación[12].

III.- LAS OBLIGACIONES CONCRETAS: ARTÍCULOS 16, 50, 53 Y 55 DEL REGLAMENTO EUROPEO

Ya he tratado el concepto de "proveedor", así como el de "responsable del despliegue" en la introducción de este trabajo. Ahora diferenciaré el régimen de las obligaciones tal y como lo hace el reglamento, en función de los niveles de riesgo asociados a los sistemas, comenzando por los de "alto riesgo", definidos en primer lugar, continuado con los de "riesgo sistémico" y terminando con los de "uso general". A mayor nivel de riesgo, más y más intensas obligaciones, pues al fin estas sirven para permitir un grado de control mayor sobre los proveedores o sujetos responsables de los sistemas de IA.

11 SADDY, André, "Códigos de buenas prácticas. Concepto, naturaleza y su configuración como fuente de Derecho administrativo", en *Regulación y competencia en servicios de interés económico genera*l, 2017.

12 DARCANULLETA GARDELLA, Mercé, Autoregulación y Derecho público: la autoregulación regulada, Marcial Pons, 2005.

Fuera de la regulación específica de cada tipo de proveedor, el reglamento contempla también obligaciones comunes y genéricas. Así, su artículo 4 les conmina a adoptar medidas de "alfabetización en materia de IA" orientadas a su personal. Así mismo, las "prácticas de inteligencia artificial prohibidas", señaladas en el artículo 5, comportan obligaciones negativas de los proveedores (no hacer). Estas prohibiciones no son triviales porque condicionan sobremanera el desarrollo técnico de los sistemas.

El primer artículo del Reglamento europeo dedicado a las obligaciones de los proveedores es el 16, dirigido a los proveedores y responsables del despliegue de sistemas de IA de alto riesgo. El Reglamento identifica los sistemas de "alto riesgo" en su artículo 6, en un modo sumamente esclarecedor si nos fijamos en la definición en negativo (lo que no son sistemas de alto riesgo), en aquellos casos que "…no plantea un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas…", así como en varios supuestos que este precepto detalla.

La primer obligación de la lista es velar por el cumplimiento de los requisitos de la sección 2 (sistema de gestión de riesgos; sometimiento a pruebas previas para determinar medidas de gestión de riesgos; prácticas de gobernanza y gestión de datos; detección y prevención de sesgos; elaboración y actualización de documentación técnica; conservación de registros; transparencia y comunicación de información a los responsables del despliegue; posibilidades de vigilancia humana mediante el desarrollo de interfaces; precisión, solidez y ciberseguridad).

Esta obligación es en realidad una "metaobligación", es decir, una obligación sobre el cumplimiento de obligaciones, refuerzo o recordatorio de los requisitos que deben cumplirse, varios de los cuales se reiteran en las siguientes obligaciones.

La siguiente obligación es indicar en el sistema o su embalaje que se trata de un sistema de alto riesgo, su nombre comercial

y dirección de contacto. Este tipo de advertencias sirven para dejar constancia del peligro propio del producto/servicio y favorecer la trazabilidad, una de las técnicas básicas de control de riesgos (con muchas aplicaciones ensayadas en el Derecho alimentario). La cuestión del etiquetado no está del todo resuelta en el Reglamento. Por eso se ha propuesto una etiqueta complementaria que advertiría de los riesgos y ofrecería más información a los usuarios[13].

La tercera obligación es cumplir con un sistema de gestión de calidad, cuyas características son detalladas por el artículo 17: estrategia de cumplimiento de la normativa (incluyendo evaluación de conformidad y gestión de las modificaciones); técnicas, procedimientos y actuaciones sistemáticas para el control y verificación del diseño; técnicas, procedimientos y actuaciones sistemáticas para el aseguramiento de la calidad; procedimientos de examen, prueba y validación antes, durante y después, con su frecuencia; especificaciones técnicas; sistemas y procedimientos de gestión de datos; sistema de gestión de riesgos; sistema de vigilancia poscomercialización; procedimientos de notificación de un incidente grave; gestión de la comunicación con las autoridades nacionales; procedimientos para llevar registro de toda la documentación; gestión de recursos y seguridad del suministro; marco de rendición de cuentas que defina las responsabilidades del personal directivo.

El deber de conservación de la documentación prevista en el artículo 18 incluye esta primera remisión y otras a los artículos 11 (sobre documentación técnica), y 17 (sistema de gestión de calidad). Además de estos documentos, deben conservarse los relativos a los cambios aprobados por los organismos, las decisiones de los organismos y la declaración UE de conformidad. Serán los Estados miembros los que establezcan los plazos de conservación de la documentación y sus condiciones.

13 STUURMAN, kees7LACHAUD, Eric, "Regulating IA. A Label to complete the proposed Act of Artificial Intelligence", *Computer Law and Security Review*, 2022.

Los proveedores también han de conservar los archivos de registros generados automáticamente por los sistemas. Esos registros dejan constancia de su operatividad y permiten trazar incidencias y comprobar si se están controlando los riesgos y cumpliendo las condiciones de seguridad, calidad y protección de datos. La regulación de esos registros se encuentra en el artículo 19 con fines de trazabilidad al fin.

La siguiente obligación prevista es la de someterse periódicamente a la evaluación de conformidad, un requisito que después analizaremos con más detalle en su régimen del artículo 43. Este tipo de certificación es una de las claves del modelo de regulación y control por el que ha optado la Unión Europea, en línea con otras modalidades de garantía de seguridad de otros muchos productos.

También están obligados los proveedores a elaborar una declaración UE de conformidad, cuyas condiciones se regulan en el artículo 47. Más adelante nos detendremos en la naturaleza de esta declaración, su vínculo al control de la seguridad y la calidad de los productos industriales y a otros ámbitos como la sanidad. Básicamente se tratan de conjuntos de información no financiera que demuestran el cumplimiento de la normativa del sector de referencia, técnicas propias de la certificación y normalización[14].

Otra obligación es colocar el marcado CE de IA de alto riesgo en el sistema o, si no fuera posible, en su embalaje o la documentación. La insistencia en que se puedan identificar los sistemas de alto riesgo es comprensible, al igual que en otros productos se advierte de su peligrosidad a quienes accedan a su uso. Tales advertencias serán relevantes también en el momento de imputar responsabilidades si se producen daños por una utilización inadecuada de los sistemas de alto riesgo.

14 BRITO MARQUINA, Avelino, "Verificaciones, la última frontera de la certificación", *Calidad. Revista mensual de la Asociación Española para la Calidad*, 1, 2020.

También han de cumplir las obligaciones de registro previstas en el artículo 49.1 del Reglamento. Deben adoptar medidas correctoras necesarias y facilitar la información prevista en el artículo 20 (retirar del mercado o desactivar sistemas que no cumplan el Reglamento, investigar las causas de riesgos e informar a las autoridades de vigilancia del mercado y a los organismos notificados. Habrán de demostrar de forma motivada ante la autoridad nacional competente la conformidad del sistema de alto riesgo con sus requisitos y velarán por que el sistema cumpla todos los requisitos de accesibilidad.

El artículo 17 añade a todas estas obligaciones la de establecer un sistema de gestión de calidad, consignado de manera sistemática y ordenada en la documentación con todos los aspectos de cumplimiento de la normativa. Este sistema debe hacer referencia a las especificaciones técnicas que se cumplirán y los procedimientos de gestión de datos, los sistemas de vigilancia postcomercialización y los procedimientos de notificación de un incidente grave, entre otras garantías para el cumplimiento de las obligaciones, moduladas en función del tamaño del proveedor.

Mención aparte merece la letra m) del apartado 1 de este artículo 17, pues exige "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". La identificación de las responsabilidades es una obligación trascendente por sus posibles consecuencias ulteriores, en línea con la tendencia actual de despliegue de sistemas de cumplimiento y asignación de funciones concretas a instancias determinadas en el organigrama de las empresas.

Otra obligación regulada en el artículo 18 es la conservación de la documentación durante un plazo de diez años desde la introducción en el mercado o la puesta en servicio del sistema de IA. Así mismo, deben conservar los archivos de registro generados automáticamente (artículo 19) y adoptar medidas correctoras sobre las que deben informar (artículo 20), cooperar

con las autoridades competentes, dando acceso a información y archivos (artículo 21) e informar a las autoridades sobre sus representantes autorizados (artículo 22).

Otros artículos que enuncian obligaciones de los proveedores se ubican en el Capítulo IV del Reglamento, sobre obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA. El artículo 50, primero de esta serie se refiere a los sistemas de IA destinados a interactuar directamente con personas físicas (obligación de informar de que están interactuando con un sistema de IA), a los sistemas que generen contenido sintético de audio, imagen, vídeo o texto (que sea posible detectar que ha sido generado o manipulado de manera artificial), a los sistemas de reconocimiento de emociones o categorizaciones biométricas (información a las personas expuestas y protección de datos). En fin, se establecen obligaciones de información para que las personas puedan identificar la IA y diferencien sus productos y servicios.

El siguiente precepto en este capítulo dedicado a obligaciones de los proveedores es el 53, que se proyecta sobre los modelos de IA de uso general. La elaboración y mantenimiento de la documentación técnica del modelo, su inteligibilidad para otros proveedores, el establecimiento de directrices para respetar los derechos de autor, puesta a disposición del contenido utilizado para el entrenamiento del modelo, y la cooperación con la Comisión y las autoridades nacionales competentes. El recurso a códigos de buenas prácticas para demostrar el cumplimiento de las obligaciones se prevé en el apartado cuarto de este precepto, demostrando la relevancia de estas pseudofuentes normativas en la regulación de la IA.

El artículo 55 del Reglamento regula las obligaciones de los proveedores de modelos de IA de uso general con riesgo sistémico, un concepto definido en el apartado 65 del artículo 3: "un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general, que tienen unas repercusiones considerables

en el mercado de la Unión debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor".

Esta lista coincide en gran parte con la anterior. Evaluación de los modelos de conformidad para reducir riesgo sistémico, evaluación y reducción de los riesgos sistémicos: vigilancia y comunicación sin demora de los incidentes graves; comunicarán las medidas correctoras para resolver esos incidentes y velarán para que se establezca un nivel adecuado de protección de la ciberseguridad.

La prueba del cumplimiento de estas obligaciones puede realizarse mediante la adhesión a códigos de buenas prácticas previstos en el artículo 56, hasta que se publique una norma armonizada. De no optar por esta forma de demostración, tendrán que seguir otro medio adecuado aprobado por la Comisión. Todo parece indicar que los códigos de buenas prácticas funcionarán por tanto con un carácter pseudonormativo complementario. No son obligatorios, pero si no se cumplen será difícil demostrar que no se ha incurrido en incumplimiento de obligaciones porque el sistema de control funciona de tal forma que es el proveedor el que tiene que demostrar que cumple (inversión de la carga de la prueba) el cumplimiento de sus obligaciones, no la autoridad de control la que debe presentar pruebas del incumplimiento[15].

Es el artículo 51 del Reglamento el que nos ofrece las reglas de clasificación de los modelos de IA con riesgo sistémico, que han de cumplir alguno de estos requisitos: capacidades de

15 Recientemente he analizado la inversión de la carga de la prueba que se observa en la práctica del Derecho administrativo sancionador, en mi artículo publicado en la REDA de 2024, RIVERO ORTEGA, Ricardo, "¿Presuntos inocentes o presuntos culpables? La prueba de la responsabilidad subjetiva en el Derecho administrativo sancionador", *Revista Española de Derecho administrativo*, 2024.

gran impacto (alertadas por grupos de expertos científicos); cantidades acumuladas de cálculo medidas en FLOP superiores a 10^25.

La identificación del riesgo sistémico se precisa en el Anexo VIII del Reglamento, que establece los criterios para la clasificación de los modelos de uso general con este nivel de riesgo en función del número de parámetros del modelo, la calidad o el tamaño del conjunto de datos, la cantidad de cálculo utilizada para entrenar el modelo, sus modalidades de entrada y salida, las capacidades del modelo, sus repercusiones sobre el mercado interior, el número de usuarios finales registrados, etc.

IV.- LA EVALUACIÓN DE CONFORMIDAD

La comercialización de productos y servicios en la Unión Europea está sujeta a unas normas comunes que incluyen un régimen de evaluación y declaración de conformidad, contenido en el Reglamento de 9 de julio de 2008 por el que se establecen los requisitos de acreditación y vigilancia del mercado relativos a la comercialización de los productos. Este reglamento establece la organización y funcionamiento para la acreditación de los organismos de evaluación de conformidad[16].

En esta norma también encontramos la definición de conceptos y herramientas que utiliza el nuevo Reglamento europeo. Así, por ejemplo, "especificación técnica", "norma armonizada", "acreditación" o "autoridad de vigilancia del mercado", entre otros. También se enuncian los principios generales de la acreditación,

16 ÁLVAREZ GARCÍA, Vicente, *Industri*a, Iustel, 2010. Más recientemente, del mismo autor, "Los instrumentos normativos reguladores de las especificaciones técnicas en la Unión Europea: un breve ensayo de identificación de nuevas fuentes del Derecho", *Revista General de Derecho Administrativo*, 63, 2023.

su funcionamiento, la acreditación transfronteriza, los requisitos de los organismos nacionales de acreditación, las obligaciones de informar y las medidas de vigilancia del mercado.

El Reglamento de IA proyecta este sistema sobre los proveedores de sistemas IA, dedicando varios considerandos a su aplicación, especialmente para los de alto riesgo, que además pueden formar parte de otros productos. Ya el Derecho europeo contempla la exigencia de la evaluación de conformidad para los productos de riesgo medio y alto.

Entre las peculiaridades de los sistemas de IA a los efectos del procedimiento de evaluación de conformidad, el Considerando 78 señala la necesidad de aplicar los requisitos esenciales de ciberseguridad de los productos digitales ("productos críticos importantes con elementos digitales"). Para ello se señala la cooperación con ENISA (Agencia de la Unión Europea para la Ciberseguridad).

Esta obligación de realizar la evaluación de conformidad está también vinculada a la de establecer un sistema de gestión de calidad y cumplir con todos los controles. Al fin, el régimen que se está proyectando sobre los sistemas de IA es el propio de la seguridad de los productos, contenido en normas transversales pensadas para que el poder público, auxiliado por entidades privadas especializadas, pueda evitar los riesgos asociados a la puesta en el mercado de determinados servicios o productos que pueden dañar a las personas o los bienes.

En el Reglamento de IA es el artículo 43 el que se dedica a la evaluación de la conformidad, a la que se deben someter los proveedores cuando apliquen las normas armonizadas (artículo 40) o las especificaciones comunes (artículo 41). Se presentan como alternativos dos procedimientos de evaluación de conformidad, respectivamente desarrollados en los anexos VI y VII. El primer está basado en el control interno, mientras el segundo se basa en la evaluación del sistema de gestión de calidad y la

evaluación de la documentación técnica, con la participación de un organismo notificado que se menciona en el anexo VII.

La sujeción al procedimiento de evaluación de conformidad basado en el sistema de gestión de calidad es obligatoria en aquellos casos en los cuales no existan normas armonizadas ni especificaciones comunes, el proveedor no haya aplicado toda la norma armonizada, no se hayan aplicado las especificaciones comunes o las normas armonizadas se hubieran aplicado con una limitación. Los sistemas de alto riesgo de los puntos 2 a 8 del Anexo III se deben atener al procedimiento de evaluación de conformidad fundamentado en control interno, sin participación de un organismo notificado. También se contemplan particularidades para los sistemas de IA de alto riesgo regulados por actos legislativos de armonización de la Unión.

Una vez un sistema ha superado un procedimiento de evaluación de conformidad, sólo tiene que someterse a un nuevo procedimiento cuando sea objeto de una "modificación sustancial". Esta previsión del apartado 4 del artículo es complementada con la precisión sobre sistemas que evolucionan ("continúen aprendiendo") a lo largo del tiempo. Si este progreso del sistema ha sido predeterminado por el proveedor en el primer procedimiento de evaluación de conformidad, no es necesario que pasen por un nuevo procedimiento porque no se consideran modificaciones sustanciales.

Los procedimientos de evaluación de conformidad dan lugar a la emisión de certificados (artículo 44) que pueden tener una validez máxima de cinco años, aunque pueden suspenderse antes si un sistema deja de cumplir los requisitos por su evolución.

El artículo 46 contempla supuestos de exención del procedimiento de evaluación de conformidad, una decisión que ha de responder a una solicitud debidamente motivada de una autoridad de vigilancia del mercado, siempre que se den circunstancias excepcionales de seguridad pública, protección de la vida

y la salud de las personas o el medio ambiente. También cabe esta exención por razones excepcionales de seguridad pública ante una amenaza específica.

La Declaración UE de conformidad es regulada en el artículo 47 y establece otra obligación de los proveedores. Redactar en un formato legible por máquina y firmado una por cada sistema de alto riesgo, poniéndola a disposición de las autoridades nacionales competentes por un período de diez años. Su contenido será afirmar que se cumplen los requisitos establecidos. Al fin estamos ante una suerte de "declaración responsable", similar a la que se prevén en otros muchos sectores de intervención administrativa.

El Anexo IV del Reglamento, detalla el procedimiento de evaluación de conformidad fundamentado en un control interno. El Anexo V explica la información que debe contener la declaración UE de conformidad: nombre y tipo del sistema de IA, nombre y dirección del proveedor o su representante, afirmación de la responsabilidad exclusiva del proveedor, declaración de conformidad con la normativa, declaración de ajuste a la normativa de protección de datos. El Anexo VII establece la secuencia de la conformidad fundamentada en la evaluación de un sistema de gestión de calidad y la evaluación de la documentación técnica.

Todos estos procedimientos son pues diseñados en su secuencia y cuentan con la experiencia de su aplicación en otros sectores de ingreso en el mercado de productos y servicios, así que sólo queda decir que representan una considerable oportunidad de negocio para las entidades especializadas en garantizar que se cumplen las especificaciones técnicas y el resto de las normas que componen un nuevo sistema de fuentes del Derecho, tal y como lo ha descrito Vicente Álvarez García[17].

17 ÁLVAREZ GARCÍA, Vicente, "Los instrumentos normativos reguladores de las especificaciones técnicas en la unión Europea…", *Revista General de Derecho administrativo*, cit.

La sujeción de los proveedores a mecanismos de evaluación de conformidad, en el caso de los de mayor riesgo, es una buena alternativa a la intensificación de controles públicos que podrían ser más gravosos. Ahora bien, esos controles van a tener lugar, y en un sector de tan rápido avance tecnológico pueden colisionar con las necesidades de las empresas y su capacidad de responder a las demandas de información y datos de las autoridades de vigilancia del mercado, así que puede haber conflictos y litigiosidad derivada de esta regulación.

V.- LA ACTUALIZACIÓN PROGRESIVA DE LAS OBLIGACIONES DE LOS PROVEEDORES Y OTRAS POSIBLES FUENTES DE LITIGIOSIDAD

Así, la seguridad jurídica suele asociarse a un conjunto de normas dado, previsible, cierto en su aplicación, presupuestos que no concurren en este régimen porque las especificaciones técnico, los códigos de conducta, los anexos del reglamento e incluso la interpretación de sus contenidos puede variar a lo largo del tiempo. Las evoluciones de las tecnologías y la comprensión de sus efectos sobre los derechos cambian efectivamente y esto puede suponer que lo que un día se considera suficiente desde el punto de vista del control y la aportación de datos deje de serlo tiempo después, con sobresalientes consecuencias.

Hay que destacar que el apartado 3 del artículo 44, sobre certificados, permite que se suspenda o retire el certificado de un sistema por incumplimiento sobrevenido de los requisitos de su concesión. Aunque todas las decisiones sobre certificados se puedan recurrir y deban tomarse respetando el principio de proporcionalidad, lo cierto es que esta posibilidad genera una cierta inseguridad en los proveedores.

Revista de Privacidad y Derecho Digital
ISSN digital 2444-5762, NÚM. 34, Madrid, Mayo-Agosto (2024), págs. 90-120

Sin duda, el concepto "estado de la técnica" es relevante en las regulaciones dirigidas al control del riesgo[18]. Las normas en estos sectores recurren a expresiones como "la mejor tecnología disponible", concepto jurídico indeterminado útil para establecer una "cláusula de progreso", proyectada en este caso sobre operadores privados (o públicos) que realizan actividades marcadas por la progresiva innovación. Los requerimientos de actualización al avance tecnológico se pueden canalizar a través de los actos delegados a la Comisión en el artículo 97, que puede actualizar los anexos VI y VII del Reglamento "a la luz del progreso técnico".

La normativa europea también contempla el proceso continuo de verificación y las técnicas concretas para garantizar la rendición de cuentas de los proveedores de servicios de IA[19]. La imputación de responsabilidades en marcos regulatorios de prevención del riesgo comporta esta constante actualización, de tal modo que la imputabilidad de daños, los deberes de diligencia y la culpabilidad en las infracciones administrativas dependerán de una ponderación que evoluciona según avanzan los conocimientos técnicos, las percepciones sociales y la experiencia[20].

Además de esta cuestión del cambio progresivo de los requisitos a los proveedores, otro punto de posible conflicto se deriva de las numerosas exigencias de aportación de datos e información, que pueden situar a las empresas ante el riesgo de perder sus ventajas competitivas. En la normativa europea se contemplan garantías compensatorias o "contradeberes", destacadamente los deberes de confidencialidad y reserva de los evaluadores, para garantizar la protección del secreto industrial. También se

18 ESTEVE PARDO, José, "La regulación de riesgos: gestionar la incertidumbre", *El Cronista del Estado Social y Democrático de Derecho*, 2021.

19 Desai, Deven R/Kroll, Joshua, "Trust but Verify: A Guide to Algoithms and the Law", *Harvard Journal of Law and Tecnology*, 1, 2017.

20 HOFMANN, Herwig, "The Duty of Care in EU Public Law – A Principle Between Discretion and Proportionality", *Review of European Administrative Law*, 13, 2020.

contemplan mínimas previsiones frente a los abusos en la exigencia de información.

Así, el artículo 20.3 del reglamento establece que: "Toda información obtenida por una autoridad nacional competente con arreglo al presente artículo se tratará de conformidad con las obligaciones de confidencialidad establecidas en el artículo 78". El artículo 78 desarrolla ese deber de confidencialidad de la Comisión, las autoridades de vigilancia del mercado, los organismos notificados y cualquier persona física o jurídica que participe en la aplicación del reglamento, obligadas a proteger los derechos de propiedad intelectual e industrial, o los secretos comerciales.

En esta línea también se manifiesta el apartado 7 del artículo 53, sobre obligaciones de proveedores de modelos de uso general: "Toda información o documentación obtenida en virtud del presente artículo, incluidos los secretos comerciales, se tratarán de conformidad con las obligaciones de confidencialidad establecidas en el artículo 78".

Ya hemos citado el apartado 3 del artículo 78, conforme al cual sólo se puede solicitar la información estrictamente necesaria para proteger la seguridad y la confidencialidad de la información, y una vez recopilados los datos, se suprimirán todos aquellos que no sean necesarios para los fines legítimos de prevención del riesgo (apartado 2 del artículo 78).

Finalmente, una posible fuente de litigiosidad y cierta rebaja de garantías puede derivarse de la inversión de la carga de la prueba a la hora de determinar la culpabilidad por el incumplimiento de obligaciones de precaución de daños. El reglamento traslada a los proveedores la obligación de demostrar, a través de la evaluación de conformidad o con sus sistemas internos de calidad, que han cumplido todos los requisitos, pero incluso si así se realiza, cuando se produzcan incidentes imprevistos o daños concretos no siempre será fácil determinar si obedecerán a errores

del proveedor o al cumplimiento de especificaciones técnicas o códigos de conducta insuficientes.

VI.- ¿SERÁ LA UNIÓN EUROPEA UN ÁMBITO MENOS PROPICIO A LA INNOVACIÓN DEBIDO A ESTE MARCO REGULADOR?

Los primeros pasos de una nueva tecnología siempre suscitan incertidumbres, incluidas las regulatorias. En torno a la regulación, su necesidad, ventajas y posibles efectos adversos existe una controversia más aguda en los Estados unidos que en Europa. Los juristas americanos se dividen entre partidarios y detractores de las intervenciones regulatorias, con opiniones muy críticas. En la Unión Europea, en cambio, parece haber consenso sobre la conveniencia y necesidad de la regulación.

Entre las distintas opciones regulatorias posibles sobre la inteligencia artificial, la Unión Europea ha aprobado una que establece intensas exigencias y multiplica los controles sobre los proveedores. Además, lo ha hecho adelantándose a otros reguladores nacionales y supranacionales, de tal modo que convierte su experiencia en un modelo sujeto a la prueba de la reacción de los operadores. Ciertamente, parece que con el reglamento se prevendrán los riesgos, pero también se va a convertir la Unión Europea en una suerte de campo de pruebas regulatoria, pudiendo avanzar en el diseño de técnicas y mecanismos de control por delante de otros países. La experiencia puede ser un éxito o generar problemas (quizás ambas cosas).

Para alcanzar el éxito deseado de la estrategia, el Capítulo VI, sobre medidas de apoyo a la innovación, es desde mi punto de vista una muestra del afán de la Unión europea por evitar que su regulación y controles asusten a las empresas. Así, se prevén espacios controlados de pruebas para la IA (artículos 57 y 58), se

regula el tratamiento de datos personales para el desarrollo de sistemas de IA (artículo 59), se prevén pruebas de sistemas de IA de alto riesgo en condiciones reales fuera de espacios controlados de pruebas para la IA (artículo 60), se regula el consentimiento informado para participar en pruebas en condiciones reales fuera de los espacios controlados de pruebas para la IA (artículo 61), se favorecen las empresas emergentes (artículo 62) y se contemplan excepciones para proveedores específicos (artículo 63).

VII.- LA APLICACIÓN A LOS PODERES PÚBLICOS PROVEEDORES O RESPONSABLES DEL DESPLIEGUE

Hasta ahora hemos analizado el régimen de obligaciones de los proveedores en el Reglamento. El concepto de proveedor no diferencia entre empresas privadas y autoridades públicas, lo cual tiene su lógica si de lo que se trata es de prevenir los riesgos y daños derivados del uso de la inteligencia artificial. Ahora bien, el modelo regulatorio por el que se opta responde a los parámetros clásicos de una relación jurídica entre un organismo de supervisión público y sujetos particulares, no tan sencillo de proyectar cuando también se trata de controlar organismos del Estado.

Debe tenerse presente, además, que los usos públicos de la inteligencia artificial suelen estar vinculados al ejercicio de potestades administrativas, algunas de las cuales sirven a la seguridad pública o la persecución de los delitos, tareas ambas de difícil compatibilidad con los principios de transparencia o explicabilidad llevados a su extremo. Así pues, algunos de estos usos no pueden reconducirse a mi juicio a los mismos controles que los aplicables a las herramientas desplegadas por las empresas privadas. Por ello se contemplan excepciones, aunque no del todo afianzadas en la literalidad de la norma.

Un ejemplo de la peculiaridad del uso de los sistemas de IA con fines públicos lo encontramos en la previsión del artículo 27 del Reglamento, que exige a los responsables del despliegue de sistemas de alto riesgo una evaluación de impacto a los derechos fundamentales. Este precepto se orienta hacia los responsables del despliegue que sean organismos de Derecho público o entidades privadas que presten servicios públicos.

El artículo 43 del Reglamento, al regular la evaluación de conformidad, contiene al final de su primer apartado un inciso que reza lo siguiente: "...cuando se prevea la puesta en servicio del sistema de IA de alto riesgo por parte de las autoridades encargadas de la aplicación de la ley, las autoridades de inmigración o las autoridades de asilo, o por las instituciones, órganos u organismos de la Unión, la autoridad de vigilancia del mercado mencionada en el artículo 74, apartado 8 o 9, según proceda, actuará como organismo notificado". Esto significa que el control de este procedimiento corresponde directamente a la autoridad pública, lo cual parece lógico. Un control privado sobre autoridades u organismos públicos no parece muy adecuado.

Las exenciones del procedimiento de evaluación de conformidad, previstas en el artículo 46, pueden aplicárseles también a los organismos públicos, toda vez que muchas de sus aplicaciones pueden tener implicaciones sobre la seguridad, la salud o la protección del medio ambiente.

La evaluación de conformidad y la sujeción a una normativa diseñada para empresas, proyectada sobre organismos públicos prestadores de servicios públicos, puede dar lugar a complejidades varias que no podemos tratar en esta breve contribución. El ámbito de aplicación y las excepciones del Reglamento no han sido demasiado precisos en su definición diferenciada de proveedores privados y públicos. Ahora sólo apuntamos que esta falta de separación puede ser fuente de dificultades interpretativas y aplicativas.

VIII.- CONCLUSIONES

PRIMERA: La regulación europea de la inteligencia artificial aspira a orientar en clave humanista y de respeto de los derechos mediante un sistema de controles y un nuevo enfoque. Si se logra un equilibrio entre las garantías, los controles y los márgenes necesarios para la innovación empresarial, esta nueva normativa será exitosa.

SEGUNDA: Es muy importante precisar la naturaleza y alcance de las obligaciones de los proveedores, así como reconocer la importancia de nuevas herramientas jurídicas para constatar su cumplimiento (códigos de buenas prácticas, especificaciones técnicas y evaluaciones de conformidad). Una adecuada comprensión de sus efectos sobre el estatuto de los proveedores será presupuesto de la seguridad jurídica en el sector de la IA europea.

TERCERA: Los proveedores de sistemas de alto riesgo están sujetos a más obligaciones que todos los demás, por razones bien comprensibles. Estas obligaciones tienen un carácter informativo e incluyen el despliegue de sistemas de calidad, prevención y trazabilidad de riesgos y evaluación de conformidad.

CUARTA: Los proveedores de IA de riesgo sistémico también están sujetos a previsiones especiales de control y suministro de información. No se enfrentan a prohibiciones tan estrictas como los de alto riesgo, pero sus deberes de cooperación con las autoridades de vigilancia son intensos. Así mismo, algunas de las garantías propias de los sistemas de alto riesgo se extienden a los de riesgo sistémico.

QUINTA: Los proveedores de IA de uso general están sujetos a las obligaciones propias de muchos fabricantes o suministradores de bienes y servicios regulados por el Derecho europeo. En este caso, el menor nivel de riesgo comporta

menos obligaciones de realización de procedimientos como la evaluación de conformidad y otros controles.

SEXTA: La actualización progresiva de las prevenciones y garantías de trazabilidad e inteligibilidad de los sistemas de IA, mediante la adaptación a la mejor tecnología disponible, suponen obligaciones adicionales y sostenidas en tiempo real. El ritmo acelerado de evolución de estas técnicas permite anticipar algunas dificultades en el cumplimiento de esa obligación, así como posibles controversias en su exigencia.

SÉPTIMA: La proyección de las normas de control, transparencia y suministro de información aplicables a los proveedores de sistemas de IA de alto riesgo a los poderes públicos puede dar lugar a complejidades interpretativas y aplicativas. Las excepciones previstas en el Reglamento no parecen suficientemente explícitas.

OCTAVA: Una regulación más avanzada de la IA en Europa que en otros mercados podría producir sobre la inversión, la investigación y la innovación. Si los requisitos de transparencia, trazabilidad y control son percibidos por los proveedores como desproporcionados, o retrasan sus desarrollos, esto perjudicará la imagen y el funcionamiento del mercado interior europea. En cambio, si los controles se muestran eficientes y adecuados, el modelo regulatorio será copiado por otros países.

NOVENA: El reconocimiento de derechos de secreto industrial y el blindaje de los deberes de confidencialidad son garantías de las empresas frente al riesgo de que el cumplimiento de sus obligaciones produzca filtraciones de su *know how* dañinas para la competencia y los incentivos de innovación. Los reguladores han de ser muy cuidadosos en sus exigencias informativas, adaptándolas a lo verdaderamente indispensable para cumplir sus cometidos.

DÉCIMA: Una cuestión de gran importancia en el régimen de las obligaciones es la relativa a las responsabilidades derivadas de su incumplimiento, especialmente cuando se produzcan daños que puedan ser sancionables e indemnizables. La carga de la prueba del cumplimiento de la obligación se traslada mediante el régimen del Reglamento en gran medida a los proveedores, en una rebaja de garantías que es común a otros sectores. En un ámbito de tecnologías tan complejas, las cuestiones probatorias sobre cumplimiento e incumplimiento de obligaciones pueden dar lugar a serios problemas.

IX.- BIBLIOGRAFÍA

ÁLVAREZ GARCÍA, Vicente, *Industria*, Iustel, 2010.

ÁLVAREZ GARCÍA, Vicente, "La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque", *Revista General de Derecho administrativo*, 63, 2023.

ÁLVARÉZ GARCÍA, Vicente, "Los instrumentos normativos reguladores de las especificaciones técnicas en la Unión Europea: un breve ensayo de identificación de nuevas fuentes del Derecho", *Revista General de Derecho Administrativo*, 63, 2023.

BRITO MARQJINA, Avelino, "Verificaciones, la última frontera de la certificación", *Calidad. Revista mensual de la Asociación Española para la Calidad*, 1, 2020.

COBBE, Jeniffer/SINGH, Jatinder, "Artificial Intelligence as a Service: Legal Responsabilities, Liabilites and Policy Challenges", *Computer Law & Security Review*, 42

COTINO, Lorenzo, "Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida", *Revista Española de Transparencia*, 2023.

DARCANULLETA GARDELLA, Mercé, Autoregulación y Derecho público: la autoregulación regulada, Marcial Pons, 2005.

DESAI, Deven R./KROLL, Joshua, "Trust but Verify: A Guide to Algorihtms and the Law", *Harvard Journal of Law and Technology*, 31, 2017.

ESTEVE PARDO, José, "La regulación de riesgos: gestionar la incertidumbre", *El Cronista del Estado Social y Democrático de Derecho*, 2021.

FERNÁNDEZ HERNÁNDEZ, Carlos, "El Reglamento de Inteligencia Artificial. Un nuevo marco regulador para una tecnología en continua evolución", *Derecho Digital e Innovación*, 19, 2024.

GARCÍA DE ENTERRÍA, Eduardo/FERNÁNDEZ RODRIGUEZ, Tomás-Ramón, *Curso de Derecho administrativo*, Cívitas,

HOFFMANN, Herwig, "The Duty of Care in EU Public Law – A Principle Between Discretion and Proportionality", *Review of European Administrative Law*, 13, 2020.

MARTÍNEZ ESPÍN, Pascual, "La propuesta de marco regulador de los sistemas de inteligencia artificial en el mercado de la UE", *Revista CESCO de Derecho del consumo*, 46, 2023.

PEÑA AMORÓS, Mª del Mar, *El deber de información*, Dykinson, 2020.

PEÑA AMORÓS, Mª del Mar, "Derechos fundamentales y deber de información", *Gaceta Fiscal*, 450, 2024.

RIVERO ORTEGA, Ricardo, *El Estado vigilante*, Tecnos, 1999.

RIVERO ORTEGA, Ricardo, *Derecho administrativo económico*, Marcial Pons, 2022.

RIVERO ORTEGA, Ricardo, *Derecho e inteligencia artificial: cuatro estudios*, Okejnik, 2023.

RIVERO ORTEGA, Ricardo, "¿Presuntos inocentes o presuntos culpables? La prueba de la responsabilidad subjetiva en el Derecho administrativo sancionador", Revista Española de Derecho administrativo, 2024.

SADDY, André, "Códigos de buenas prácticas. Concepto, naturaleza y su configuración como fuente de Derecho administrativo", en *Regulación y competencia en servicios de interés económico genera*l, 2017.

STUUMAN, kees/Lachaud, Eric, "Regulating iA. A Label to complete the proposed Act of Artificial Intelligence", *Computer Law and Security Review*, 2022.

USUARIOS DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL Y SUS OBLIGACIONES (*)

USERS OF ARTIFICIAL INTELLIGENCE SYSTEMS AND THEIR OBLIGATIONS

Por MERCEDES FUERTES

Catedrática de Derecho Administrativo.
Universidad de León

(*) Este trabajo se recibió el 28 de mayo de 2024 y fue aceptado en junio.

REVISTA DE
PRIVACIDAD Y DERECHO DIGITAL

RESUMEN

La autora analiza en este estudio las obligaciones específicas que han de cumplir los usuarios de los sistemas de inteligencia artificial teniendo en cuenta las diversas categorías de riesgos que presentan tales sistemas y las moratorias para la efectiva aplicación del Reglamento europeo previstas.

PALABRAS CLAVE: *Sistemas de Inteligencia Artificial. Usuarios de los sistemas de inteligencia artificial. Obligaciones de los usuarios. Responsables del despliegue de sistemas de inteligencia artificial. Obligaciones de los responsables del despliegue. Sistemas de inteligencia artificial de riesgo alto. Moratoria del Reglamento.*

ABSTRACT

In this study, the author analyses the specific obligations to be met by users of artificial intelligence systems, taking into account the various categories of risks posed by such systems and the expected moratoriums on the effective application of the European Regulation.

KEYWORDS: *Artificial Intelligence Systems. Users of artificial intelligence systems. Obligations of users. Deployers of artificial intelligence systems. Obligations of those responsible for deployment. High-risk artificial intelligence systems. Moratorium on the Regulation.*

SUMARIO

Agradezco a la dirección de esta Revista la invitación para realizar algunas consideraciones sobre las diversas obligaciones de los usuarios que precisa el Reglamento de inteligencia artificial. Conviene analizar y, sobre todo, debatir el contenido de este texto complejo -más complejo que extenso- por su relevancia. Pero también porque le seguirán otras disposiciones que tratarán de embridar la desaforada carrera tecnológica que estamos presenciando. Por ello, resulta oportuno, a raíz de un

debate constructivo, precisar los conceptos, atinar con un régimen jurídico ante los numerosos y dispares conflictos que suscitan los sistemas de inteligencia artificial, así como aclarar la redacción evitando la compleja lectura, las reiteraciones y otros defectos. Pues desde hace años se insiste en mejorar la "técnica legislativa" además de resultar imprescindible un correcto uso del español[1].

Sin perjuicio de la posible formulación de algunas críticas, la culminación de esta regulación ha de valorarse en términos generales de manera positiva. Los obstáculos durante su tramitación se hicieron evidentes: dificultades de atender a un ámbito tan innovador y de desarrollo tan acelerado, recordemos que avanzada la tramitación aparecieron los sistemas que generan contenidos y se introdujeron nuevas precisiones; tensiones entre las perspectivas diversas de los Gobiernos; posiciones distintas entre las grandes corporaciones empresariales, algunas voces solicitando incluso moratorias técnicas y normativas. Las negociaciones en el seno de las instituciones europeas consiguieron limar notables aristas, sortear significativos escollos y ofrecer un acuerdo final sobre el marco de regulación. No obstante, insisto en la concurrencia de dos circunstancias que generan dureza en la lectura: una, que todos los debates, audiencias y trabajos se desarrollaron en un inglés tecnológico, por lo que las expresiones jurídicas no son siempre afortunadas e, incluso, quizá la urgencia en facilitar versiones en todos los idiomas abuse de asociaciones con expresiones inapropiadas (me refiero a esas "traducciones" calificadas como "falsos amigos"); dos, y es un problema general de muchas normas europeas, la insistente adición de enmiendas genera una exposición con un hilo conductor no siempre coherente.

1 Santiago Muñoz Machado, como Director de la Real Academia de la Lengua Española, ha impulsado la Red Panhispánica del lenguaje claro que ha elaborado guías para un lenguaje claro y accesible, además de los libros de estilo, de ortografía y buen uso del español que se facilitan a través de la página de esa Real Academia. Vid. también su reciente libro *Fundamentos del lenguaje claro*, Ed. Espasa, 2024.

Partiendo de esta complejidad, adentrémonos en esta disposición con ánimo de analizar un aspecto concreto: qué obligaciones han de cumplir determinados usuarios con el fin de prevenir, reducir o corregir riesgos.

Porque su finalidad, la que motivó su impulso y ha persistido durante toda su tramitación, es atender a los riesgos, tratar de evitarlos o, al menos, minorarlos. Que los trepidantes avances, en la medida de lo posible, no se desparramen de manera desenfrenada pues hay que ser conscientes de los peligros que la frenética expansión de los sistemas de inteligencia artificial pueden traer.

Este Reglamento sigue la lógica de tantas previsiones normativas sobre la seguridad de los productos. Sin impedir la innovación, ni el desarrollo de estos sistemas, las instituciones europeas han fijado unas mínimas reglas con el fin de reducir el desconcierto, la preocupación ante las consecuencias desconocidas de un desenvolvimiento tan rápido y, con ello, asentar cierta confianza de los ciudadanos. Un marco jurídico que se irá completando y actualizando por la Comisión Europea pues son varias, e importantes, las delegaciones normativas establecidas.

No se establece, por tanto, un régimen jurídico integral de tales sistemas de inteligencia artificial. Ha quedado pospuesto, por ejemplo, entre otros aspectos trascendentes, el debate de la regulación del régimen de responsabilidad patrimonial[2]. Pero este marco jurídico sí acoge, como veremos, ciertas pautas éticas de comportamiento.

Centrado el objetivo en prevenir los riesgos, se abre la lente visual, un gran angular, para englobar el amplio ámbito de aplicación subjetivo que está afectado por esta disposición. Detengámonos en el mismo pues a quienes abarque ese foco serán los destinatarios de determinadas obligaciones.

2 Me refiero a la Propuesta de Directiva que adaptará las normas de responsabilidad civil extracontractual a la inteligencia artificial, COM (2022) 303, de 28 de septiembre de 2022.

I.- USUARIOS Y RESPONSABLES DEL DESPLIEGUE

Junto a los más directos partícipes en la creación y extensión de esta tecnología, quienes la diseñan, la desarrollan o comercializan, concurren circunstancias peculiares e intensas (la complejidad y sofisticación ínsita en su corazón, la ingente información de la que se nutre, cómo se interpreten sus resultados, entre otras), que reclaman la necesidad de contemplar cómo se maneja y cómo va funcionando. Por ello, interesa atender también a quienes utilizan los sistemas de inteligencia artificial, los usuarios.

En los documentos iniciales de trabajo, así como en la propuesta que finalmente presentó la Comisión Europea, hace ya tres años, se anunciaba que el Reglamento se aplicaría "a los usuarios del sistema de inteligencia artificial que se encuentren en la Unión"[3]. Vocablo este de usuarios que fue variando en los debates del Parlamento Europeo y del Consejo. En determinado momento, una enmienda precisó la condición de determinados usuarios, que se reflejó en algunos textos con el palabro "implementador" y en otros con la locución "responsable del despliegue". Es esta una mejor expresión dentro de la complejidad de delimitar las diferencias de *status* jurídicos de todos aquellos que utilizan un sistema informático.

Tales matices lingüísticos, de usuario a responsable del despliegue, no han afectado a la esencia de la definición legal inicial. Se ha mantenido con las mismas palabras desde la primera propuesta de la Comisión. Porque es ahora responsable del despliegue -como antes era el usuario- cualquier persona que haya incorporado y utilice para el ejercicio de funciones públicas, para su actividad profesional o comercial, un sistema de inteligencia artificial (art. 3.4).

3 Art. 29 de la Propuesta de la Comisión Europea de Reglamento de normas armonizadas en materia de inteligencia artificial COM (2021) 206, de 21 de abril.

Dos notas esenciales son las que delimitan el concepto. Una, que la utilización del sistema de inteligencia artificial esté bajo su responsabilidad, porque sea quien dirige la organización, o supervisa o vigila su funcionamiento. Dos, que el uso tenga por finalidad el ejercicio de sus competencias, que atienda al desenvolvimiento de su actividad profesional o mercantil. La utilización de estos sistemas en el ámbito exclusivamente particular y privado queda fuera de los focos de este Reglamento.

Esta definición legal acoge tanto a las personas físicas o como a las jurídicas. Alberga cualquier entidad, ya sea privada o de naturaleza pública porque se trata de una norma de carácter general, común a todos.

Cualquier profesional, empresario, compañía, que incorpore un sistema para el desarrollo de sus actividades, prestación de servicios, relaciones comerciales..., ha de atender a estas previsiones. Cualquier persona, aunque se trate de un único profesional o empresario; cualquier empresa, grande o pequeña. Hace años que se han difundido sistemas asequibles para ser utilizados sin tener especiales conocimientos de programación: sistemas creativos para el diseño particular de páginas web, de servicios permanentes de atención al cliente, de disponibilidad de repuestos y suministros, de reserva de citas previas, de precisar las mejores rutas de reparto, las distintas predicciones de ingresos y costes, y otras muchas utilidades que facilitan la llevanza del negocio a un profesional, empresario o a pequeñas empresas.

Y también cualquier "autoridad pública, órgano u organismo". La definición incluye a cualquier Administración u organismo público, lógicamente también a las instituciones europeas, a sus órganos y organismos (cosa que recuerda el considerando 23)[4].

4 Sobre las transformaciones que exige en el ámbito de las actuación administrativa la inteligencia artificial recomiendo desde este momento: Menéndez, E. *From bureaucracy to artificial intelligence: the tension between effectiveness and guarantees,* Wolters Kluwer, CEDAM. 2023; y el colectivo dirigido por Gamero, E. *Inteligencia artificial y sector público: retos, límites y medios,* Tirant lo Blanch, 2023.

No se aplicará a las autoridades de terceros países ni a organismos internacionales (art. 2.4)

Esa alusión a la "personalidad" no ha de interpretarse, a mi juicio, en el sentido de excluir a comunidades de bienes, a otros grupos o a colectivos sin personalidad jurídica, a fondos públicos carentes de personalidad, así como a otras formas pintorescas dentro del amplio vestuario de disfraces que ofrece el sector público. Asumirán personalmente sus miembros las específicas obligaciones atendiendo a la clasificación de riesgos del sistema, salvo que concretas previsiones en sus estatutos o acuerdos establezcan otras reglas de organización y funcionamiento que determinen la responsabilidad. Con relación a los fondos públicos sin personalidad jurídica u otros entes públicos, la responsabilidad será del organismo al que esté adscrito.

Quedan únicamente fuera de este régimen jurídico quienes utilicen esos sistemas de inteligencia artificial para determinados usos. Así, con fines militares, de defensa o de seguridad nacional, pues se mantiene dentro de la exclusiva competencia de los Estados miembros sin perjuicio de la política europea de seguridad y defensa (arts. 4.2 y 42 y ss del TUE).

Igualmente quedan fuera de este marco jurídico los sistemas que se usen en el ámbito privado, así como aquellos utilizados de manera exclusiva para la investigación y el desarrollo científico, (art. 2. apartados 6 y 10). Por tanto, en la Universidad, los profesores que utilicen estos sistemas dentro de sus actividades propias de estudio o investigación científica, quedarán fuera del ámbito de aplicación. Sí veremos que deberán -deberemos- atender a específicas obligaciones con relación al uso de algunos sistemas, como los que controlan los exámenes con el fin de evitar fraudes. Lo mismo que estará lógicamente obligada la Universidad como institución al utilizar estos sistemas de inteligencia artificial en el ejercicio de sus funciones, en la gestión de su actividad.

Tampoco se aplicarán estas disposiciones a la investigación, actividades de prueba y desarrollo de los modelos de inteligencia artificial antes de su puesta a disposición o comercialización (art. 2.8). Y, en algunos casos, los sistemas que se difundan con códigos abiertos o licencias libres, pues no podrán eludir las prohibiciones establecidas ni el régimen de los sistemas de alto riesgo y deberán garantizar su transparencia (art. 2.12).

Ese gran angular que se ha abierto para contemplar el horizonte subjetivo de aplicación es extraordinariamente amplio. Mira lógicamente a todos los Estados miembros. Por ello, todas aquellas personas que residen en la Unión, así como aquellas instituciones y compañías que tengan establecimiento en la Unión Europea estarán sujetas a estas disposiciones. Pero abarca más: aún localizándose en otros Estados, se exige el respeto a las disposiciones de este Reglamento cuando los resultados generados por el sistema de inteligencia artificial se utilicen dentro de la Unión Europea (art. 2 letras b y c).

Anótese la singularidad: incluso aunque el sistema de inteligencia artificial no se comercialice ni preste servicios en la Unión Europea. Y ello porque se es consciente de las múltiples relaciones que se traban con corporaciones extranjeras a raíz de las cuales se utilizan de manera lícita información y datos procedentes de la Unión. Si tales compañías extranjeras ofrecen los resultados del sistema, incluso sin comercializarlos, habrán de cumplir las previsiones de este Reglamento. El propósito de esta amplia extensión es evitar que se eludan las obligaciones por la mera circunstancia de estar localizada una empresa fuera de la Unión (considerando 22). El sistema de inteligencia artificial no solo es relevante por su diseño y estructura, es trascendente por el nutriente y, en consecuencia, si hay información y datos que proceden de Europa, han de respetarse las reglas de la Unión.

Dentro de la determinación del ámbito subjetivo de aplicación aparece otra locución diferente al "usuario" y al "responsable del despliegue". A saber, "persona afectada". Tiene tal consideración

cualquier ciudadano que, estando en el territorio de la Unión Europea, quede bajo el influjo de un sistema de inteligencia artificial (art. 1 *g)*). La existencia de afectados será un límite para el funcionamiento de algunos sistemas de inteligencia artificial, además de que se recuerdan los derechos que tienen ante los riesgos de estos sistemas.

Expuesto ese amplio ámbito de aplicación subjetivo, se estructura la regulación clasificando los sistemas de inteligencia artificial en sucesivas categorías por los riesgos que se intuyen. Se parte de aquellos que aparentemente presentan menos peligro -la inexistencia de riesgos no cabe en el ámbito informático-, porque se entiende que las tareas automatizadas se desenvuelven en un marco bastante controlado. En otro rango se alojan aquellos donde hay una interacción que genera particularidades, además de cierta incertidumbre. El siguiente nivel distingue aquellos sistemas que presentan riesgos notables y que, por ello, exigen mayores cautelas y, en último lugar, se señalan aquellos cuyo uso se considera inaceptable por la intromisión en la vida de los ciudadanos o porque, aunque se ignora qué riesgos pueden generar, se barrunta su notable peligro. De ahí que se prohíban.

Conozcamos esas obligaciones cuyo cumplimiento confiemos que minore los riesgos.

II.- HONESTE VIVERE, ALTERUM NON LAEDERE

El uso de una tecnología es desbordante, inunda todos los espacios vitales, se incorpora a las relaciones jurídicas, a los hábitos profesionales, son unos sistemas con infinitas posibilidades y, en consecuencia, tienen infinitos riesgos... Estas y otras consideraciones conducen a que se subrayen desde el primer

momento los básicos deberes y obligaciones para una prudente utilización.

En este sentido, procede un recordatorio básico. Vivimos tiempos en que parecen desvanecerse los perfiles de instituciones jurídicas, se agrietan las columnas que dan solidez al Estado de Derecho, hay ciudadanos que se desentienden de los deberes constitucionales, de esas responsabilidades elementales que se asumen por vivir en sociedades abiertas que se amparan en las garantías del Estado social y democrático. Parece olvidarse que los sistemas democráticos son frágiles y que resulta imprescindible un mínimo comportamiento respetuoso y tolerante con los conciudadanos para que esta sociedad abierta se mantenga. La fuerza de una Constitución reside en la firme determinación de los ciudadanos de defenderla, pues únicamente cuando cada persona se siente obligado en preservar ese marco, se pueden defender la libertad y los derechos fundamentales.

Apunto esta idea que, en principio, resulta elemental, porque parte de la crisis institucional y democrática que presenciamos tiene su causa en la débil educación cívica, en que no se insiste en el comportamiento responsable y en el cumplimiento de deberes con la misma convicción con la que se exigen y defienden los derechos.

Y parto de este recordatorio elemental porque ha de estar muy presente en el uso de las nuevas tecnologías que han explosionado ofreciendo un universo de posibilidades reales, efectivas, unas posibilidades que solo se habían soñado o imaginado por algunos escritores[5].

Ha de hacerse hincapié en que la utilización de estos sistemas que generan tantas ventajas, beneficios, un caudal voluminoso

5 En otras ocasiones he recordado los textos de Homero, Herón de Alejandría... y, partir de ahí podríamos recorrer tantas novelas que incluyen sistemas de automatización y extraños seres (las sucesivas obras basadas en la leyenda de Fausto, Frankestein, el hombre mecánico de Melville, los robots ajedrecistas del siglo XIX, las novelas de Julio Verne, etc.

de facilidades... tiene como presupuesto una actitud responsable, el respeto a unos deberes generales, el cumplimiento de unas obligaciones específicas. Todo derecho tiene su previo o simultáneo deber. Toda ventaja o beneficio procede de una previa responsabilidad. Porque, en caso contrario, habrá, como mínimo, un enriquecimiento injusto.

En estos sistemas de inteligencia artificial ha de exigirse con rigor esa responsabilidad, una exquisita diligencia en su uso. Son sistemas que se extienden, anegan y empapan muchas actuaciones sin que se adviertan la colosal información que ha deglutido, los vínculos que ha establecido, los pasos intermedios que ha realizado, las alternativas que ha descartado... hasta llevar en breve tiempo a un resultado. Un resultado que despertará la admiración al procesar ingente volumen de información en unos segundos. Admiración que la rutina atenuará. Pero la eventualidad de que en ese complejo sistema se hayan producido asociaciones de datos no buscadas, que se hayan desechado otras opciones, que arroje soluciones distintas a las esperadas, en fin, así como que hayan existido intromisiones mediante ciberataques, hace necesaria esa mayor atención y conciencia en su utilización[6].

Por ello, además de seguir las instrucciones específicas de uso, es imprescindible mantener una actitud cabal que se concreta en respetar unos deberes generales, al menos, como condensó Ulpiano en la expresión que he utilizado para este epígrafe: *honeste vivere, alterum non laedere* (Digesto 1.1.10.1).

Este clásico aforismo encuentra dos claras manifestaciones en el Reglamento de inteligencia artificial. Por un lado, una relación de usos prohibidos para evitar daños a otras personas; por otro, el respeto a unos códigos de conducta.

6 Resultan preocupantes las noticias que se suceden sobre ciberataques a estos sistemas que, por ejemplo, facilitan los controles de calidad del abastecimiento de agua a las poblaciones o el tráfico viario. Así, hace unos días se difundió el ataque a determinados sistemas de los vehículos lo que les condujo a ignorar las señales de tráfico.

Como tantos otros instrumentos y herramientas que utilizamos, los sistemas de inteligencia artificial resultan extraordinariamente versátiles y, por ello, lógicamente, no deberán utilizarse como medios de aquellas conductas reprochables, tipificadas como delitos o infracciones administrativas. En este sentido, el Reglamento prevé en su artículo 5 prohibiciones de uso por considerarlos inaceptables en una sociedad civilizada. Enuncia en apartados -sin una acertada sistemática- varios supuestos teniendo en cuenta que esta regulación, lógicamente, no desplaza otras previsiones de usos prohibidos y delimitación de infracciones ("*no afectará a las prohibiciones aplicables cuando una práctica de inteligencia artificial infrinja otra legislación de la Unión*").

En dos grandes grupos cabe clasificar estas prohibiciones. Uno, aquellas que se dirigen a proteger la identidad de cada persona, su privacidad e intimidad. Por ello, se veta la captación y tratamiento de imágenes para ampliar bases de datos de reconocimiento facial, ya provengan de circuitos cerrados de televisión o de un "raspado no selectivo" de otras que circulan por Internet. Igualmente se restringe el análisis de las expresiones con el fin de inferir emociones, creencias o identificar las intenciones de los actos personales (se admite alguna excepción por motivos de seguridad como los sistemas que alertan del cansancio del piloto de una aeronave). Tampoco catalogar a las personas por sus comportamientos y darles puntuaciones, previsión que repele el sistema de crédito social chino[7]; así como, fuera del estricto marco de requisitos en el ámbito judicial, realizar predicciones sobre el riesgo de comisión de delitos.

Un segundo grupo de prohibiciones se dirigen a garantizar la libertad personal, a evitar que los sistemas manipulen el comportamiento al distorsionar la verdad, adulterar o engañar... Esta injerencia de la tecnología en la esencia de la personalidad, que

7 Junto a documentales que se han difundido por los medios de comunicación, realiza una exposición detallada de este sistema Avaro, D. *El Sistema de Crédito Social chino. Vigilancia, paternalismo y autoritarismo,* Ed. Biblos, Buenos Aires, 2023.

podría tanto identificar los procesos mentales como alterarlos, es lo que ha de protegerse mediante un firme reconocimiento de los "neuroderechos" como desde hace tiempo están defendiendo varios científicos y, entre ellos, descato la labor de Valentín Fuster[8].

Procede insistir en las actitudes leales de *pro rei veritate*, de programar con autenticidad la información, que no haya estímulos subliminales. La manipulación es el abono de graves peligros, tanto en las relaciones sociales, por lo que corrompen la sociedad, como en las relaciones privadas o económicas.

El retorcimiento de los conceptos, la falsificación de las ideas llevaría a encerrarse en la "cueva" que denunció **Platón** en su obra *La República*[9]. La manipulación origina la percepción de una imágenes deformadas, inadecuadas, aquellas que en cada momento muestren quienes mueven los hilos del teatro de sombras chinescas en que se convertiría la sociedad. Los ciudadanos quedarían recluidos inconscientemente en un espacio turbio, sin la luz del sentido crítico, y con sus movimientos limitados ante el estrechamiento cada vez más angosto de las opciones que generan los filtros de los sistemas informáticos. Y, lo que no es menos grave, personas con una actitud sumisa, muy distinta a la enfurecida del Segismundo de **Calderón de la Barca**.

El riesgo es grave, tanto en las relaciones privadas (un ejemplo: las empresas de seguros ya han alertado de cómo se están trastocando con rapidez de manera muy sencilla las fotos tras un accidente de tráfico), como en el ámbito social ante las corrientes manipuladoras de la opinión pública. De ahí que la obligación

8 Resultan ilustrativas, y por ello recomendables, las conferencias de Valentín Fuster muchas de las cuales están disponibles a través de Internet. La Unión Europea suscribió en 2023 en León, una Declaración sobre la neurotecnología que debe centrarse en la persona y respetar los derechos humanos. Entre las últimas publicaciones jurídicas destaco la monografía de Beltrán de Heredia, I. *Inteligencia artificial y neuroderechos: la protección del yo inconsciente de la persona*, Ed. Aranzadi, 2023.

9 La República, Libro VII.

de informar con claras marcas o etiquetas cuando se generen o manipulen imágenes, sonidos, audios o vídeos... Se excluyen, lógicamente, los contextos creativos, artísticos, ya sea de sátira, ficción "análogos".

Resulta urgente adoptar medidas eficaces contra estos peligros. No obstante, se ha previsto una moratoria de seis meses de este artículo 5 desde la entrada en vigor del Reglamento y, desde ese momento, podrá denunciarse su uso. Recordemos que quienes informen de la existencia de tales sistemas prohibidos cuentan con la protección del régimen desplegado por la Directiva de protección de los denunciantes (arts. 87 del Reglamento)[10].

Ese plazo es más breve que los dos años que con carácter general se establece como moratoria para la aplicación del Reglamento. Aún así, me parece excesivo. Si se insiste en las actualizaciones periódicas de los sistemas informáticos ante los graves riesgos de ciberataques ¿por qué, entonces, no se pueden actualizar en menos de seis meses esos sistemas tan invasivos y perniciosos? ¿por qué no se puede impedir que no se distribuyan ya los nuevos sistemas que se están ultimando con rapidez para aprovecharse de esa moratoria?

Una escueta consideración, pues ya otro trabajo en esta Revista (firmado por Martín Razquin) profundiza en estos preceptos: la transgresión deberá tener una contundente contestación en el Ordenamiento jurídico.

El régimen sancionador ha de completarse por los Estados miembros pues la regulación en este Reglamento es limitada: señala las sanciones cuando quien incurre en esas reprochables conductas es un "operador" y tienen esa consideración a sus efectos: el "proveedor, fabricante de productos, implantador, representante autorizado, importador o distribuidor" (art. 3.8),

10 Directiva 2019/1937, de 23 de octubre que, como sabemos, se ha incorporado al Ordenamiento español mediante la Ley 2/2023, de 20 de febrero.

así como también a los responsables del despliegue que incumplen determinadas obligaciones (art. 99). Pero eso no impide que el Ordenamiento jurídico español precise tipos delictivos o infracciones administrativas graves por conculcar las prohibiciones si es que tales conductas no quedan ya incluidas en los tipos existentes como la suplantación de personalidad, las injurias y calumnias, los delitos contra el honor u otros.

El Reglamento de inteligencia artificial se ocupa de precisar las multas que podrá imponer el Supervisor europeo de protección de datos a las instituciones y organismos europeos (art. 100). Una regulación relevante porque puede abrir la puerta a extender la imposición de multas a Administraciones y organismos públicos ante graves incumplimientos. Algo que, como es conocido, está excluido en muchos sectores[11].

Junto a tales prohibiciones hay que insistir en otra idea.

La conciencia de los riesgos, el avance hacia lo desconocido, la convicción de mantener las bases de la civilización ha generado numerosas llamadas de atención recordando unos compromisos éticos básicos. Porque hay que honrar la dignidad humana, defender la libertad individual, proteger los derechos fundamentales y de las libertades públicas. Sólo así la musa de la confianza acompañará el uso de los sistemas de inteligencia artificial.

Entre otros muchos documentos propiciados por las instituciones europeas, el Grupo de expertos para asesorar la estrategia sobre inteligencia artificial presentó unas directrices éticas como base insustituible[12]. Concretaron principios generales con el fin de garantizar el diseño y el uso de estos sistemas. Entre otros:

11 Vid. art. 40 del Decreto-Ley 12/2018, de 7 de septiembre, de seguridad de las redes y sistemas de información o art. 77 de la Ley orgánica de protección de datos. Precisiones oportunas sobre este singular régimen jurídico realiza Domínguez Álvarez, J.L., *Iusdata y Administración pública,* Civitas, 2023.

12 Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, Directrices éticas para una IA fiable, Oficina de Publicaciones, 2019, disponible en https://data.europa.eu/doi/10.2759/14078.

el respeto a la autonomía humana, a la equidad y no discriminación, su solidez técnica y seguridad, su transparencia, garantizar el bienestar social y ambiental, así como prevenir el daño.

Tales principios hay que enmarcarlos en su específico contexto ante la complejidad y sofisticación de algunos sistemas[13]. Así, la transparencia se reconduce a la posibilidad de seguir el trazado de los procesos y de contar con una explicación adecuada. Pero sabemos que hay sistemas con una especie de "caja negra" integrada por modelos de aprendizaje interconectados, por miles de millones de instrucciones que dan resultados sin ser posible advertir el por qué, el cómo llegaron a esa conclusión. Se ha difundido el siguiente ejemplo: el modelo de "chat Gpt4" incluye en su caja negra más de ciento setenta y cinco mil billones de parámetros. Quizá en cifras se aprecie mejor la profunda inmensidad de los ceros: 175.000.000.000.000.000.

Es probable -al menos deseable- que el propio desarrollo tecnológico permita ofrecer aclaraciones más asequibles y sencillas de algunos parámetros. No obstante, habrá de incrementarse el esfuerzo de los programadores con vistas a facilitar esa información, la trazabilidad de los procesos, el seguimiento de las huellas porque también resulta necesario evaluar y realizar auditorías periódicas.

El Parlamento europeo insistió igualmente en 2020 en esa conciencia ética en el diseño de los sistemas de inteligencia artificial: han de respetar la dignidad humana, la autonomía individual, la seguridad, los derechos fundamentales reconocidos en la Carta europea[14]. Del mismo modo que se insistió en la preocupación

13 Entre otros comentarios, me remito al trabajo de Oliver, N *Inteligencia artificial, naturalmente*, ONTSI 2020, en el que con gran claridad resume la distinción que ha generalizado Jenna Burrell, quien diferencia: a) una opacidad derivada de la propiedad intelectual, b) otra porque sólo los entendidos conocen el lenguaje y lo entiendan y c) el aprendizaje profundo de la máquina que sólo los muy versados pueden interpretarlo

14 Resolución del Parlamento Europeo de 20 de octubre de 2020, de recomendaciones a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL), (2021/C 404/04).

por el impacto ambiental de la tecnología. En muchos de sus apartados se insiste una y otra vez la necesidad de mitigar y remediar el impacto general sobre "los recursos naturales, el consumo de energía, la producción de residuos, la huella de carbono, la emergencia climática y la degradación del medio ambiente". Pues es ingente el consumo de energía de estos sistemas ante las exigencias de la supercomputación[15]. Además de sus altos costes económicos[16].

Pero la exigencia ética ha de ser atendida en primer lugar. Y así se reitera en distintos foros internacionales[17].

Un compromiso ético que no se circunscribe a la estructura del sistema de inteligencia artificial por la labor de quienes lo diseñan. Ha de presidir el comportamiento de los usuarios. Recordemos que algunos de estos sistemas, con su mera puesta en marcha, coleccionan datos del usuario y, sobre todo, intensifican su "entrenamiento", de tal modo que pueden generar nuevas vinculaciones y otros resultados. De ahí la necesidad de que el comportamiento del usuario haya de ser correcto con el fin de evitar distorsiones en ese entrenamiento.

15 Algunos medios de comunicación han difundido el resumen de un artículo publicado en febrero de 2024 en la revista Nature sobre los impactos ambientales de estos sistemas de inteligencia artificial, lo que ha generado la presentación de un proyecto de ley en el Senado norteamericano para investigar y medir tales impactos. También en España, se han divulgado informaciones sobre el coste energético y la huella de CO2. En resumen, el entrenamiento de más de 200 mil millones de parámetros en los sistemas de inteligencia artificial generan 752 toneladas de CO2. Y hay sistemas donde los parámetros se han multiplicado (Megatron-Turing tiene 530 mil millones, PaLM tiene 540 mil millones de parámetros, y el recientemente aparecido GPT-4 aumenta hasta 170 billones de parámetros). Para advertir la densidad de esta cifra sirva la referencia de que un vuelo de ida entre Madrid a Múnich que genera media tonelada de CO2. En consecuencia, el entrenamiento equivale a miles y miles de vuelos.

16 Aspecto que ha sido considerado por el Artificial Intelligence Index Report 2024 del Instituto de la Universidad de Stanford HAI Human-centered Artificial Intelligence.

17 Sirva el mero recordatorio el Acuerdo del G7 de código de conducta bautizado como "Proceso de inteligencia artificial de Hiroshima, de 29 de octubre de 2023); las conclusiones de la Cumbre internacional sobre inteligencia artificial celebrada en Londres en noviembre de 2023; o de la Resolución de la Asamblea General de las Naciones Unidas reclamando el compromiso de respeto a los derechos humanos en el uso de los sistemas de inteligencia artificial (Resolución 21 de marzo de 2024).

En consecuencia ha de avanzarse en una especie de alfabetización con relación a estos sistemas de inteligencia artificial. Alfabetización que incluye el glosario del Reglamento para referirse a unos conocimientos básicos que, por supuesto, no han de llegar a "leer y escribir el sistema", sino a "*tomar conciencia de las oportunidades y los riesgos que plantea la inteligencia artificial, así como de los perjuicios que puede causar*" (art. 3 apartado 56).

Y, de ahí la importancia de la concreción y difusión de "códigos de conducta". Se confía en la difusión de pautas que vayan incorporándose como hábitos voluntarios (considerando 165).

Los proveedores de los sistemas son quienes están en mejor posición para su redacción, pero también podrán elaborarse por los responsables del despliegue que advierten cómo funcionan los sistemas en el seno de una organización y las incidencias que suscita su uso. Del mismo modo que pueden impulsarse tales pautas por investigadores o asociaciones civiles. En los considerandos se citan a las organizaciones de consumidores o a los sindicatos, pero no hay que ignorar que existen institutos científicos especializados, entre otros, el *Future of Life Institute* del MIT[18], el *Future of Humanity Institute* de Oxford[19], y, de manera más intensa, el *Alan Turing Institute*[20] que han difundido relevantes estudios sobre cómo deben los sistemas de inteligencia artificial atender a los valores y principios aceptados como éticos en las sociedades civilizadas.

Junto a tales iniciativas privadas, el Reglamento destaca el papel que pueden desempeñar la Oficina europea de inteligencia artificial, así como de las autoridades de los Estados miembros con el fin de difundir las mejores prácticas y soluciones técnicas (art. 95). Es más, se prevé que esa Oficina europea precise tales códigos de

18 https://futureoflife.org/

19 https://www.fhi.ox.ac.uk/

20 https://www.turing.ac.uk/

buenas prácticas en el plazo máximo de nueve meses a partir de la entrada en vigor del Reglamento, el 2 de mayo de 2025 (art. 56.9).

Pautas éticas y códigos de conducta resultan imprescindibles en el uso de los sistemas de inteligencia artificial "que generan contenidos", esto es, aquellos que no se han diseñado para una finalidad específica, sino como asistentes "sabelotodo". Estos sistemas son objeto de una atención especial en los artículos 51 y siguientes del Reglamento. Requieren mayor supervisión por los proveedores y, como veremos, obligaciones de los responsables en las instituciones y empresas que los utilizan. Pero también de los usuarios porque con ellos se interactúa. Quien utilice estos sistemas no debe incorporar noticias falsas ni erróneas ni realizar otro tipo de "usos indebidos", cosa que el Reglamento incluso considera "razonablemente previsible".

Recordemos que estos sistemas devoran toda la información y la relacionan siguiendo una lógica interna de predicción, considerando todas las circunstancias, valorando todas las alternativas y de ahí que la introducción de yerros puede dar lugar a sofismas necios, artificios temerarios, dislates mayúsculos... lo que llevará a perder fiabilidad y seguridad. Y, en este caso, el usuario será corresponsable. Un comportamiento honesto, siguiendo las instrucciones dadas por el proveedor, es el que debe presidir el uso de esos sistemas de capacidad espectacular, mientras se incorporan en su diseño pautas de corrección de tales disparates a través de los mecanismos de gestión de riesgos (art. 9) o se corrigen y evitan por la supervisión humana (art. 14).

III.- OBLIGACIONES DE LOS RESPONSABLES DEL DESPLIEGUE

Hemos de detenernos ya en las obligaciones específicas impuestas a quienes sean responsable del despliegue de un sistema de inteligencia artificial en una institución, profesión o negocio.

III.1.- ADVERTIR DEL ARTIFICIO

La primera precaución ante la extensión de estos sistemas es que los ciudadanos sean conscientes de que están relacionándose con un asistente artificial o que están bajo su radio de acción. El artículo 50 impone varias obligaciones, no solo a los proveedores para que en el diseño existan marcas o etiquetas específicas, sino también a los responsables del despliegue, quienes nos interesan en estos momentos[21].

Así, los sistemas que facilitan una relación recíproca, una directa interactuación con el fin de denunciar delitos u otras infracciones, han de dar noticia previa, salvo, dispone el Reglamento, que resulte "evidente" que se trata de un mecanismo elaborado, artificioso. Tal "evidencia" queda delimitada por el comportamiento de una persona "atenta y perspicaz", "razonablemente informada", consideraciones estas que pueden resultar indubitadas en ocasiones, pues se muestran dibujos animados especialmente significativos pero que, en otras ocasiones, aquellas que se han diseñado para facilitar la confianza de los usuarios, pueden generar una falsa percepción de que ciertamente el asistente es una persona con alguna responsabilidad o autoridad.

Quienes utilicen sistemas para generar textos, imágenes o vídeos han de avisar de tal circunstancia para evitar que los ciudadanos incurran en el error de considerarlos reales y auténticos. Si esa composición artificial tiene una finalidad artística, creativa, será suficiente en estos casos con hacer pública la existencia de elementos producidos por tales sistemas al inicio de la obra.

21 Entre otros, resultan ilustrativos los estudios de Miranzo Díaz, J. *Inteligencia artificial y Derecho Administrativo*, ICA-Tecnos, 2023; Boix A. y Soriano A., "Transparencia y control de uso de la inteligencia artificial por las Administraciones públicas", en la obra coordinada por Balaguer Callejón F. y Cotino Hueso, L. *Derecho público de la inteligencia artificial*, Fundación Giménez Abad, Zaragoza, 2023; Cotino, L, "Discriminación, sesgos e igualdad de la inteligencia artificial en el sector público" y Martín Delgado, I., "La aplicación del principio de transparencia a la actividad administrativa algorítmica", ambos en obra colectiva Gamero, E., *Inteligencia artificial …*, cit. págs. 260 y ss., y págs. 132 y ss., respectivamente, sobre la transparencia y la corrección de discriminaciones y sesgos por las Administraciones públicas, a los que me remito.

Del mismo modo, cuando se utilicen sistemas de inteligencia artificial para generar textos e informar de asuntos de "interés público" ha de difundirse tal circunstancia. Quedan fuera de esta obligación aquellos casos en que ha existido una supervisión humana, un control editorial o esté prevista ya la responsabilidad por la publicación de ese contenido. Como, con carácter general, quedan excluidas de esta obligación de información el uso de sistemas que hayan sido autorizados legalmente para "detectar, prevenir, investigar o enjuiciar delitos".

Ha de advertirse en aquellas situaciones en que se admite alguna identificación biométrica. Por ejemplo, caso de ciertos controles de entrada como el sistema de acceso al espacio Schengen: un programa que comprueba los datos de identificación de los pasaportes, con otros faciales o huellas dactilares. Igualmente en consultas médicas donde ese análisis biométrico puede resultar determinante para una adecuada exploración. Como también si el tratamiento médico atiende al reconocimiento de emociones. En estos casos, ha de comunicarse del funcionamiento del sistema a las personas afectadas además, lógicamente, ha de garantizarse el escrupuloso respeto a la normativa de protección de datos personales pues estamos en ámbito de materia sensible.

Se exige, además, que se dé de manera clara y nítida en el primer momento, en la "primera interacción o exposición".

La Oficina europea deberá precisar buenas prácticas de "etiquetado y detección" que garanticen la rápida percepción por el usuario, que se adviertan con claridad que ha intervenido un sistema de inteligencia artificial. Es imprescindible que los ciudadanos se percaten desde el primer momento, por ello se ha habilitado a la Comisión Europea para que apruebe códigos de buenas prácticas como Derecho derivado, remisiones normativas denominadas "actos de ejecución" en la terminología del Tratado y que están sujetos a un control del Parlamento y el Consejo que pueden, en cualquier momento, requerir información,

así como la previa consulta a los "comités" integrados por representantes de los Estados miembros[22].

III.2.- CONJURAR LOS PELIGROS

Algunos sistemas de inteligencia artificial inciden en ámbitos sensibles de tal modo que, a pesar de las significativas ventajas de su uso, persiste la preocupación ante los riesgos que genera. Procede prestar mayor atención en su utilización pues el sistema opera, podríamos decir, "por su cuenta", pero no "por su cuenta y riesgo", en el sentido de asumir su responsabilidad. Esta recae en la institución que utiliza el sistema cuyos efectos, en algunos casos, pueden ser imprevisibles. De ahí que el Reglamento imponga obligaciones específicas que minoren los riesgos y, con ello, la responsabilidad.

Las obligaciones se concretan en actuaciones de prevención y supervisión, nuevas "cargas" como reconoció la Comisión Europea en la propuesta normativa donde apuntó, incluso, una aproximación de los posibles costes.

Es cierto que desde hace años la legislación europea y española están insistiendo en aligerar las cargas administrativas, especialmente de los empresarios, de tal modo que "*las Administraciones Públicas que en el ejercicio de sus respectivas competencias creen nuevas cargas administrativas para las empresas eliminarán al menos una carga existente de coste equivalente*" y que las memorias de impacto normativo han de valorar las nuevas cargas impuestas[23].

22 Vid. Reglamento 182/2011, de 16 de febrero, que establece las normas y los principios generales relativos al control por parte de los Estados miembros del ejercicio de competencias de ejecución de la Comisión Europea.

23 Art. 37 de la Ley 14/2013, de 27 de septiembre, de apoyo a los emprendedores y su internacionalización y art. 2 del Decreto 931/2017, de 27 de octubre, sobre la memoria de impacto normativo.

Sin embargo, en este ámbito donde se requiere un especial cuidado, estas obligaciones se consideran razonables y proporcionadas. La Comisión Europea señaló que no se han encontrado otras medidas que incidan menos en el comportamiento de estos responsables.

Para paliar en alguna medida su coste, el Reglamento anuncia el apoyo, especialmente a las pequeñas y medianas empresas por parte de la Unión Europea y de los Estados miembros. Se dotarán cuantiosos fondos europeos con este fin, además de la asistencia que llevarán a cabo la nueva Oficina europea de inteligencia artificial y las autoridades nacionales [24].

Presupuesto indispensable para el cumplimiento de estas obligaciones es que los sistemas estén diseñados para que sean "interpretados" por estos responsables y que se les traslade suficiente información sobre su funcionamiento con el fin de advertir los usos previstos, aquellos que están excluidos, las ventajas y limitaciones, las pautas o cambios predeterminados... Tales aspectos se relacionan a lo largo del artículo 13 del Reglamento, donde se establece que dicha información ha de ser "concisa, completa, correcta y clara, que sea pertinente, accesible y comprensible". Los proveedores deberán hacer un esfuerzo significativo para conseguir esa accesible comprensión. Una cauta previsión es que se ilustre con ejemplos. Ha de evitarse que las expresiones técnicas sean un obstáculo para la adecuada interpretación.

Estas obligaciones más específicas se exigirán en todos aquellos usos de los sistemas calificados de "alto riesgo". ¿Cuáles tienen tal calificación? Por un lado, aquellos que afecten a "la salud, la seguridad y los derechos fundamentales", como reiteran varios

24 En España, tanto el organismo RED.ES como INCIBE tienen una atención especial para las pequeñas y medianas empresas. Además, se acaba de aprobar una Estrategia de inteligencia artificial en la que se anuncian nuevas ayudas para las empresas (se cifran en 700 millones de euros), además de otras dirigidas al centro de supercomputación MareNostrum, el mayor centro de supercomputación de España, alojado en Barcelona.

preceptos. Así, aquellos sistemas que se integren dentro de los elementos de seguridad de determinados productos y estén sometido a una evaluación para ser comercializados o prestados de conformidad con la normativa europea mencionada en el Anexo I, que relaciona un largo listado que atiende a juguetes infantiles, ascensores, vehículos, diversa maquinaria...

Por otro lado, aquellos que se utilizan en ámbitos específicos a los que alude el Anexo III pero con una importante salvedad: quedan excluidos de manera explícita aquellos sistemas que se utilicen para actividades auxiliares o preparatorias, que no influyan en las decisiones del sistemas o que se utilicen con posterioridad a las actividades humanas con el fin de comprobar sus resultados y mejorarlas. Porque no es lo mismo, por ejemplo, utilizar sistemas para localizar bibliografía o jurisprudencia que para redactar resoluciones judiciales.

¿Cuáles son esos ámbitos a los que alude el citado Anexo III? Algunos sistemas que utilizan datos biométricos ya que es necesario prevenir los riesgos de que se difundan tales datos. Por ello, se distinguen aquellas situaciones en que su captación y uso están radicalmente prohibidos, caso de pretender clasificar por categorías a las personas, facilitar la identificación remota, analizar emociones... de aquellas otras permitidas como la identificación de un sospechoso o en el control interno de acceso a determinadas instalaciones.

Son también de alto riesgo los sistemas de inteligencia artificial que incorporan medidas de seguridad para la protección de infraestructuras críticas, esto es, elementos indispensables para el adecuado funcionamiento de servicios esenciales como el abastecimiento de agua, el suministro eléctrico u otros a los que atiende la Directiva 2022/2557, de 16 de diciembre[25].

25 Esta Directiva de protección de las entidades críticas ha dado un gran impulso a la armonización de normas mínimas de ciberseguridad en ámbitos esenciales y estratégicos. Explico las razones que motivaron su adopción, modificando la anterior perspectiva de la Unión en mi monografía *Metamorfosis del Estado. Maremoto digital y ciberseguridad*, Marcial Pons, 2021, págs.

También se consideran de alto riesgo los que se utilizan en el sistema educativo y de formación con el fin de valorar las solicitudes de admisión, la calificación de las pruebas y exámenes realizados, también la vigilancia para evitar comportamientos prohibidos por parte de los estudiantes. En el ámbito de las relaciones laborales, aquellos que analizan las solicitudes de empleo, los que distribuyen tareas teniendo en cuenta los comportamientos individuales, los que evalúan el rendimiento en el trabajo... Igualmente, los sistemas dirigidos a analizar las solicitudes de servicios sociales, así como aquellos otros que supervisan las prestaciones con el fin de revisarlas o suprimirlas; los que valoran las solvencia financiera o una calificación crediticia; los utilizados para precisar las condiciones de los seguros sanitarios o de vida; aquellos que ordenan la preferencia de las llamadas de emergencia o el triaje de pacientes para su atención sanitaria; en fin, aquellos utilizados para valorar el riesgo de comisión de un delito, así como para verificar la veracidad de las declaraciones...

Y reitero: quedan fuera los sistemas que contribuyen a realizar actividades meramente auxiliares o preparatorias, los que no influyen en las decisiones del sistemas o que se utilicen con posterioridad a las intervenciones humanas con el fin de comprobar sus resultados y mejorarlas, caso, por ejemplo, de la comprobación de las medidas adoptadas para proteger una infraestructura crítica.

La necesidad de mantener al día este marco jurídico ante la imparable aceleración del desarrollo de estos sistemas, así como también, en términos generales, de las técnicas que pueden incorporar medidas de seguridad, ha generado que se habilite de manera explícita a la Comisión Europea para actualizar este Anexo III (arts. 7 y 97 del Reglamento).

Lógicamente se precisan los presupuestos que ha de tener en cuenta la Comisión, tanto para retirar esa alta precaución de riesgo de algún sistema, como para incluir otro del que se advierten sus peligros. La incorporación de ajustes en el diseño puede conducir a que se desvanezcan los temores considerables para

la inicial calificación siempre que satisfagan las reglas generales de seguridad del Derecho europeo. Y, a la inversa, la difusión de nuevos sistemas de inteligencia artificial, que puedan utilizarse en esos ámbitos sensibles y que presenten riesgos similares o más altos que los sistemas ya clasificados, originará su incorporación en esa categoría de alto riesgo. Para ello se exige que la Comisión realice una valoración detallada de diversos elementos como la finalidad del sistema, datos de los que se nutre y las probabilidades de generar tanto beneficios como perjuicios desproporcionados (art. 7).

Hay que ser conscientes de la rápida extensión y generalización de estos sistemas ante las ventajas de su utilización por todo tipo de organismos públicos, así como por las empresas, incluso pequeñas. ¿No conviene que los colegios, institutos y universidades utilicen sistemas para garantizar el comportamiento honesto en los exámenes? ¿han de descartarse las ventajas de las exploraciones médicas a distancia o la completa información que ofrecen programas sanitarios a la hora de graduar las urgencias? Los sistemas que ayudan a valorar las solicitudes de empleo, el análisis de los curricula, el rendimiento en el trabajo, la productividad de los funcionarios públicos, serán cada vez más frecuentes. Como aquellos que facilitan la supervisión del correcto uso de las ayudas y subvenciones o de los servicios sociales.

Señalados estos presupuestos, detengámonos ya en las concretas obligaciones que deberán cumplir todas aquellas instituciones o responsables del despliegue que incorporen sistemas de alto riesgo.

III.2. A.- De manera previa a su instalación

- Análisis de su incidencia en los derechos fundamentales.- El Reglamento exige evaluar el posible impacto sobre los derechos fundamentales de quienes se vean afectados por el sistema (art. 27). Esta obligación no se impone con

carácter general. Solo se dirige a determinadas instituciones. En concreto: todos los organismos públicos, las empresas privadas que prestan servicios públicos (estupendo recordatorio de una categoría clásica), así como quienes, ya tengan naturaleza pública o privada, evalúen las circunstancias personales en dos ámbitos muy sensibles, a saber, la prestación de servicios esenciales de asistencia pública, incluida la sanitaria, y la solvencia patrimonial o calificación crediticia, con la excepción de que se persiga el fraude financiero.

No será exigible tal análisis cuando la autoridad de vigilancia competente haya acordado, ante situaciones excepcionales, dispensar al proveedor del sistema de tal evaluación (art. 46.1).

Quedan igualmente excluidos todos aquellos organismos y empresas que gestionan un servicio estratégico y cuentan con infraestructuras críticas. Recordemos que, tras la nueva regulación de protección de entidades críticas, la Unión Europea ha ampliado los sectores considerados estratégicos precisando una evaluación rigurosa y detallada, un concienzudo análisis de riesgos... y, de ahí, su preferencia, que desplaza la atención a este precepto[26].

Junto a esta lógica exclusión, sin embargo no me parece tan razonable que no se exija la realización de una evaluación previa sobre el impacto en los derechos fundamentales al resto de "responsables del despliegue", esto es, a otras personas o empresarias ya que nos encontramos con sistemas que se han calificado de alto riesgo. Si ello se ha debido a no querer incrementar las cargas de los empresarios con otra obligación, me parece poco acertada la decisión.

26 Me refiero a la Directiva 2022/2557, de 14 de diciembre. Sobre su extensión, así como la divergente aplicación del régimen europeo previo que ha motivado la reforma, me remito a las consideración que realizo en Metamorfosis del Estado..., cit.

Los derechos fundamentales deben ser respetados por todos los ciudadanos y empresarios, y asumir la utilización de un sistema de inteligencia artificial que eventualmente puede lesionar algún derecho fundamental es causa suficiente para una actitud más cuidadosa. Además, hay que tener en cuenta que esta carga tiene un peso, podríamos decir, "liviano". Por un lado, la obligación solo se exige en el "primer uso del sistema". Pero, si con anterioridad ya se han realizado por otros organismos o, lo que es más probable, ha realizado minuciosas evaluaciones de impacto el proveedor, así como si ya se han realizado evaluaciones de impacto en el ámbito de la protección de datos, se entiende que queda satisfecha esta obligación. Por otro lado, la Oficina europea de inteligencia artificial tiene el encargo de elaborar un modelo de cuestionario que sea sencillo rellenar "mediante una herramienta automatizada", lo que facilitará tal evaluación.

El Reglamento precisa el contenido mínimo al que deben atender tales evaluaciones: a) procedimientos en los que se utilizarán; b) durante cuanto tiempo y frecuencia; c) qué personas pueden verse afectadas por su utilización, ya sean personas individuales o grupos de personas; d) qué tipo de riesgos les pueden perjudicar; e) qué medidas de supervisión humana se aplican; y f) qué medidas se adoptarán en caso de producirse perjuicios, incluidos los medios de reclamación. Aspectos que pueden describirse en asequibles formularios de ayuda.

- Información previa a los representantes de los trabajadores.- Previsión general en la legislación laboral es la información que ha de darse a los representantes de los trabajadores de manera previa[27]. El Reglamento obliga a

27 Sirva el recordatorio de la Directiva 2002/14, de 11 de marzo, que establece el marco general de información y consulta a los trabajadores, así como al artículo 64 del Estatuto de los trabajadores.

los proveedores a facilitar información con el fin de que se entienda el funcionamiento del sistema, un diseño que permita cierto conocimiento, unas mínimas exigencias de transparencia (art. 13). Por tanto, deberá ponerse a disposición de los representantes de los trabajadores tales explicaciones, así como responderse a sus dudas y, por su trascendencia, debido a los altos riesgos del sistema, debería generar el rápido traslado de esta información a todos los trabajadores.

Del mismo modo que debería facilitarse la información inversa como empieza a preocupar a las empresas. Esto es, qué sistemas de inteligencia artificial están empezando a utilizar los trabajadores por su cuenta y riesgo, pues conocen por su uso diario y doméstico algunas de sus ventajas.

- Registro previo.- Otra obligación a la que se sujeta solo a determinados "responsables del despliegue" es el registro en la base de datos que ha configurado la Unión Europea (arts. 26.8 y 49 del Reglamento). Afecta a las autoridades y organismos públicos. Quedan fuera las compañías privadas u otros profesionales que utilicen estos sistemas.

  La Comisión Europea, con la colaboración de los Estados miembros, creará la base de datos con el fin de ofrecer información accesible a cualquier ciudadano sobre los sistemas de alto riesgo que se utilicen en toda la Unión Europea: un resumen sencillo sobre su finalidad, los datos que colecciona y su funcionamiento. Servirá para el adecuado ejercicio de las funciones encomendadas a la Comisión y otras autoridades con competencias en este ámbito porque incorporará datos del proveedor y del sistema (art. 71 del Reglamento). De tal modo que, cuando estos concretos responsables, estas instituciones públicas, comprueban que el sistema no está debidamente registrado, se le impide su utilización (art. 26.8). Y aquí se abre un interrogante:

¿por qué no se impide igualmente la utilización a empresarios y profesionales?

Sabemos que los proveedores de los sistemas de alto riesgo han de registrarse en esa base de datos europea antes de ponerlo a disposición o comercializarlo (art. 6.4). En consecuencia, ¿por qué pueden utilizarlo empresarios y profesionales? o, incluso, el sector público empresarial cuando se impide a los organismos públicos. Dualidad que, a mi entender, no está justificada y generará problemas.

La información que ha de facilitarse para tal inscripción por los responsables del despliegue se recoge en el Anexo VIII sección C del Reglamento.

Quedan fuera de esta base de datos europea los sistemas de inteligencia artificial que se utilicen para proteger infraestructuras críticas. Las instituciones europeas carecen de información sobre qué entidades estratégicas tienen infraestructuras críticas, pues son datos que afectan a la seguridad de cada Estado miembro. De ahí que, desde la primera regulación europea del régimen jurídico para su protección especial, se excluyó informar a la Comisión sobre tales elementos cruciales. Recordemos que, a pesar de las políticas europeas de seguridad, del principio de solidaridad y ayuda mutua, también el Tratado admite que "*ningún Estado miembro estará obligado a facilitar información cuya divulgación considere contraria a los intereses esenciales de su seguridad*" (art. 346 TFUE). Por ello, en estos casos, el registro es nacional. En España, ejercen estas competencias el Comité de protección de las infraestructuras críticas.

III.2.B.- Adopción de medidas técnicas y organizativas

Incorporado un sistema de inteligencia artificial ha de garantizarse que se siguen las instrucciones de uso. Cada institución,

empresa o profesional, atendiendo a su propia estructura y a la complejidad del sistema, establecerá las medidas organizativas y técnicas que considere. El Reglamento alude a la propia "libertad", a la particular consideración de cada institución para "organizar sus propios recursos y actividades" porque tales medidas dependen de elementos internos y particulares (art. 26.3). Criterio éste distinto al que se ha seguido en otras disposiciones europeas, por ejemplo, en la normativa sobre seguridad de las redes y los sistemas de información, donde se exige la designación de un responsable de ciberseguridad que supervise las medidas adoptadas o notifique los incidentes. Este Reglamento de inteligencia artificial reconoce la autonomía para la organización de las atribuciones y responsabilidades.

Algunas organizaciones o empresas están creando ya un nuevo departamento, una unidad administrativa, un comité e, incluso, un director que asume la responsabilidad de la supervisión de estos sistemas. Junto a los acrónimos ingleses que salpican desde hace años los organigramas empresariales para hacer referencia al consejero o director ejecutivo (CEO), al responsable de la oficina de supervisión de cumplimiento normativo (CCO), al director de ciberseguridad (CISO), aparece ahora también el director de inteligencia artificial (CIAO, *chief intelligente artificial officer*).

Del mismo modo, se podrá establecer una lista de actividades previas antes del uso, como otros hábitos simultáneos a su utilización y medidas de control posterior con el fin de que todos aquellos funcionarios o todos aquellos trabajadores de la empresa que utilicen dispositivos con estos sistemas tengan un conocimiento suficiente de las mismas.

Además, al ser tan amplio el universo de aplicaciones que ofrece la multitud de sistemas de inteligencia artificial, será cada institución la que determine quién haya de manejarlos. Así, en un quirófano únicamente determinados especialistas, con una cualificación relevante, serán quienes puedan ser asistidos por sistemas quirúrgicos o los que atiendan a la modificación de las

dosis químicas; mientras que, por ejemplo, en los procedimientos de supervisión de las ayudas públicas u otros programas que ahorran la realización de un sin número de tareas repetitivas, todos aquellos funcionarios previamente aleccionados lo utilizarán y pronto lo incorporarán con soltura a sus hábitos.

Tales usuarios deberán seguir de manera rigurosa las instrucciones recibidas y para ello se exige que cuenten con unos conocimientos adecuados, una formación y competencia suficiente. Obligación en la que insiste el artículo 4 del Reglamento al señalar que "en la medida de lo posible" todas aquellas personas que se encarguen del funcionamiento y de la utilización de sistemas de inteligencia artificial tengan suficiente formación, conocimientos técnicos y experiencia.

Dentro de la institución se podrán adoptar pautas de utilización, códigos de conducta, además de reiterar aquellos difundidos por los proveedores de los sistemas (art. 95.3 del Reglamento). Con relación a estos códigos que inciden en sistemas de alto riesgo, la Comisión Europea ha de evaluarlos de manera periódica (cada cuatro años) con el propósito de considerar su extensión a otros sistemas que no sean de alto riesgo (art. 112.7).

Impone el Reglamento con carácter general, como medida técnica específica, la conservación de aquellos archivos que el propio sistema de inteligencia artificial genera de manera automática, aquellos en los que quedan registradas las sucesivas operaciones y que permitirán interpretar, en su caso, la nítida huella temporal de la actividad realizada. Porque estos sistemas de alto riesgo han de permitir su seguimiento a lo "largo de todo el ciclo de vida", además de otras capacidades tan relevantes como: detectar modificaciones en sus instrucciones internas o facilitar la vigilancia durante su funcionamiento (artículo 12 del Reglamento).

En principio, salvo otra previsión específica, tal conservación ha de preservarse durante seis meses (apartado seis del art. 26 del Reglamento).

Como singularidad, se reitera la conservación de tales archivos por las entidades financieras. No obstante, sabemos que existe abundante normativa sectorial que exige la conservación de documentación: en el ámbito tributario, en el ámbito contable, los registros de órdenes de operaciones bursátiles, también para prevenir el blanqueo de capitales o facilitar la lucha contra el terrorismo, determinados archivos policiales, etc.

En resumen, junto a la *lex specialis* que despliega importantes obligaciones que inciden en la organización y en las actuaciones técnicas, este Reglamento insiste en la adopción de medidas específicas para garantizar el correcto funcionamiento de los sistemas. Unas medidas que se completan con el deber de vigilancia en el que hemos de detenernos.

III.2.C.- Vigilancia y supervisión humana

El Reglamento insiste en mantener una constante vigilancia con el fin de que los sistemas se utilicen de manera adecuada, que se sigan de manera rigurosa las instrucciones, que se introduzcan datos correctos... Datos que, se señala de manera explícita, han de ser "pertinentes" y "representativos" (art. 26.4).

Los sistemas tragan con voracidad la información, la retienen, trituran, desmenuzan, consideran el amplio abanico de variables para, gracias a la potencia de la computación, ofrecer en poco tiempo resultados. El incremento de la capacidad de "supercomputación" es una de las causas del desarrollo de estos sistemas de inteligencia artificial. Pero que los frutos sean útiles dependerá, no sólo de cómo se ha configurado el sistema, también de los elementos introducidos. Informaciones erróneas, anotaciones falsas, apuntes inexactos...generarán otro desenlace distinto. En algunos ámbitos, la existencia de una insignificante variación en los datos iniciales incide de tal manera que origina propuestas drásticamente diferentes, imprevistas. Es lo que se estudia por los matemáticos y físicos bajo el nombre de

"teoría del caos", que ha de entenderse en su sentido científico y no en su acepción de desorden mayúsculo.

De ahí otra trascendental obligación: que los datos que se incorporen, siempre que se "ejerza control" sobre los mismos sean, por un lado, pertinentes y, por otro, "suficientemente representativos" (art. 26.4 del Reglamento). La obligación surge cuando existe algún control sobre los datos. Cuando estos provienen de hechos indubitados, de solicitudes presentadas, de otras circunstancias o informaciones externas sobre las que no se puede incidir la obligación queda matizada. En tales casos, habrá que actuar siguiendo meramente las instrucciones y persiguiendo la finalidad buscada.

Cuando exista alguna actividad de expurgo previo, los datos han de ser pertinentes y representativos. Pertinentes, porque han de ser ciertos y contener la información que conviene analizar para el adecuado cumplimiento de las funciones; representativos, porque se exige que sean significativos y adecuados, descartando aquellos atípicos, extraños, irregulares que pueden distorsionar la perspectiva del resultado, así como el sucesivo aprendizaje del sistema.

Como en tantas otras ocasiones, dependerá de las concretas funciones de la organización, de la complejidad del sistema y su finalidad determinar el alcance de esta obligación. En ocasiones, funcionarios o trabajadores incorporan datos fijos, únicos, que no generan especiales problemas. Sin embargo, en otros ámbitos, desde la construcción de grandes obras de ingeniería al análisis minucioso de reacciones a sustancias químicas, habrá de ponderarse cuáles son los datos más pertinentes y representativos.

Con relación a los sistemas de inteligencia artificial "generales", aquellos que no tienen un propósito específico, conviene insistir en la incorporación de datos representativos para evitar que el sistema se emborrache y revele, como explican los técnicos, "alucinaciones", errores derivados de los múltiples procesos de

análisis de probabilidades y predicciones con tal cúmulo de intoxicación por la desinformación.

Del funcionamiento del sistema ha de darse cuenta al proveedor. La justificación es comprensible. Resulta conveniente conocer si responde a las expectativas generadas y, sobre todo, cómo evoluciona. Tal información facilitará al proveedor cumplir con sus obligaciones de seguimiento dentro de la planificación cuyos contenidos básicos detallará la Comisión europea (art. 72 del Reglamento).

Si se advirtiera la presentación efectiva de un riesgo, deberá suspenderse la utilización y comunicar de inmediato al proveedor tal incidencia ("sin demora indebida" es la imprecisa expresión del Reglamento en la versión española). Esa misma información ha de remitirse a la autoridad competente con el fin de analizar los riesgos (art. 79). Autoridad con la que se exige una leal cooperación (art. 26.12).

Lógicamente cualquier incidente grave ha de comunicarse, también al distribuidor y al importador. A los efectos de la aplicación de este Reglamento se define el "incidente grave" como aquel que produce unas consecuencias de trascendental entidad: fallecimiento, daños que comprometen la salud, incumplimientos de obligaciones que protegen los derechos fundamentales, alteración crucial del funcionamiento de una infraestructura crítica y otras específicas mencionadas en el apartado 49 del artículo 3.

Como en otros preceptos, se singulariza un recordatorio en el ámbito de las entidades financieras cuya regulación más detallada impone unas especiales reglas de vigilancia pues se entiende satisfecha esta obligación de vigilancia y comunicación si se han cumplido las normas específicas ya establecidas.

También en otros muchos sectores existen precisiones detalladas tanto sobre la vigilancia del funcionamiento del negocio, como de los sistemas de comunicaciones electrónicas. Esa *lex specialis*, ley propia del sector, puede satisfacer plenamente es-

tas obligaciones que ahora recoge el Reglamento de inteligencia artificial y, en consecuencia, debería aplicarse la misma razón: su correcto seguimiento servirá para acreditar también que se satisfacen estas obligaciones. Así, por ejemplo, en todos los servicios esenciales y estratégicos que han de cumplir unas reglas de seguridad específicas y han de notificar incidentes, muchos más que los considerados en este Reglamento[28].

Sin duda se desarrollarán e incorporarán sistemas informáticos que auditen, verifiquen y supervisen otros sistemas. No obstante, la precaución ante lo imprevisto, en ámbitos donde se han considerado riesgos altos, hace que se obligue también a una supervisión humana (art. 26.2). Lo que requiere un presupuesto básico: que el proveedor del sistema lo diseñe de tal modo que permita tal supervisión (art. 14).

A los efectos que ahora nos interesan, esto es, qué obligaciones tiene el responsable del despliegue han de destacarse algunas consideraciones.

En primer lugar, se concreta el ámbito mínimo de la supervisión en estos sistemas de alto riesgo, a saber, la preocupación por cómo afectan a la "salud, seguridad y derechos o fundamentales". Dentro de la institución el responsable puede decidir lógicamente extender la verificación del funcionamiento a otros extremos pero, en estos ámbitos tan sensibles se imponen controles específicos con el propósito de reducir al mínimo tales riesgos teniendo que considerar, además, que alguien haga un "uso indebido" del sistema[29]. Ayudará, sin duda, que el proveedor especifique medidas para tal supervisión en las instrucciones.

28 Me refiero a la Directiva 2022/2555, de seguridad de las redes, conocida como NIS 2. Sobre la evolución de ese régimen jurídico y sus carencias me remito a las consideraciones que realicé en Metamorfosis..., cit.; y Heredero, C., "Nueva Directiva Europea NIS2: un avance en la regulación de la ciberseguridad para las actuales sociedades digitalizadas," *Derecho Digital e Innovación,* núm. 16, 2023, realiza un análisis de esta Directiva.

29 Vid. Gallone, G., *Riserva di umanitá e funzioni amministrative,* Wolters Kluwer, CEDAM, 2023.

En segundo lugar, resulta indispensable para garantizar la eficacia de esta obligación que quienes supervisen cuenten con la "competencia, formación y autoridad necesarias". Se establecen algunos aspectos que permiten delimitar el conocimiento de los supervisores. En concreto: han de entender de manera adecuada el sistema, su capacidad y limitaciones de tal modo que detecte y resuelva "anomalías, problemas y comportamientos inesperados"; se exige que no incurra en el "sesgo de automatización", que no confíe en exceso en el funcionamiento automático; ha de interpretar correctamente los resultados y, además, ha de contar con facultades específicas para intervenir en el funcionamiento o decidir sobre los resultados. Ámbitos sobre los que se realizan relevantes precisiones.

Porque se exige que el supervisor pueda "decidir, en cualquier situación concreta, no utilizar el sistema", así como "descartar, invalidar o revertir los resultados de salida" que se hayan generado e, incluso, "interrumpir el sistema" bien pulsando un botón de parada u otro procedimiento que permita que se detenga de manera segura (letras d y e del art. 14.4). Ello supone unos conocimientos técnicos especiales para, como hemos visto, detectar y resolver anomalías, problemas y comportamientos inesperados.

Tales previsiones han de superar un obstáculo: en la actualidad existe una notable carencia de profesionales con conocimientos informáticos suficientes para ocupar la demanda de puestos de trabajo tanto en las Administraciones públicas como en las empresas[30].

Sin perjuicio de que en los próximos años se incremente de manera muy significativa la formación de especialistas, anoto dos observaciones. Es la primera que, como en tantas ocasiones a lo largo de la lectura del Reglamento, dependerá de la comple-

30 Entre los muchos documentos e informes que resaltan esta carencia, me remito al publicado por la Asociación DigitalEs, que agrupa a las principales empresas españolas de telecomunicaciones, tecnología e innovación digital: "Anatomía de la brecha de talento tecnológico" publicado en mayo de 2024.

jidad interna del sistema de inteligencia artificial la exigencia de la concreta capacitación del supervisor. Algunos sistemas calificados de alto riesgo siguen unas pautas absolutamente uniformes, como sujetos a raíles ferroviarios, siempre las mismas, sin capacidad de modular otras circunstancias, de tal modo que puede resultar más asequible esa actividad de supervisión del funcionamiento. Otros, sin embargo, incorporan tal sofisticación, tantísima información, tal cúmulo de posibilidades, de valoración de alternativas, que resulta complejo desentrañar hasta qué punto, en un proceso, el sistema se ha "salido del libro", esto es, ha creado, ha inventado otra pauta que es la que ha llevado al resultado.

Situaciones singulares como la perplejidad de matemáticos al intentar comprender todos los pasos que ha dado un sistema de inteligencia artificial para explicar proposiciones irresolubles durante siglos o por qué se ganaron y por qué se perdieron algunas partidas del juego del go en el mítico enfrentamiento entre AlphaGo y Sedol han sido bien narradas[31]. Pueden considerarse acontecimientos especiales, anecdóticos, pero no hay que descartar que la extraordinaria celeridad en cómo se incrementa la complejidad de los sistemas reduzca la posibilidad de explicar sus pasos.

En fin, deberá intensificarse la formación de especialistas para garantizar esa supervisión cuya falta, en caso de originarse daños, generará la correspondiente exigencia de responsabilidad[32].

31 Vid. Satoy, M., *Programados para crear. Cómo está aprendiendo a escribir, pintar y pensar la inteligencia artificial,* Acantilado, 2020; y Labatut, B., *Maniac,* Anagrama, 2023.

32 Coincido por ello con Ponce Solé, J. que hace años ya advirtió de estas exigencias de supervisión y responsabilidad, "Reserva de humanidad y supervisión humana de la inteligencia artificial", *El Cronista,* núm. 100, págs. 58 y ss.; y "Seres humanos e inteligencia artificial: discrecionalidad artificial, reserva de humanidad y supervisión humana" dentro de la obra colectiva *Inteligencia artificial*..., cit. págs. 196 y ss.

III.2.D.- Información a los afectados

Ha de informarse a los ciudadanos del uso de sistemas de alto riesgo cuando les afecten los resultados de los procesos. Noticia que ha de darse, incluso, cuando se ha utilizado el sistema como mera ayuda a lo largo del procedimiento para adoptar la decisión (art. 26.11).

Esta previsión requiere, sin embargo, varias puntualizaciones.

Hace años que el Reglamento de protección de datos ha reconocido, con carácter general, el derecho a los ciudadanos de no ser objeto de una decisión que les afecte basada de manera exclusiva en un proceso automatizado (art. 22 RGPD). Derecho que el mismo precepto matizaba al admitir decisiones automatizadas en varias situaciones, entre otras, cuando existe un consentimiento previo, así como cuando esté autorizado tal proceso por el Derecho de la Unión Europea (apartado segundo)[33].

Por tanto, los sistemas de alto riesgo que incidan en datos personales encontrarán su amparo en este Reglamento si satisfacen tales exigencias del consentimiento previo o una habilitación europea.

Además, hay actuaciones y procesos en los que se utilizan tales sistemas de alto riesgo y que no inciden en esos datos personales, de tal modo que podrán ser utilizados y, en ese caso, ha de informarse. Como en aquellas otras ocasiones en que el procedimiento ha sido impulsado o ha contado con la participación voluntaria de las personas: análisis de pruebas físicas, resolución de solicitudes de ayudas, convocatorias de oposiciones, asesoramiento financiero...

Sin embargo, no surge tal deber de información en el ámbito de la protección de las entidades e infraestructuras críticas, ni

33 Vid. Ballesteros, L.A., *Las fronteras de la privacidad. El conflicto entre la seguridad jurídica y los datos personales en una sociedad amenazada y tecnológica*, Comares, 2020; así como Gamero E., "Sistemas automatizados en la toma de decisiones en el Derecho administrativo español", *RGDA*, núm. 63, 2023.

tampoco cuando el Derecho de la Unión haya previsto excepciones o restricciones específicas, como establece el artículo 86 del propio Reglamento de inteligencia artificial. Tal es el caso de las previsiones establecidas en el ámbito de las investigaciones policiales y la justicia penal (Directiva 2016/680, de 26 de abril).

Interesa subrayar que la obligación de informar no queda circunscrita únicamente a señalar que se ha utilizado un sistema de alto riesgo. Deberá indicarse la finalidad, el ámbito de la ayuda o decisión, así como, en determinados casos, una explicación "clara y significativa" (art. 86 del Reglamento).

Es cierto que este precepto tiene carácter supletorio, pues "se aplicará únicamente" si no existe otra regulación comunitaria y, sobre todo, puede quedar desplazado ante excepciones y restricciones previstas en el Derecho de la Unión (86.2).

En conclusión, cuando no exista otra *lex specialis,* habrá de informarse y aclararse a los afectados por qué se confía en ese sistema, en ese avance en los procesos de análisis o decisión: qué ventajas aporta frente al anterior modo de actuar más tradicional, qué experiencia se ha conseguido... comentarios que pueden realizarse de manera previa o posterior a su utilización y que resultarán insuficientes si no se completan con "explicaciones claras y significativas" sobre el resultado. Porque explicar implica esclarecer qué denota el resultado obtenido, algo que lógicamente ha de realizarse *ex post.*

IV.- LARGAS MORATORIAS Y DERECHO TRANSITORIO

La *vacatio legis* se extendió, como es frecuente en el Derecho de la Unión, durante veinte días tras la publicación en el Diario Oficial. Sin embargo, se ha decidido posponer la aplicación de este nuevo régimen jurídico. Se ha establecido una larga moratoria,

con matices, precisiones y excepciones que retrasarán el despliegue efectivo del Reglamento.

Así, con carácter general, será exigible a los dos años de su entrada en vigor, el dos de agosto de 2026.

Sin embargo, y aquí está la primera precisión, la aplicación de los dos primeros capítulos que aluden a las disposiciones generales, así como a las prácticas prohibidas, será efectiva a los seis meses de la entrada en vigor, el 2 de febrero de 2025. Pero, una excepción importante, se excluye el régimen de calificación de los sistemas de alto riesgo para los que se ha fijado un plazo de tres años. Plazo también con varios matices.

Existen especialidades con relación a otros capítulos pues se aplicará el Reglamento al año de su entrada en vigor, según establece el artículo 113, con relación a: la nueva organización en la Unión Europea, esto es, la creación de la Oficina de inteligencia artificial, la constitución del Consejo europeo de inteligencia artificial y del Grupo científico de expertos, así como de un Foro consultivo (capítulo VII); la obligación de que los Estados miembros designen autoridades de supervisión (sección IV del capítulo III); la obligación de confidencialidad (art 78); el régimen de los sistemas generativos o de "propósito general" (capítulo V); el régimen sancionador con excepción del régimen del artículo 101 que se refiere a los sistemas generales o de propósito general (capítulo XII). En esos casos, el régimen jurídico será exigible a partir del 2 de agosto de 2025 (art. 113).

Interesa igualmente saber que el Reglamento tiene una eficacia retroactiva media. Esto es, con relación a los sistemas de inteligencia artificial existentes, así como los que aparezcan durante esos meses hasta la efectiva aplicación, quedan sujetos al régimen de transitoriedad establecido en el artículo 111. La finalidad: facilitar una adaptación progresiva a las nuevas obligaciones, buscar cierto equilibrio, cierta proporcionalidad entre el esfuerzo y coste que implica construir algunos sistemas,

su extensa comercialización y difusión, y la seguridad jurídica que ofrece el eficaz cumplimiento de este régimen jurídico.

Así, todos los "operadores", también por tanto quienes ahora nos interesan, los responsables del despliegue, deberán garantizar este marco jurídico europeo con relación a los sistemas calificados de alto riesgo cuando, después de la moratoria, una actualización del sistema implique "cambios significativos de diseño", esto es, modificaciones substanciales que trastoquen su finalidad, incidan en elementos no previstos inicialmente en su diseño y, en consecuencia, afecten al cumplimiento de los requisitos esenciales exigidos por el Reglamento en los artículos 9 y ss.

Los responsables del despliegue deberán atender, en principio, al contenido de las actualizaciones de los sistemas de alto riesgo para advertir, en caso de que no lo indique de manera explícita el proveedor, cómo afectan los cambios al sistema, si son sustanciales.

Este régimen transitorio apunta otras precisiones relevantes mediante una cadena de excepciones, particularidades y puntualizaciones.

Porque se establece el objetivo de que las autoridades públicas deberán satisfacer las exigencias del Reglamento relativas a todos los sistemas de alto riesgo que utilicen a más tardar a los seis años de la entrada en vigor (2 de agosto de 2030). Esto afectará a aquellos que se hayan introducido durante estos meses de moratoria y no sufran modificaciones sustanciales. Sin embargo, y empiezan las reglas especiales, con relación a los sistemas de alto riesgo que se utilicen en las actividades de seguridad a las que alude el anexo X, esto es, sistemas de información Schengen, de información de visados, sistemas de interoperabilidad para la cooperación judicial y policial y otros semejantes. En concreto, se ha fijado la fecha del 31 de diciembre de 2030 para su total adaptación al Reglamento con relación a aquellos sistemas que se comercialicen o pongan a disposición durante

el plazo de tres años desde la entrada en vigor del Reglamento. Esto es, la moratoria inicial de dos años se prorroga un año más, hasta mediados de 2027. Aunque, nuevo matiz, cuando se actualice la regulación de esa normativa de seguridad citada en el anexo X, tales sistemas serán objeto de evaluación.

A mi juicio, resultan plazos muy largos: seis años. De ahí que la Comisión Europea haya impulsado un "Pacto" con el fin de planificar la aplicación del Reglamento, en la medida de lo posible, con antelación. Con este fin ha convocado a los interesados para difundir las prácticas que se están analizando e incorporando, así como fomentar compromisos y códigos de conducta[34].

Es cierto que la celeridad con la que se producen los avances tecnológicos generará, por un lado, que a los nuevos sistemas que aparezcan dentro de dos años se les aplicará íntegramente el Reglamento. Por otro lado, que también las actualizaciones de los sistemas se suceden con notable rapidez y, en caso de que afecten a elementos sustanciales, en caso de modificaciones esenciales, implicará su sujeción al Reglamento.

En este sentido, y como mero elemento de comparación, recuerdo las fechas de evolución de uno de los sistemas generativos conocidos, el ChatGPT. Según las noticias que se han publicado sobre su proceso de desarrollo relativas a los parámetros básicos de las sucesivas versiones, el diseño inicial se concretó en 2018. Al año siguiente, ya se había configurado una segunda versión que también se actualizó tras otro año de entrenamiento y que se presentó públicamente en noviembre de 2022. A partir de ahí, hay una aceleración en los cambios, pues en unos meses, en marzo del 23, hay una nueva versión, y tras otros meses de entrenamiento, en diciembre se bautiza otra.

En resumen, el desarrollo y los cambios son constantes y, por ello, tras la moratoria los sistemas de inteligencia artificial deberán someterse a este régimen jurídico. Pero ello no ha de minorar

34 Vid. https://digital-strategy.ec.europa.eu/es/policies/ai-pact

la preocupación por el desparrame durante esta larga moratoria porque además de ser continuos los cambios, lo trascendente es que se producen con extraordinaria celeridad. Nuestros tiempos, nuestra percepción del tiempo se volatiliza al advertir las referencias de la supercomputación.

¿Por qué facilitar que durante dos años funcionen sin restricciones sistemas de alto riesgo? ¿Por qué permitir que durante seis meses se usen sistemas que se considerarán prohibidos? ¿Se es consciente de los billones de cálculos que en un solo día realiza una computadora?[35]

Sin dejar de lado estas inquietudes, hay que realizar unas últimas consideraciones porque no he concluido con las obligaciones de los responsables del despliegue de los sistemas de inteligencia artificial. Hemos de detenernos en otro aspecto.

V.- CONTRIBUCIÓN AL DESARROLLO DE LOS SISTEMAS

Las instituciones que incorporan sistemas de inteligencia artificial en el ejercicio de sus funciones o en su negocio pueden quedar sujetas a otras obligaciones, además de las que acabamos de repasar. En concreto, a las obligaciones, lógicamente más rigurosas, establecidas para los proveedores (art. 25 del Reglamento). Y ello porque se entiende que en sus comportamientos están "desarrollando" ciertamente el sistema, están acrecentando sus aplicaciones y ese incremento de capacidad del sistema impone una responsabilidad mayor.

35 Excede al propósito de este trabajo incidir en qué ámbitos resulta más peligrosa la moratoria de algunos sistemas de alto riesgo. Pero ha de tenerse mínima noción de lo que suponen los procesos de supercomputación. Las unidades de esas anotaciones se llaman "comas flotantes" a los que alude el glosario del Reglamento de inteligencia artificial (art. 3.67) y la básica, "peta-FLOTS-día" implica mil billones de cálculos por segundo. Un ejemplo expresivo que facilita Jordi Torres (2023) para comparar la celeridad es que un ordenador personal necesitaría, al menos, un año para alcanzar ese peta-FLOTS-día.

¿En qué situaciones, quien ha adquirido un sistema y lo utiliza, se sujeta a las obligaciones de los proveedores? En aquellos casos que, tras comprobar su funcionamiento, advertir las ventajas de su uso, los efectos prácticos, repara en otras posibilidades del sistema, en la facultad de nuevas utilidades, en la viabilidad de otros fines... y ello le lleva a incorporar nuevas instrucciones de tal manera que lo modifica de manera sustancial.

Como ya he señalado en otras ocasiones anteriores a lo largo de este texto, son tan dispares los sistemas de inteligencia artificial que, a efectos de esta regulación, lo primero que hay que advertir es que algunos sistemas generan internamente constantes cambios y alteraciones, a raíz del mayor cúmulo de datos introducidos o de una combinación de los mismos, pues tienen tales instrucciones previas para ese aprendizaje. En este sentido, importa subrayar que aquellos parámetros predeterminados por el proveedor en la evaluación inicial y que originarán cambios a lo largo de su funcionamiento no son considerados "modificaciones sustanciales" a los efectos de incrementar las obligaciones de los responsables del despliegue (art. 43.4 del Reglamento).

En el glosario de conceptos que recoge el artículo 3, aparecen delimitadas las "modificaciones sustanciales" como los cambios que no estén previstos o planificados en la evaluación inicial y por los cuales, bien se altera la finalidad original o bien afectan a "la conformidad del sistema" con los requisitos establecidos para estos sistemas de alto riesgo tales como los presupuestos de ciberseguridad, solidez, gestión de riesgos, transparencia y otros que se exigen en los artículos 8 y siguientes del Reglamento (apartado 23).

Para que ciertamente opere una mayor responsabilidad de las Administraciones, empresas o profesionales resulta necesario que haya una decisión de añadir nuevas instrucciones al sistema. Las nuevas cargas y responsabilidades han de derivar de una previa voluntad consciente de modificación. Del mismo modo que se somete a esas obligaciones a la institución que

añade su nombre o marca al sistema. Esa confesada muestra de que entiende que está contribuyendo a su desarrollo, probablemente por cómo está nutriendo de datos al sistema, genera la mayor responsabilidad.

Junto a esta conciencia, hay otro dato objetivo clave que determina ese salto hacia una mayor responsabilidad: el propio sistema de inteligencia artificial advierte de ese cambio sustancial. Y es que, entre los requisitos del diseño, el Reglamento exige que estos sistemas cuenten con la capacidad de indicar tal cambio (art. 12).

En todo caso, hay que insistir en que los cambios e instrucciones en el sistema han de incidir en elementos sustanciales del sistema para que puedan originar nuevas obligaciones. Añadidos que pudieran calificarse de formales, auxiliares, accesorios, no pueden conducir a imponer obligaciones como si estas instituciones fueran ciertamente diseñadoras, proveedoras de los sistemas. Un requisito que ha de interpretarse en sentido riguroso por la consecuencia de incrementar la responsabilidad. Manifestaciones de esta contribución serán las que resulten de realizar pruebas y ensayos en espacios controlados.

Las nuevas obligaciones se extienden a ofrecer mayor información, realizar evaluaciones, entre otras que son objeto de análisis específico en otro trabajo de esta misma revista firmado por Ricardo Rivero. Me remito al mismo.

Concluyo con una precisión: el uso de los sistemas de inteligencia artificial se ha extendido notablemente, ciertamente estamos como sumergidos en ese entramado de procesos automáticos. Como si al navegar por Internet los sistemas de inteligencia artificial nos han empujado a bucear y, a partir de ahí, estamos ya rodeados de agua. Esperemos que no nos introduzcan en un saco amniótico.

Por ello termino con una advertencia: para la realización de este estudio no he utilizado sistemas de inteligencia artificial. No he

consultado a ningún asistente informático, a ningún sistema de generación de contenido. Pero sé que, una vez que este texto se digitalice, habrá sistemas que lo incorporen. Les deseo una buena digestión si ello contribuye a promover el debate entre juristas sobre la necesaria regulación de la inteligencia artificial. Porque somos los juristas, junto con otros técnicos y profesionales, quienes tenemos que precisar el régimen jurídico de los sistemas para poder confiar en su uso.

VI.- CONCLUSIONES

La publicación de trabajos en esta Revista de privacidad y Derecho digital exige terminar con unas sucintas conclusiones enumeradas. De manera escueta recuerdo algunas de mis consideraciones:

1º Todos los usuarios de sistemas de inteligencia artificial hemos de respetar unas mínimas normas éticas, así como seguir los códigos de conducta e instrucciones que se difundan de tales sistemas.

2º Las Administraciones, empresas y profesionales que incorporen sistemas en el ejercicio de sus funciones, negocios o actividades han de garantizar unas pautas de transparencia para que los ciudadanos estén advertidos de su uso.

3º Además, cuando tales sistemas sean calificados de alto riesgo, deberán adoptar específicas medidas organizativas, de carácter técnico, ser objeto de supervisión humana e informar a quienes afecten tales procesos de su utilización.

4º El Reglamento de inteligencia artificial prevé una moratoria extensa para exigir su efectiva aplicación. Cosa que, a mi juicio, puede poner en riesgo el esfuerzo conseguido de sistematizar este régimen jurídico.

Revista de Privacidad y Derecho Digital
ISSN digital 2444-5762, NÚM. 34, Madrid, Mayo-Agosto (2024), págs. 121-171

5º He reflexionado y he escrito este trabajo sin ninguna asistencia de sistema inteligencia artificial pero soy consciente de que la digitalización del mismo será alimento para alguno de esos sistemas. Espero que tal digestión sea nutritiva con el fin de generar un debate entre juristas sobre la regulación de la inteligencia artificial.

VII.- BIBLIOGRAFÍA

Son numerosos los libros y artículos que analizan el régimen jurídico de los sistemas de inteligencia artificial. A continuación únicamente completo los datos de las referencias mencionadas a lo largo del texto y me remito, pues su lectura ilustrará, a los números monográficos de las Revistas El Cronista del Estado social y democrático de Derecho, núm. 100/2022; a la Revista Jurídica de Asturias, núm. 45/2022; al libro colectivo dirigido por Eduardo Gamero y coordinado por Francisco Pérez Guerrero, *Inteligencia artificial y sector público: retos, límites y medios,* Tirant lo Blanch 2023; así como a las ponencias y comunicaciones presentadas en el Congreso de la Asociación Española de Profesores de Derecho Administrativo celebrado en Vigo en 2024.

ALONSO GARCÍA, C., "Sistema Viogén: fallos y algunas propuestas de mejora", Congreso de la Asociación Española de Profesores de Derecho Administrativo, Vigo, 2024

AVARO, D., El Sistema de Crédito Social chino. Vigilancia, paternalismo y autoritarismo, Ed. Biblos, Buenos Aires, 2023.

BALLESTEROS MOFFA, L.A., *Las fronteras de la privacidad. El conflicto entre la seguridad jurídica y los datos personales en una sociedad amenazada y tecnológica*, Comares, 2020.

BELTRÁN DE HEREDIA, I., *Inteligencia artificial y neuroderechos: la protección del yo inconsciente de la persona*, Ed. Aranzadi, 2023.

BOIX PALOP, A. Y SORIANO ARNANZ, A., "Transparencia y control de uso de la inteligencia artificial por las Administraciones públicas", en la obra coordinada por F. Balaguer Callejón y L. Cotino Hueso *Derecho público de la inteligencia artificial*, Fundación Giménez Abad, Zaragoza, 2023.

COTINO, L., "Discriminación, sesgos e igualdad de la inteligencia artificial en el sector público" en la obra colectiva dirigida por Eduardo Gamero, *Inteligencia artificial y sector público: retos, límites y medios,* Tirant lo Blanch, págs. 260 y ss., 2023.

DOMÍNGUEZ ÁLVAREZ, J.L., *Iusdata y Administración pública,* Civitas, 2023.

FUERTES, M., *Metamorfosis del Estado. Maremoto digital y ciberseguridad,* Marcial Pons, 2022.

GALLONE G., *Riserva di umanitá e funzioni amministrative,* Wolters Kluwer, CEDAM, 2023.

GAMERO E., "Sistemas automatizados en la toma de decisiones en el Derecho administrativo español", *RGDA*, núm. 63, 2023.

GARCÍA MEXÍA, P., "Inteligencia artificial: una mirada desde el Derecho", *Anales de la Academia Matritense del Notariado*, tomo 60, pág. 117, 2020.

HEREDERO, C., "Nueva Directiva Europea NIS2: un avance en la regulación de la ciberseguridad para las actuales sociedades digitalizadas", *Derecho Digital e Innovación,* núm. 16, 2023.

LABATUT, B., *Maniac,* Anagrama, 2023.

MARTÍN DELGADO, I., "La aplicación del principio de transparencia a la actividad administrativa algorítmica", en la obra colectiva Eduardo Gamero, *Inteligencia artificial y sector público: retos, límites y medios,* Tirant lo Blanch, págs. 132 y ss., 2023.

MENÉNDEZ, E., *From bureaucracy to artificial intelligence: the tension between effectiveness and guarantees,* Wolters Kluwer, CEDAM, 2023.

MIRANZO DÍAZ, J. *Inteligencia artificial y Derecho Administrativo*, ICA-Tecnos, 2023

OLIVER, N., *Inteligencia artificial, naturalmente*, ONTSI 2020

PONCE, J., "Reserva de humanidad y supervisión humana de la inteligencia artificial", *El Cronista,* núm. 100, págs. 58 y ss. 2022;

"Seres humanos e inteligencia artificial: discrecionalidad artificial, reserva de humanidad y supervisión humana" dentro de la obra colectiva *Inteligencia artificial y sector público*, coord. por E. Gamero y F.L. Guerrero, Tirant lo Blanch, págs. 196 y ss., 2023.

SATOY, M., *Programados para crear. Cómo está aprendiendo a escribir, pintar y pensar la inteligencia artificial,* Acantilado, 2020.

TORRES, J., *La inteligencia artificial explicada a los humanos,* Plataforma editorial, 2023.

SISTEMAS DE IA PROHIBIDOS, DE ALTO RIESGO, DE LIMITADO RIESGO, O DE BAJO O NULO RIESGO[1] (*)

PROHIBITED, HIGH-RISK, LIMITED RISK, OR MINIMAL OR NO RISK AI SYSTEMS

Por MARTÍN MARÍA RAZQUIN LIZARRAGA

Catedrático de Derecho Administrativo.
Universidad Pública de Navarra

(*) Este trabajo se recibió el 28 de mayo de 2024 y fue aceptado en septiembre.

1 Este trabajo se enmarca en el Proyecto "Biometría, Derecho Administrativo y Datos -BIODATA", PID2021-125170NB-I00, financiado por MCIN/AEI/10.13039/501100011033/ y por FEDER Una manera de hacer Europa, del que soy investigador principal.

REVISTA DE
PRIVACIDAD Y DERECHO DIGITAL

RESUMEN

La LIA regula los sistemas de IA desde la perspectiva del riesgo y por tanto los clasifica en sistemas de IA prohibidos, de alto riesgo o de bajo o nulo riesgo. Quedan prohibidas las prácticas de IA que supongan riesgos inadmisibles, salvo algunas excepciones. La categoría principal es la de sistemas de IA de alto riesgo, que vienen enumerados en función de su inclusión en normas sobre productos o en razón de criterios horizontales. Se permite su introducción en el mercado, comercialización y uso de los sistemas de IA de alto riesgo, previo cumplimiento de diversos requisitos y obligaciones relevantes para los operadores. Por el contrario, los sistemas de bajo o nulo riesgo quedan sólo vinculados al cumplimiento voluntario de códigos de conducta.

PALABRAS CLAVE: *Ley de Inteligencia Artificial, inteligencia artificial, sistemas de IA prohibidos, sistemas de IA de alto riesgo.*

ABSTRACT

AIA regulates AI systems from a risk perspective and therefore classifies them into prohibited, high risk, or low or no risk AI systems. AI practices that suppose unacceptable risks are prohibited, with some exceptions. The main catogery is high-risk AI systems, which are listed based on their inclusion in products standard or based on horizontal criteria. High-risk systems are permitted to be introduced into the market, commercialized and used, subject to compliance with various relevant requirements and obligations for operators. On the contrary, minimal or no risk systems are only linked to voluntary compliance with codes of conduct.

KEY WORDS: *Artificial Intelligence Act, artificial intelligence, prohibited AI systems, high-risk AI systems.*

SUMARIO

I.- INTRODUCCIÓN: ENFOQUE Y OBJETO DE LA LIA

Una explicación de los sistemas de IA desde la perspectiva del riesgo, no puede entenderse sin atender al enfoque que la LIA efectúa de su regulación sobre la inteligencia artificial. Sólo desde esta premisa inicial e integral podrá comprenderse por qué se ha adoptado la perspectiva del riesgo y los condicionamientos que se irán estableciendo respecto de cada una de las categorías de sistemas de IA. Y dicha premisa es el valor superior de la persona sobre la inteligencia artificial[2].

2 La posición de la Comisión Europea, ahora plasmada en la LIA, proviene de los diversos documentos que ha ido aprobando hasta llegar a su propuesta de LIA. Por citar los

El Considerando 1 de al LIA muestra este enfoque "de una inteligencia artificial (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 (en lo sucesivo, «Carta»), incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, proteger frente a los efectos perjudiciales de los sistemas de IA en la Unión, así como brindar apoyo a la innovación". Asimismo, su Considerando 2 insiste en esta premisa: "El presente Reglamento debe aplicarse de conformidad con los valores de la Unión consagrados en la Carta, lo que facilitará la protección de las personas físicas, las empresas, la democracia, el Estado de Derecho y la protección del medio ambiente y, al mismo tiempo, impulsará la innovación y el empleo y convertirá a la Unión en líder en la adopción de una IA fiable". Y en su Considerando 6 afirma que "como requisito previo, la IA debe ser una tecnología centrada en el ser humano. Además, debe ser una herramienta para las personas y tener por objetivo último aumentar el bienestar humano".

más cercanos y relevantes: la Comunicación "Generar confianza en la inteligencia artificial centrada en el ser humano" (COM 2019-640, de 8.4.2019), el "Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza" (de 19.2.2020) y "Una Estrategia Europea de Datos" (de 19.2.2020). Así la Comunicación citada afirma lo siguiente: "La Estrategia europea de IA y el plan coordinado dejan claro que la confianza es un requisito previo para garantizar un enfoque de la IA centrado en el ser humano: la IA no es un fin en sí mismo, sino un medio que debe servir a las personas con el objetivo último de aumentar su bienestar. Para ello, la fiabilidad de la IA debe estar garantizada. Los valores en los que se basan nuestras sociedades han de estar plenamente integrados en la evolución de la IA. La Unión se fundamenta en los valores de respeto de la dignidad humana, la libertad, la democracia, la igualdad, el Estado de Derecho y el respeto de los derechos humanos, incluidos los derechos de las personas pertenecientes a minorías. Estos valores son comunes a las sociedades de todos los Estados miembros, en las que prevalecen el pluralismo, la no discriminación, la tolerancia, la justicia, la solidaridad y la igualdad. Además, la Carta de los Derechos Fundamentales de la UE reúne, en un único texto, los derechos individuales, civiles, políticos, económicos y sociales de que gozan los ciudadanos de la UE" (pág. 2).

Sólo teniendo en cuenta este enfoque ético[3] se podrán entender las prácticas prohibidas de IA, por ser consideradas como inadmisibles para el ser humano, así como los requisitos y obligaciones impuestas para los sistemas de IA de alto riesgo, principalmente, la evaluación de impacto.

Partiendo de dicha premisa, procede advertir desde el principio que la LIA no es una ley general de los sistemas de IA, puesto que no regula todos ellos y menos todos sus aspectos. El enfoque adoptado del riesgo provoca que la regulación de la LIA sea limitada en cuanto a su objeto, puesto que disciplina sólo aquellos sistemas de IA que tengan riesgo inadmisible o elevado, dejando fuera de su regulación el resto de los sistemas de IA. Dentro de este riesgo se incluyen también los sistemas y modelos de IA de uso general, conocidos como IA generativa, en la medida que son sistemas que dada su amplitud, profundidad y alcance pueden comportar un elevado riesgo. No obstante, en la presente exposición no se van a analizar estos sistemas y modelos de IA de uso general, por ser objeto de otro Estudio.

Sin embargo, en contrapartida, la LIA afecta a todos los sectores que utilizan o pueden utilizar un sistema de IA[4].

El enfoque del riesgo viene ya precisado en el artículo 1.2 LIA que, aunque afirma como primer objetivo, el de establecer normas armonizadas para la introducción, la puesta en servicio y la utilización de sistemas de IA en la Unión, sin embargo, seguidamente precisa que va a establecer prohibiciones de determinadas

3 SALAZAR GARCÍA, I., explica que el primer reto ético de la inteligencia artificial es el del ser humano como centro, siendo ésta "la premisa de las principales organizaciones e instituciones, a nivel internacional, que estudian la regulación ética y normativa de la IA" ("Retos actuales de la ética en la Inteligencia Artificial", en *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, pág. 59).

4 SIMÓN CASTELLANO resalta la concepción y visión amplia de la LIA por su enfoque horizontal, de modo que no hay un *numerus clausus* de sectores a los que afecte ("Allende una teoría general de las garantías jurídicas para una inteligencia artificial confiable", en *Derecho Público de la Inteligencia Artificial*, Fundación Manuel Giménez Abad, Zaragoza, 2023, pág. 119).

prácticas de IA y los requisitos específicos para los sistemas de IA de alto riesgo y obligaciones para los operadores de dichos sistemas.

Una segunda advertencia es necesaria. La LIA parte de la relación a veces inescindible entre sistemas de IA y productos. Como es bien sabido, la LIA es la primera regulación sobre IA de Europa, e incluso del mundo (a salvo de China y de EEUU). Sin embargo, la UE ha venido aprobando, desde hace tiempo, un importante número de disposiciones normativas relativas a los productos, como puede verse en la larga relación contenida en el Anexo I de la LIA. Pues bien, el artículo 2.2 limita, de forma casi total[5], la aplicación de la LIA para aquellos sistemas de IA de alto riesgo que estén asociados a productos que estén sometidos al sistema de armonización total de la UE, en concreto, la referenciada en la sección B del Anexo I. El art. 2.3 contiene, también, algunas exclusiones, como la relativa a los sistemas de IA en materia de seguridad nacional, fines militares o de defensa.

Así pues, la LIA no regula la "fabricación" o producción de sistemas de IA, sino solamente su comercialización (introducción al mercado o posterior) y su uso, pero sólo cuando genere riesgo inadmisible o alto[6]. La producción o fabricación que esté integrada en productos queda sometida a las regulaciones mencionadas en el Anexo I. Incluso cabe añadir que más propiamente la LIA regula las posibles consecuencias que se pueden derivar o pueden ser provocadas por el uso de un sistema de IA, es decir, sus efectos, que es lo que explica el enfoque del riesgo que adopta. Como afirma el Considerando 27 de la LIA, ésta adopta

5 Sólo se les aplican los arts. 6.1, 102 a 109 (disposiciones finales sobre modificación de diversos Reglamentos y Directivas europeos), y 112 (evaluación y revisión), y el art. 57 (espacios controlados de pruebas para la IA) en la medida en que los requisitos de las LIA estén integrados en la legislación de armonización.

6 Como indica SIMÓN CASTELLANO, "la prohibición por riesgo inadmisible no se refiere al diseño tecnológico, sino a la introducción en el mercado, la puesta en servicio y el uso de los sistemas de inteligencia artificial en cuestión en el conjunto de la Unión" ("Allende una teoría general", op. cit, pág. 123).

el enfoque basado en el riesgo que "es la base de un conjunto proporcionado y eficaz de normas vinculantes".

Además, aunque el objetivo de la LIA sea evitar riesgos inadmisibles o condicionar los riesgos elevados para las personas, falta una regulación de los derechos y garantías de los ciudadanos respecto de los sistemas de IA[7], y ello a pesar de que el art. 2.1 g) extiende el ámbito de aplicación a las personas afectadas que estén ubicadas en la UE. Sin embargo, en la fase legislativa se han introducido tres preceptos en orden a paliar esta deficiencia. Por un lado, desde una perspectiva más general, toda persona física o jurídica tiene derecho a presentar una reclamación ante una autoridad de vigilancia del mercado si considera que se ha infringido la LIA (art. 85). Por otro, las personas afectadas tienen derecho a exigir una explicación de las decisiones tomadas individualmente por los responsables del despliegue, cuando entiendan que tienen un efecto perjudicial para su salud, su seguridad o sus derechos fundamentales (art. 86)[8]. Y, por último, se aplica la Directiva (UE) 2019/1937 sobre protección del denunciante a las denuncias de infracciones de la LIA (art. 87).

Por último, la regulación de la LIA se acerca, a veces, más a una disposición de principios, que a una regulación precisa y concreta, por más que la propuesta de la Comisión Europea haya recibido aportaciones en la fase legislativa en orden a su mayor grado de precisión y aplicabilidad directa[9]. Ello provoca que la

7 Así lo advierten COTINO et al. ("Un análisis crítico constructivo de la propuesta de Reglamento de la Unión Europea por la que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act)", *Diario La Ley, sección Ciberderecho*, 2 de julio de 2021) y FERNÁNDEZ HERNÁNDEZ ("La futura regulación europea de la inteligencia artificial: objetivos, principios y pautas", en *Claves de inteligencia artificial y derecho*, La Ley, Madrid, 2022, págs. 129-131).

8 Bien es verdad que este derecho es subsidiario, puesto que se aplica únicamente cuando no esté previsto de otro modo en el Derecho de la UE (apartado 3 del art. 86).

9 HUERGO LORA, A., enfatiza en esta cuestión, poniendo el ejemplo que la LIA exige un sistema de gestión de riesgos que admite determinados riesgos residuales y también que la evaluación de conformidad es como regla general un autocontrol ("Gobernar con algoritmos,

LIA precisará de una concreción posterior a través de una diversidad de actuaciones de la Comisión europea o del Comité Europeo de Inteligencia Artificial[10].

II.- LA NOCIÓN DE RIESGO: CLASIFICACIÓN DE LOS SISTEMAS DE IA

La LIA adopta el enfoque del riesgo como noción central y vertebral de su regulación[11]. El riesgo viene definido en el punto 2 del artículo 3 LIA como "la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio"[12].

El Diccionario de la Lengua Española define el riesgo como contingencia o proximidad de un daño. Y considera como sinónimos los de "peligro, amenaza, ventura, risco" y como antónimo el de seguridad. Así puede verse cómo el art. 9.5 LIA se refiere a riesgos asociados a cada peligro.

gobernar los algoritmos", *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, 2022, pág. 89).

10 HERNÁNDEZ PEÑA, J. C., indica al respecto que "el reglamento recoge normas de intensidad regulatoria reducida e incluso escasa. Por tanto, es esperable que corresponda al comité -al menos parcialmente- contribuir a completar el programa normativo sectorial. Por tanto, explícitamente se le atribuye la aprobación de normas de Soft Law, con el fin de contribuir a uniformar prácticas administrativas de los Estados miembros" ("Organización y gobernanza de la inteligencia artificial: marco general", en *Inteligencia artificial y sector público. Retos, límites y medios*, Tirant lo blanc, Valencia, 2023, pág. 614).

11 COTINO HUESO, L., afirma que este enfoque del riesgo que adopta la propuesta de LIA supone "la mayor imposición de obligaciones y garantías cuanto mayor riesgo implique el tratamiento de datos o el sistema de IA" ("Nuevo paradigma en la garantía de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivos de la inteligencia artificial", en *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Mayor, 2022, pág. 93).

12 Cabe recordar en este sentido la normativa legal y reglamentaria sobre prevención de riesgos laborales. Así el artículo 4 de la Ley de prevención de riesgos laborales define el riesgo laboral en los siguientes términos: "Se entenderá como «riesgo laboral» la posibilidad de que un trabajador sufra un determinado daño derivado del trabajo. Para calificar un riesgo desde el punto de vista de su gravedad, se valorarán conjuntamente la probabilidad de que se produzca el daño y la severidad del mismo".

BECK, U., acuñó el concepto de la sociedad del riesgo en la época de la postmodernidad y la postindustrialización, puesto que desde hace tiempo la sociedad vive en una situación continuada y permanente de riesgo, que plantea incluso el dilema de la política tecnológica[13]. También ESTEVE PARDO, J., ha profundizado sobre la sociedad del riesgo, señalando que los riesgos son riesgos tecnológicos, que se deben a la intervención humana, y advierte cómo la legislación europea, especialmente la relativa a la seguridad y calidad industrial, pretende dominar, encauzar y unificar, en especial, a partir del denominado nuevo enfoque en el que ha optado por la armonización de normas técnicas[14].

La UE conoce desde hace tiempo el concepto de riesgo, así como el principio de precaución, como, por ejemplo, muestra el Derecho ambiental. Por ejemplo, La Directiva 2004/35/CE, sobre responsabilidad medioambiental en relación con la prevención y reparación de daños medioambientales está repleta de referencias al riesgo, e incluso con menciones al riesgo significativo o al bajo riesgo. Así también la Directiva 2011/92/UE, sobre evaluación de impacto ambiental ha ido incorporando referencias a los riesgos ambientales que sirven como criterio para la realización de una EIA.

También es un concepto esencial en los actos legislativos de armonización a que se refiere el Anexo I de la LIA. A título ejemplo, cabe reparar en la Directiva 2006/42/CE sobre máquinas, (derogada y sustituida por el Reglamento (UE) 2023/1230, a partir de

13 BECK, U. afirma que "los riesgos y peligros de hoy se diferencian esencialmente de los de la Edad Media (que a menudo se les parecen exteriormente) por la globalidad de su amenaza (seres humanos, animales, plantas) y por sus causas modernas. Son riesgos de la modernización. Son un producto global de la maquinaria del progreso industrial y son agudizados sistemáticamente con su desarrollo ulterior" (p. 33). Más adelante señala que los riesgos tienen que ver con la previsión, con lo que todavía no ha llegado (p. 48). Y frente al autocontrol, defiende una generalización, con ciertas garantías jurídicas, de ciertas capacidades de influencia de la subpolítica (*La sociedad del riesgo. Hacia una nueva modernidad*, Paidós, Barcelona, 2010, p. 371).

14 ESTEVE PARDO, J., se refiere a los riesgos tecnológicos derivados de la acción humana y no de causas naturales (p. 29) y al avance del Derecho europeo en aprobar normas de armonización para superar las barreras técnicas (*Técnica, riesgo y Derecho. Tratamiento del riesgo tecnológico en el Derecho ambiental,* Ariel, Barcelona, 1999, págs. 170-171).

14 de enero de 2027). Este Reglamento recoge un concepto de riesgo muy similar al de la LIA: "«riesgo»: una combinación de la probabilidad y la gravedad de una lesión o de un daño a la salud que pueda surgir en una situación peligrosa" (Anexo III, Parte A). Además, cabe referirse a la similitud respecto del control sobre aquellos productos introducidos en el mercado que entrañen riesgos para la salud o la seguridad de las personas (art. 45 del Reglamento). Por su parte, el Reglamento (UE) 2019/1020, de productos contiene una definición similar de riesgo y diferencia entre producto que presenta un riesgo y producto que presenta un riesgo grave (art. 3, apartados 18, 19 y 20).

También en el ámbito de la protección de datos personales, el RGPD utiliza el concepto de riesgo, e incluso de alto riesgo. Por un lado, en todo momento los responsables del tratamiento deben tener en cuenta los riesgos, en especial, en cuanto a su seguridad (art. 32), pero, además, en el caso de que implique un alto riesgo, deberán realizar una evaluación de impacto relativa a la protección de datos (art. 35 RGPD)[15].

El Reglamento (UE) 2022/2065 (conocido como DSA) también se refiere en numerosas ocasiones al riesgo, e incluso establece cuatro categorías de sistémico (Considerando 80).

Ya el Libro Blanco sobre la Inteligencia Artificial de la Comisión Europea de 2020 apuntaba como elemento central de una futura regulación el riesgo, que se configura como punto de equilibrio entre el fomento de la IA y la protección de las personas[16].

15 El art. 35 citado es muy ilustrativo respecto de la noción de alto riesgo: "Cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas, el responsable del tratamiento realizará, antes del tratamiento, una evaluación del impacto de las operaciones de tratamiento en la protección de datos personales. Una única evaluación podrá abordar una serie de operaciones de tratamiento similares que entrañen altos riesgos similares" (apartado 1).

16 MORAL SORIANO, L., señala que "el riesgo es la piedra angular del modelo de gobernanza ética, tal y como está recogida en el Libro Blanco de la Comisión, reflejada también en la propuesta que hace el Parlamento Europeo sobre el Marco ético de la IA, y por supuesto en el borrador de Reglamento sobre la IA de la Comisión. Un enfoque basado en los

Por tanto, como indica el Considerando 26 de la LIA, es la intensidad y el alcance de los riesgos que pueden generar los sistemas de IA, lo que determina que éstos sean clasificados en tres tipos: a) sistemas o prácticas prohibidas; y b) sistemas de alto riesgo; y c) sistemas de bajo o nulo riesgo[17]. Ello provoca que los sistemas de bajo o nulo riesgo queden, en la práctica, fuera de la regulación de la LIA[18]. La doctrina, y también la Comisión europea en su página web, consideran como una categoría diferenciada la de los sistemas de IA de riesgo limitado, referida a aquellos sistemas para los que se imponen obligaciones de información y transparencia (art. 50 LIA)[19]. Respecto de los sistemas o modelos de IA de uso general se contempla otra categoría, la del peligro sistémico, definido en el art. 3.65 LIA, que no va a ser objeto de análisis aquí, por corresponder su examen a otro Estudio.

riesgos asegura, como sostiene la Comisión, una intervención proporcionada: la equidistancia necesaria entre el principio de precaución y el principio de innovación, si es que la UE quiere ser un *hub* de IA" ("Modelos de gobernanza global de la inteligencia artificial", en *Inteligencia artificial y Derecho. El jurista ante los retos de la era digital,* Aranzadi, Cizur Menor, 2021, pág. 248). HERNÁNDEZ PEÑA, en idéntica línea, apunta la vinculación de los riesgos con el principio de precaución, y afirma que esta posición está alineada con lo dispuesto en el RGPD (*El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza*, Aranzadi, Cizur Menor, 2022, pág. 100).

17 COTINO HUESO, L., lo ejemplifica muy bien, indicando que "teóricamente el AIA supone un sistema de semáforo. Rojo: prohíbe algunos usos de IA (art. 5). Amarillo: fija algunos usos de "alto riesgo" (art. 6 y Anexos II y III). Verde: no es obligatorio cumplir la regulación del AIA" ("Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal", en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, 2022, pág. 72).

18 FERNÁNDEZ HERNÁNDEZ afirma que para la Comisión Europea la gran mayoría de los sistemas de IA implican un riesgo bajo o mínimo, con base en el documento de la Comisión titulado Nuevas normas sobre la inteligencia artificial: preguntas y respuestas, aunque a la vista del Anexo III se pone de relieve que el campo de aplicabilidad de la LIA es amplísimo ("La futura regulación europea de la inteligencia artificial: objetivos, principios y pautas", *Claves de inteligencia artificial y derecho*, op. cit., p. 144).

19 Es por ello que SIMÓN CASTELLANO se refiera a cuatro clases de sistemas de IA, añadiendo la categoría de sistema de IA de riesgo limitado para referirse a estos supuestos del art. 50 ("Allende una teoría general", op. cit, págs. 124-125).

Así pues, puede hablarse de una construcción piramidal[20] de cinco escalones, en los que se encuentran cada una de las categorías de sistemas de IA:

1) Sistemas de IA prohibidos, ya que el riesgo es inadmisible.
2) Sistemas de IA de alto riesgo, permitidos pero sometidos al cumplimiento de requisitos y obligaciones.
3) Sistemas de IA de riesgo sistémico, referidos a los sistemas de IA de uso general, que son permitidos bajo cumplimiento de los requisitos establecidos en la LIA.
4) Sistemas de IA de riesgo limitado, permitidos y únicamente sujetos a obligaciones de transparencia e información.
5) Sistemas de IA de bajo o nulo riesgo, excluidos en la práctica de la regulación de la LIA; y respecto de los cuales los operadores pueden cumplir voluntariamente los códigos de conducta.

Conviene tener en cuenta, asimismo, el sistema de gestión de riesgos al que se refiere el art. 9, y que será abordado más adelante.

La noción central de riesgo requiere que el sistema de IA se ajuste a lo previsto en la LIA. Así el art. 79 regula el procedimiento nacional aplicable a los sistemas de IA que presenten riesgo, que exige una evaluación a fin de comprobar si el sistema se ajusta a las exigencias de la LIA. Dicho precepto debe completarse con lo dispuesto en el art. 82 que se dirige a aquellos sistemas de alto riesgo que, a pesar de ser conformes con la LIA, sin embargo, presentan riesgos, en orden a que los proveedores adopten las medidas correctoras pertinentes.

Por último, la clasificación tiene una relevante consecuencia en materia de sanciones, en orden a la determinación de su cuantía (art. 99. 3 y 4).

20 Para una explicación de la pirámide de los puestos que cada categoría ocupa en ella véase CHRISTAKIS, T. y KARRATHANASIS, T., "Tools for Navigating the EU AI Act (2) Visualisation Pyramid", AI Regulation Papers 24-03-5, AI-Regulation.com, March 8th, 2024.

III.- SISTEMAS DE IA PROHIBIDOS

Dentro de la escala de riesgo antes expuesta la LIA examina, en primer lugar, aquellos sistemas que suponen riesgos inaceptables y, por tanto, deben estar prohibidos. Sin embargo, incluso dentro de los sistemas prohibidos se efectúan algunas excepciones que permiten su utilización. El art. 1.2 de la LIA señala que establece "prohibiciones de determinadas prácticas de IA" (letra b)). Así pues, estas prácticas de sistemas de IA no se pueden comercializar ni usar, salvo que concurra alguna de las excepciones que permitan su realización.

Téngase en cuenta que lo dispuesto en el Capítulo II, compuesto únicamente por el art. 5, ambos titulados "Prácticas de IA prohibidas", será de aplicación a los 6 meses de la fecha de entrada en vigor de la LIA, que se producirá a los 20 días de su publicación en el DOUE (art. 113 LIA).

Como indica el Considerando 45 de la LIA, ésta no afecta a las prácticas prohibidas por otras normas del Derecho de la Unión Europea, singularmente en materia de protección de datos, de no discriminación, de protección de consumidores y sobre competencia. Y así el art. 5.8 LIA afirma que "El presente artículo no afectará a las prohibiciones aplicables cuando una práctica de IA infrinja otro acto legislativo de la Unión".

Resulta necesario apuntar la idea, repetida en diversos lugares por la LIA, de su complementariedad con el RGDP (y con la Directiva 2016/680) y con las demás disposiciones de la UE que dispongan normas de protección de las personas o de la competencia.

El art. 5.1 LIA contempla ocho supuestos de prácticas de IA prohibidas. Debe advertirse que se trata, por un lado, de una regulación que constituye un *numerus clausus*, dado que no se contiene una disposición que determine su posible modificación, como ocurre en el art. 7 LIA para los sistemas de alto riesgo.

Tampoco el art. 97 LIA al referirse a los actos delegados de la Comisión contiene mención alguna al art. 5.

Por otra parte, la redacción de los diferentes supuestos del art. 5.1, a pesar de la introducción durante la fase legislativa de ciertas precisiones, adolece de numerosos conceptos jurídicos indeterminados, que dejan un campo abierto, tal vez demasiado extenso, en orden a su interpretación. Y debe advertirse que la interpretación tendrá una incidencia muy relevante porque supondrá, en su caso, la consideración de un sistema de IA como prohibido o como de algo riesgo. En este sentido el Considerando 28 señala que la IA tiene usos beneficiosos y también perversos, cuando lleva a cabo prácticas perjudiciales e incorrectas de manipulación, explotación y control social, que son las que deben resultar prohibidas.

En tercer lugar, en la fase legislativa ha cobrado importancia la determinación de los sistemas biométricos, en orden a su introducción y mayor precisión, pero también en las últimas lecturas se ha realizado una interpretación más restrictiva de los supuestos de prohibición.

Con estas precisiones cabe examinar cada uno de los ocho supuestos legales de prohibición.

III.1.- USOS QUE SUPONGAN MANIPULACIÓN O ENGAÑO O ALTERACIÓN DEL COMPORTAMIENTO

La letra a) del art. 5.1 recoge como primer supuesto de usos prohibidos el siguiente: "la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una

decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas".

Su explicación se contiene en el Considerando 29, donde se examinan las posibilidades de la IA para manipular a las personas e inducirlas en su comportamiento, en suma, que provocan la pérdida de autonomía de las personas. Incluso se describen algunas técnicas de manipulación y de componentes subliminales, o las interfaces cerebro-máquina.

Esta letra a) debe ser interpretada con arreglo al Considerando 29, que permite efectuar varias observaciones. La primera consistente en que, dado que el enfoque es el de riesgo, en este caso de un riesgo inadmisible, no tiene importancia la conducta del proveedor o del responsable del despliegue, sino que "el perjuicio se derive de las prácticas de manipulación o explotación que posibilita la IA".

La segunda que, dado que se trata de evitar la anulación de la autonomía personal[21], no están prohibidas aquellas prácticas que cuenten con el consentimiento explícito de las personas o de sus representantes legales.

En tercer lugar, debe darse una relación de causalidad entre el sistema de IA y el objetivo o efecto que provoca la prohibición[22].

Incluso para el caso de prácticas comerciales y legítimas habrá de demostrarse que incurren en esta manipulación, puesto que por sí mismas, siempre que cumplan el Derecho aplicable, no son prácticas de manipulación perjudiciales a los efectos del art. 5 LIA.

21 Así las Directrices para una IA fiable de 2019 recogían como primer principio ético el respeto de la autonomía humana.

22 HERNÁNDEZ PEÑA señala que "un aspecto objeto de controversia es la causalidad entre la exposición a un sistema de IA y el impacto sobre el comportamiento, de forma tal que se llegue a distorsionar o perturbar el comportamiento" (*El marco jurídico…*, op. cit., pág. 124).

III.2.- USOS DE MANIPULACIÓN DE PERSONAS VULNERABLES

La letra b) del art. 5.1 contiene un segundo supuesto de este tenor literal: "la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que explote alguna de las vulnerabilidades de una persona física o un determinado colectivo de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona o a otra".

Es fácil de percibir que este segundo supuesto constituye una variante agravada del primero[23]. Aquí se acentúa la prohibición por tratarse de personas vulnerables, en razón de su edad, discapacidad, situación social o económica, o pertenencia a una minoría étnica o religiosa.

III.3.- USOS PARA LA EVALUACIÓN O CLASIFICACIÓN DE PERSONAS FÍSICAS

El tercer supuesto de uso prohibido de IA se explica en la letra c) del art. 5.1. Dicho supuesto puede ser dividido en dos partes: la premisa y sus circunstancias, toda vez que no todos los usos que incurran en lo que se determina en la premisa quedan prohibidos, sino solamente aquellos que incurran en alguna de las situaciones perjudiciales o desfavorables para las personas. Así también se trata de proteger no sólo a las personas físicas individualmente consideradas sino también a los colectivos de personas, que pueden padecer estas evaluaciones o clasificaciones negativas o despectivas como tales colectivos. Se trata de evitar sistemas de "crédito social" (*social scoring*), como los existentes en China[24].

23 De ahí que también sea examinada conjuntamente con el primer supuesto en el Considerando 29 LIA.

24 Así lo indica HERNÁNDEZ PEÑA (*El marco jurídico…*, págs. 125-127).

La premisa es la siguiente: "la introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA para evaluar o clasificar a personas físicas o a colectivos de personas durante un período determinado de tiempo atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas…".

Como indica el Considerando 31 sólo deben prohibirse "los sistemas de IA que impliquen esas prácticas inaceptables de puntuación y den lugar a esos resultados perjudiciales o desfavorables. Esa prohibición no debe afectar a prácticas lícitas de evaluación de las personas físicas que se efectúen para un fin específico de conformidad con el Derecho de la Unión y nacional".

Tal como se ha advertido más arriba, de nuevo se trata de un supuesto limitado, puesto que debe concurrir alguna de las dos condiciones de perjuicio a las personas o colectivos que se relatan en el citado art. 5.1 c) y que dicen así:

"i) un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos de personas en contextos sociales que no guarden relación con los contextos donde se generaron o recabaron los datos originalmente,

ii) un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos de personas que sea injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este".

Así pues, cualquier otra práctica de IA de evaluación o clasificación de personas o colectivos de personas que no comporte estos efectos perjudiciales o desfavorables no se encuentra prohibida.

III.4.- USOS PARA EVALUACIONES DE RIESGOS DE COMISIÓN DE DELITOS

La LIA acoge como cuarto supuesto de prácticas prohibidas de IA la relativa a la realización de evaluaciones de riesgos de personas físicas en orden a la comisión de delitos. En concreto la

letra d) del art. 5.1 dice así: "la introducción en el mercado, la puesta en servicio para este fin específico o el uso de un sistema de IA para realizar evaluaciones de riesgos de personas físicas con el fin de valorar o predecir el riesgo de que una persona física cometa un delito basándose únicamente en la elaboración del perfil de una persona física o en la evaluación de los rasgos y características de su personalidad".

La explicación de esta prohibición se contiene en el Considerando 42 LIA: "En consonancia con la presunción de inocencia, las personas físicas de la Unión siempre deben ser juzgadas basándose en su comportamiento real. Las personas físicas nunca deben ser juzgadas a partir de comportamientos predichos por una IA basados únicamente en la elaboración de sus perfiles, en los rasgos o características de su personalidad, como la nacionalidad, el lugar de nacimiento, el lugar de residencia, el número de hijos, el nivel de endeudamiento o el tipo de vehículo, sin una valoración humana y sin que exista una sospecha razonable, basada en hechos objetivos comprobables, de que dicha persona está implicada en una actividad delictiva".

Este precepto plantea diversas cuestiones. En primer lugar, la prohibición alcanza tanto a los sistemas de IA que se hayan creado con este fin específico, como también a aquellos que permitan en su utilización conseguir este resultado. En segundo término, la referencia a delitos excluye las infracciones administrativas. Y, en tercer término, se acota a la elaboración de perfiles o rasgos o características de la personalidad, cuando este sea el fundamento único del sistema de IA, lo que puede significar que en otro caso (es decir, cuando no sea "únicamente") dicho sistema no está prohibido.

De nuevo este supuesto contiene una excepción consistente en que no están prohibidos los usos de IA en el caso de personas sospechosas, cuando la sospecha se fundamente en hechos objetivos y verificables. Como señala el propio precepto: "esta prohibición no se aplicará a los sistemas de IA utilizados para apoyar la

valoración humana de la implicación de una persona en una actividad delictiva que ya se base en hechos objetivos y verificables directamente relacionados con una actividad delictiva".

III.5.- USOS DE RECONOCIMIENTO FACIAL

La letra e) del art. 5.1 LIA comienza con los supuestos de usos prohibidos de sistemas biométricos, que ha sido una de las grandes preocupaciones de la fase legislativa.

El art. 3.34 LIA define qué se entiende por datos biométricos: "los datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física, como imágenes faciales o datos dactiloscópicos". Estos datos se encuadran dentro de las categorías especiales de datos del art. 9 del RGPD[25] y del art. 10 de la Directiva 2016/680 (a la que se remite el apartado 37 del art. 3 LIA), cuyo tratamiento está prohibido salvo que concurra alguna de las bases jurídicas expresadas en dichos preceptos.

Como señala el Considerando 14, este concepto de datos biométricos debe interpretarse conforme a la normativa europea de protección de datos antes citada. Este Considerando diferencia los diferentes usos de los datos biométricos: 1) Autenticación; 2) Identificación; 3) Categorización; y 4) Reconocimiento de emociones.

Así el primer supuesto está dedicado al reconocimiento facial: "la introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante la extracción

25 El art. 4 apartado 14 del RGPD contienen la siguiente definición de datos biométricos: "datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona, como imágenes faciales o datos dactiloscópicos".

no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión".

La LIA no prohíbe cualquier reconocimiento facial, sino solamente aquel que se de en alguna de las condiciones descritas en esta letra e). Como señala el Considerando 43 se protege a las personas frente a la vigilancia masiva, por lo que se hace referencia a una extracción no selectiva.

Esta consideración de vigilancia masiva explica las condiciones limitativas de este supuesto para la extracción de imágenes: internet o circuitos cerrados de televisión.

Pero, además, lo que se prohíbe es la creación o ampliación de bases de datos con estas imágenes no selectivas, no el uso inmediato de las imágenes que no comporte incorporación de las mismas a una base de datos.

III.6.- SISTEMAS DE INFERENCIA DE EMOCIONES

Este sexto supuesto de uso prohibido recoge un nuevo caso de uso de los datos biométricos, ahora con la finalidad de detectar o deducir emociones. Como señala el Considerando 44 de la LIA, su prohibición se fundamenta en que pueden tener resultados discriminatorios e invadir los derechos y libertades de las personas afectadas, en especial, en determinados contextos[26].

El art. 3 apartado 39 lo define así: "«sistema de reconocimiento de emociones»: un sistema de IA destinado a distinguir o inferir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos"[27].

26 El Parlamento Europeo impulsó la redacción de un informe sobre esta cuestión, que le fue entregado en septiembre de 2021 y lleva como título "Biometric Recognition and Behaviour Detection", siendo sus autores externos WENDEHORST, C., y DULLER, Y.

27 El Considerando 18 LIA enumera las emociones que pueden reconocerse: "El concepto de «sistema de reconocimiento de emociones» a que hace referencia el presente Reglamento debe definirse como un sistema de IA destinado a distinguir o deducir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos. El

La redacción de la letra f) del art. 5.1 LIA es la siguiente: "la introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de IA para inferir las emociones de una persona física en los lugares de trabajo y en los centros educativos".

La inferencia de emociones se limita en este supuesto a los lugares de trabajo o centros educativos, no a otros ámbitos, y ello porque el carácter intrusivo de estos sistemas de inferencia de emociones tiene un resultado inadmisible en estos ámbitos.

Nuevamente, esta disposición recoge una excepción a este supuesto de prohibición: "excepto cuando el sistema de IA esté destinado a ser instalado o introducido en el mercado por motivos médicos o de seguridad".

Ambos motivos de excepción deben tener como punto común la protección de las personas, de ahí que se trata de aspectos de salud (motivos médicos) o de seguridad (prevención de riesgos laborales, por ejemplo).

III.7.- USOS DE CATEGORIZACIÓN BIOMÉTRICA

El apartado 40 del art. 3 LIA define qué se entiende por categorización biométrica: "un sistema de IA destinado a incluir a las personas físicas en categorías específicas en función de sus datos biométricos, a menos que sea accesorio a otro servicio comercial y estrictamente necesario por razones técnicas objetivas".

concepto se refiere a emociones o intenciones como la felicidad, la tristeza, la indignación, la sorpresa, el asco, el apuro, el entusiasmo, la vergüenza, el desprecio, la satisfacción y la diversión. No incluye los estados físicos, como el dolor o el cansancio, como, por ejemplo, los sistemas utilizados para detectar el cansancio de los pilotos o conductores profesionales con el fin de evitar accidentes. Tampoco incluye la mera detección de expresiones, gestos o movimientos que resulten obvios, salvo que se utilicen para distinguir o deducir emociones. Esas expresiones pueden ser expresiones faciales básicas, como un ceño fruncido o una sonrisa; gestos como el movimiento de las manos, los brazos o la cabeza, o características de la voz de una persona, como una voz alzada o un susurro".

Como señala el Considerando 16 de la LIA, estas categorías pueden referirse "a aspectos como el sexo, la edad, el color del pelo, el color de los ojos, los tatuajes, los rasgos conductuales o de la personalidad, la lengua, la religión, la pertenencia a una minoría nacional o la orientación sexual o política".

No tienen esta consideración determinados usos por su nota de accesoriedad a otro uso o servicio principal, tal como observa el considerando 16, con algún ejemplo.

Pues bien, los sistemas de IA que tengan como finalidad específica o permitan el uso de sistemas de categorización biométrica están prohibidos, tal como establece la letra g) del art. 5.1 LIA: "la introducción en el mercado, la puesta en servicio para este fin específico o el uso de sistemas de categorización biométrica que clasifiquen individualmente a las personas físicas sobre la base de sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida sexual u orientación sexual"[28].

Como puede verse, la protección alcanza a las finalidades o usos que afectan a datos de categoría especial del RGPD, que son los datos especiales y más relevantes de las personas[29].

También aquí este supuesto contiene una excepción: "esta prohibición no incluye el etiquetado o filtrado de conjuntos de datos biométricos adquiridos lícitamente, como imágenes, basado en datos biométricos ni la categorización de datos biométricos en el ámbito de la garantía del cumplimiento del Derecho".

El Considerando 30 señala como supuestos excluidos de la prohibición "la clasificación de imágenes en función del color del

28 Así se explica en el Considerando 30 LIA.

29 Procede recordar que ya el RGPD prohíbe el tratamiento de datos personales con esta finalidad en su art. 9.1: "Quedan prohibidos el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientación sexuales de una persona física".

pelo o del color de ojos, que pueden utilizarse, por ejemplo, en el ámbito de la garantía del cumplimiento del Derecho".

III.8.- SISTEMAS DE IDENTIFICACIÓN BIOMÉTRICA REMOTA EN TIEMPO REAL EN ESPACIOS DE ACCESO PÚBLICO

Este supuesto se revela como uno de los más debatidos dentro del uso de sistemas biométricos. Venía incorporado en la propuesta de la Comisión europea, pero ha recibido varias aportaciones en la fase legislativa, que por un lado precisan este supuesto y por otro lo acotan.

Por otra parte, la LIA se ha hecho eco de la crítica social a los sistemas masivos de control de las personas, uno de cuyos ejemplos es el de China, que se trata de evitar a toda costa en la Unión Europea.

El supuesto requiere, en primer término, una explicación del mismo, que viene dada en las definiciones que recoge la LIA. El apartado 35 del art. 3 LIA explica, por un lado, lo que es la identificación biométrica: "el reconocimiento automatizado de características humanas de tipo físico, fisiológico, conductual o psicológico para determinar la identidad de una persona física comparando sus datos biométricos con los datos biométricos de personas almacenados en una base de datos".

El Considerando 15 de la LIA se ocupa de definir con mayor precisión el concepto de identificación biométrica y su diferenciación con la verificación o autenticación biométricas[30]. Es esencial

30 Resulta, por su importancia, obligado transcribir el Considerando 15 citado: "El concepto de «identificación biométrica» a que hace referencia el presente Reglamento debe definirse como 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, a fin de determinar la identidad de una persona comparando sus datos biométricos 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. Quedan excluidos los sistemas de IA destinados a la verificación biométrica, que comprende la autenticación, cuyo único propósito es confirmar que una persona física

entender estas diferencias para comprender que la LIA sólo se refiere en el art. 5 a los sistemas de identificación y no a los de verificación y autenticación.

Y luego los apartados 41 a 44 del art. 3 LIA explican con detalle qué se entiende por cada uno de los conceptos adicionales:

1) Sistema de identificación biométrica remota (se explica en el Considerando 17).
2) Sistema de identificación biométrica remota en tiempo real (se explica en el Considerando 17).
3) Sistema de identificación biométrica en diferido (se explica en el Considerando 17).
4) Espacio de acceso público (se explica en el Considerando 19).

La letra h) del art. 5.1 LIA impone la prohibición del "uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de garantía del cumplimiento del Derecho". El fundamento de esta prohibición se recoge en el Considerando 32: "El uso de sistemas de IA para la identificación biométrica remota «en tiempo real» de personas físicas en espacios de acceso público con fines de garantía del cumplimiento del Derecho invade de forma especialmente grave los derechos y las libertades de las personas afectadas, en la medida en que puede afectar a la vida privada de una gran parte de la población, provocar la sensación de estar bajo una vigilancia constante y disuadir indirectamente a los ciudadanos de ejercer su libertad de reunión y otros derechos fundamentales. Las imprecisiones técnicas de los sistemas de IA destinados a la identificación biométrica remota de las personas físicas pueden dar lugar a resultados sesgados y tener efectos discriminatorios. Tales posibles resultados sesgados y efectos discriminatorios son especialmente pertinentes por lo que respecta a la edad, la

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".

etnia, la raza, el sexo o la discapacidad. Además, la inmediatez de las consecuencias y las escasas oportunidades para realizar comprobaciones o correcciones adicionales en relación con el uso de sistemas que operan «en tiempo real» acrecientan el riesgo que estos conllevan para los derechos y las libertades de las personas afectadas en el contexto de actividades de garantía del cumplimiento del Derecho, o afectadas por estas".

Puede advertirse de la "excepcionalidad" de la propia salvedad, que provoca que la LIA impida incluso la obtención de determinadas consecuencias de su uso cuando éste sea posible de forma excepcional, tal como indica el inciso final del apartado 3: "Dicha autoridad no podrá adoptar ninguna decisión que produzca efectos jurídicos adversos para una persona exclusivamente sobre la base de los resultados de salida del sistema de identificación biométrica remota «en tiempo real»".

Como se desprende de la literalidad del precepto, debe tratarse de identificación biométrica (no verificación o autenticación) y, además, deben concurrir tres aspectos esenciales:

1) Que la identificación sea remota. De ahí que no alcance a la identificación en cercanía, es decir, la que se realice con consciencia de la persona física como cuando se acerca voluntariamente a un sistema de identificación. Ello hace referencia, por ejemplo, a las grabaciones indiscriminadas de imágenes de las personas.
2) Que la identificación sea en tiempo real. Por tanto, su diferenciación con la identificación en diferido, tal como explica el Considerando 17.
3) Que la identificación sea en espacios de acceso público, lo que no alcanza a otro tipo de espacios de acceso no público, sean de titularidad pública o privada[31]. Aquí se hace

31 Así lo indica la definición de espacio de acceso público contenida en el art. 3.44 LIA, tras la precisión añadida en el procedimiento legislativo. En cambio, HERNÁNDEZ PEÑA consideraba que "No parece que la definición incluya instalaciones privadas de uso público,

> referencia especial al acceso a las calles o lugares de tránsito público, como también aeropuertos o estaciones de transporte terrestre o marítimo.

Así pues, en este supuesto caracterizado con las notas antedichas, el uso de la identificación biométrica recibe una prohibición absoluta.

Sin embargo, la LIA contempla un supuesto excepcional que permite utilizar la identificación biométrica remota en espacios de acceso público con fines de garantía del cumplimiento del Derecho. Este supuesto excepcional constituye una "lex specialis", aplicable frente a lo dispuesto en la normativa de protección de datos personales, singularmente, el art. 10 de la Directiva (UE) 2016/680. Por tanto, lo dispuesto en el artículo 10 citado, así como en el art. 9.1 del RGPD sólo se aplica a los demás tratamientos de datos personales diferentes de la citada excepción (apartado final de la letra h) y Considerando 39).

No obstante, esta excepción se somete a unos requisitos muy estrictos, con el objetivo de limitar al máximo estos usos con la finalidad expresada, en orden a evitar un Estado policial de continuada vigilancia sobre los ciudadanos.

La LIA establece importantes límites y condiciones para permitir el uso de este tipo de prácticas.

El más importante y previo es su aceptación por los Estados. Para poder hacer uso de la posibilidad contenida en la excepción antedicha, los Estados deben introducir en su normativa interna la posibilidad de utilizar, total o parcialmente, esta excepción (apartado 5, y también apartado 2, segundo párrafo)[32].

como podría ser el caso de universidades o clínicas privadas, así como comercios abiertos al público, entre otros" (*El marco jurídico…*, op. cit., pág. 130).

32 Como señala el Considerando 37 este uso excepcional es posible "cuando el Estado miembro de que se trate haya decidido contemplar expresamente la posibilidad de autorizarlo en las normas detalladas de su Derecho nacional, y en la medida en que lo haya contemplado". Y el Estado miembro puede acogerse a la totalidad de los objetivos de la excepción o sólo a algunos de ellos.

Y si así lo hicieran, comunicarán a la Comisión las normas adoptadas al respecto en el plazo de 30 días siguientes.

Por otra parte, la regulación que efectúa la LIA sobre esta excepción de uso de sistemas de identificación biométrica es una regulación de mínimos y por tanto obligada para los Estados. No obstante, éstos podrán adoptar leyes más restrictivas sobre el uso de sistemas de identificación biométricas (apartado 5 *in fine*).

Veamos, pues, los límites para el uso de esta excepción.

En primer lugar, no basta cualquier fin de garantía del cumplimiento del Derecho, sino que debe perseguir alguno de los objetivos específicos establecidos en el apartado 1), párrafo primero, de la letra h), siempre que se cumplan los límites y condiciones que se fijan en los apartados siguientes. Como expresa el Considerando 33, la salvedad persigue lograr un interés público esencial cuya importancia compense los riesgos.

Y es preciso constatar que el uso de estos sistemas de IA constituye tratamiento de datos biométricos (Considerando 38, párrafo 1º)[33].

Aquí comienza el acotamiento de los tres supuestos excepcionales previstos. En primer lugar, se contiene un límite general, aplicable a todos ellos, "en la medida en que dicho uso sea estrictamente necesario". Y seguidamente se describe cada uno de los objetivos que permiten salvar la prohibición.

Los límites y condiciones cabe enumerarlos del siguiente modo:

1) Concurrencia de uno de los tres objetivos señalados en la letra h) del apartado 1.
2) Necesidad de una finalidad específica: confirmar la identidad de la persona que constituya el objetivo específico, y con las condiciones de las letras a) y b) del apartado 2.

33 De ahí que la regulación de la LIA sea calificada como *lex specialis* respecto del art. 10 de la Directiva (UE) 2016/680.

3) Evaluación de impacto de protección de datos y registro en la base de datos de la UE[34].

4) Autorización previa[35] de autoridad judicial o de autoridad administrativa independiente[36]. A tal fin se fijan condiciones para la concesión de dicha autorización previa.

5) Notificación del uso a la autoridad de vigilancia del mercado y a la autoridad nacional de protección de datos. Estas autoridades deben efectuar informes anuales a la Comisión sobre este uso. La Comisión realizará informes anuales con dicha información agregada.

Los objetivos que habilitan para el uso de esta excepción son los tres siguientes:

"i) la búsqueda selectiva de víctimas concretas de secuestro, trata de seres humanos o explotación sexual de seres humanos, así como la búsqueda de personas desaparecidas,

ii) la prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas

34 Excepcionalmente el registro en la base de datos puede efectuarse *a posteriori*.

35 El Considerando 33 excluye de esta autorización previa determinados controles de identidad: "Además, el presente Reglamento debe preservar la capacidad de las autoridades garantes del cumplimiento del Derecho, de control fronterizo, de la inmigración o del asilo para llevar a cabo controles de identidad en presencia de la persona afectada, de conformidad con las condiciones establecidas en el Derecho de la Unión y en el Derecho nacional para estos controles. En particular, las autoridades garantes del cumplimiento del Derecho, del control fronterizo, de la inmigración o del asilo deben poder utilizar sistemas de información, de conformidad con el Derecho de la Unión o el Derecho nacional, para identificar a las personas que, durante un control de identidad, se nieguen a ser identificadas o no puedan declarar o demostrar su identidad, sin que el presente Reglamento exija que se obtenga una autorización previa. Puede tratarse, por ejemplo, de una persona implicada en un delito que no quiera revelar su identidad a las autoridades garantes del cumplimiento del Derecho, o que no pueda hacerlo debido a un accidente o a una afección médica".

36 Excepcionalmente puede utilizarse el sistema con carácter de urgencia sin autorización previa, siempre que la misma se solicite inmediatamente dentro del plazo de 24 horas. No obstante, también se contempla que en caso de denegación de la autorización debe interrumpirse de inmediato el uso y además desecharse y suprimirse los datos obtenidos (Considerando 35).

físicas[37] o de una amenaza real y actual o real y previsible de un atentado terrorista,

iii) la localización o identificación de una persona sospechosa de haber cometido una infracción penal a fin de llevar a cabo una investigación o un enjuiciamiento penales o de ejecutar una sanción penal por alguno de los delitos mencionados en el anexo II[38] que en el Estado miembro de que se trate se castigue con una pena o una medida de seguridad privativas de libertad cuya duración máxima sea de al menos cuatro años[39]".

IV.- SISTEMAS DE IA DE ALTO RIESGO

La LIA presta especial y amplia atención a la regulación de los sistemas de IA de alto riesgo. Frente al art. 5 que es la única disposición dedicada a los sistemas de IA prohibidos, el Capítulo III relativo a los sistemas de IA de alto riesgo está compuesto de 44 artículos, e incluso los Capítulos siguientes se refieren en gran medida a este tipo de sistemas. Ello permite concluir que la regulación de los sistemas de IA de alto riesgo constituye el eje central de la LIA.

37 Según el Considerando 33 también se engloba en esta excepción la perturbación grave de las infraestructuras críticas, definidas en el art. 2.4 de la Directiva (UE) 2022/2557.

38 El Anexo II enumera los siguientes delitos: "terrorismo, trata de seres humanos, explotación sexual de menores y pornografía infantil, tráfico ilícito de estupefacientes o sustancias psicotrópicas, tráfico ilícito de armas, municiones y explosivos, homicidio voluntario, agresión con lesiones graves, tráfico ilícito de órganos o tejidos humanos, tráfico ilícito de materiales nucleares o radiactivos, secuestro, detención ilegal o toma de rehenes, delitos que son competencia de la Corte Penal Internacional, secuestro de aeronaves o buques, violación, delitos contra el medio ambiente, robo organizado o a mano armada, sabotaje, y participación en una organización delictiva implicada en uno o varios de los delitos enumerados en esta lista".

39 El número de 4 años es un mínimo, que puede ser elevado por los Estados miembros, tal como señala expresamente el párrafo final del Considerando 33.

Dentro del Capítulo III la clave se encuentra en su primera Sección que determina qué sistemas de IA se clasifican como de alto riesgo, que viene complementada por los Anexos I y III. A partir de ello, se fijan, luego, los requisitos (Sección 2) y las obligaciones (Secciones 3, 4 y 5).

Los sistemas de IA de alto riesgo se pueden introducir en el mercado y utilizar, siempre que reúnan los requisitos exigidos. Sólo en aquellos supuestos en que el riesgo no supere la evaluación de conformidad o, en su caso, la evaluación de impacto (art. 27), los sistemas de IA de alto riesgo no podrán ser utilizados.

IV. 1.- CLASIFICACIÓN DE SISTEMA DE ALTO RIESGO

IV.1.A.- La noción de alto riesgo: criterios de determinación

Como se ha advertido más arriba, la LIA parte de la noción de riesgo para su regulación, de modo que hay cuatro clases de sistemas de IA, además de la de riesgo sistémico para los sistemas de uso general: a) sistemas prohibidos por ser inadmisibles dado su riesgo; b) sistemas que se pueden autorizar con los requisitos y obligaciones que se determinan por tener un riesgo alto; c) sistemas de riesgo limitado, sujetos a obligaciones de transparencia; y d) sistemas de IA que quedan prácticamente fuera de la regulación de la LIA, dado que su nivel de riesgo es bajo o nulo. Procede recordar que la noción de riesgo es un cálculo de probabilidad, tal como define el art. 3 LIA. Y es dicho cálculo en razón de la intensidad y alcance de los riesgos el que provoca su clasificación, en este caso, su encuadramiento como sistemas de alto riesgo (Considerando 26).

La primera idea que se desprende de la LIA es que la clasificación de un sistema como de alto riesgo es restrictiva, es decir, sólo cabe su encuadre en esta categoría cuando el sistema se encuentre incluido dentro de las determinaciones expresadas en su art. 6. Lo enuncia de forma clara el Considerando 46 al

afirmar que "La clasificación de un sistema de IA como «de alto riesgo» debe limitarse a aquellos sistemas de IA que tengan un efecto perjudicial considerable en la salud, la seguridad y los derechos fundamentales de las personas de la Unión, y dicha limitación debe reducir al mínimo cualquier posible restricción del comercio internacional".

El concepto de "alto riesgo" se encuentra vinculado a la "magnitud" o "gravedad" de sus consecuencias o perjuicios adversos (Considerandos 47 y 48), en relación con los derechos fundamentales, y con especial atención a los menores y también al medio ambiente. Se trata de que los productos que se introduzcan en el mercado o se comercialicen sean seguros y conformes.

La dificultad de definir la noción de alto riesgo justifica que el art. 6.5 ordene a la Comisión la aprobación de Directrices, junto con ejemplos prácticos, de cuáles sean sistemas de alto riesgo y cuáles no lo sean.

La LIA diferencia entre dos tipos de sistemas de IA para su encuadramiento como sistemas de alto riesgo. Por un lado, los que están vinculados o forman parte de un producto, y, por otro, los que son independientes de un producto. Como señala el art. 6.2, hay dos supuestos de sistemas de alto riesgo: a) los del apartado 1, vinculados a productos; y b) los del apartado 2, independientes de productos y relacionados en el Anexo III.

IV.1.B.- Sistemas de IA de alto riesgo vinculados a un producto: el Anexo I

El primer supuesto de sistemas de IA de alto riesgo es el de aquellos sistemas que son componentes de un producto o incluso constituyen por sí mismos un producto. Los Considerandos de la LIA ponen diversos ejemplos: componentes de robots de uso en fábricas o en sanidad (Considerando 47) o ciertos productos como "máquinas, juguetes, ascensores, equipo y sistemas de protección para uso en atmósferas potencialmente

explosivas, equipos radioeléctricos, equipos a presión, equipos de embarcaciones de recreo, instalaciones de transporte por cable, aparatos que queman combustibles gaseosos, productos sanitarios, productos sanitarios para diagnóstico *in vitro, automoción y aviación"* (Considerando 50).

El art. 6.1 LIA comienza con una advertencia que indica la amplitud de este supuesto: "Con independencia de si se ha introducido en el mercado o se ha puesto en servicio sin estar integrado en los productos que se mencionan en las letras a) y b)". De ahí que se utilice el concepto de vinculación al producto, puesto que no precisa estar integrado en el mismo, sino solamente que pueda ser utilizado en relación con un producto.

La Directiva de máquinas establecía la definición de componente de seguridad[40]. En el futuro dicha definición será sustituida por la contenida en el Reglamento (UE) 2023/1230: "un componente físico o digital, incluido el software, de un producto incluido en el ámbito de aplicación del presente Reglamento que esté diseñado o destinado a desempeñar una función de seguridad y que se introduzca en el mercado por separado, cuyo fallo o funcionamiento defectuoso ponga en peligro la seguridad de las personas, pero que no sea necesario para que dicho producto funcione o cuyos componentes normales puedan ser sustituidos para que dicho producto funcione" (art. 3, apartado 3).

Ahora el art. 2.14 LIA define los componentes de seguridad en los siguientes términos: "un componente de un producto o un sistema que cumple una función de seguridad para dicho producto o

40 Su art. 2 c) contiene la siguiente definición de componente de seguridad: "componente:
- que sirva para desempeñar una función de seguridad,
- que se comercialice por separado;
- cuyo fallo y/o funcionamiento defectuoso ponga en peligro la seguridad de las personas, y
- que no sea necesario para el funcionamiento de la máquina o que, para el funcionamiento de la máquina, pueda ser reemplazado por componentes normales.
En el anexo V figura una lista indicativa de componentes de seguridad que podrá actualizarse con arreglo al artículo 8, apartado 1, letra a)".

sistema, o cuyo fallo o defecto de funcionamiento pone en peligro la salud y la seguridad de las personas o los bienes".

Los productos son los regulados en la normativa de la UE, que viene explicitada en el Anexo I de la LIA. Así pues, este Anexo, dividido en dos Secciones A y B, recoge la doble lista de actos legislativos de armonización de la UE, que obedece a la categorización de productos sometidos al denominado *old approach* o sistema de armonización total o a los componentes de seguridad de productos sometidos al nuevo enfoque o *New Framework Legislation*[41]. Procede recordar que a los sistemas de IA referidos a la Sección B sólo se les aplica la LIA de forma muy limitada, únicamente en lo relativo a sus arts. 6.1, 102 a 109 (disposiciones finales sobre modificación de diversos Reglamentos y Directivas europeos), y 112 (evaluación y revisión), y el art. 57 (espacios controlados de pruebas) (art. 2.2).

Por tanto, para que un sistema de IA se encuadre en el art. 6 como de alto riesgo, se exige, en primer lugar (letra a) que su regulación armonizada esté comprendida en el Anexo I, en definitiva, constituye un "numerus clausus" de actos legislativos.

Como segundo requisito, el art. 6.1 requiere que el sistema de IA "deba someterse a una evaluación de la conformidad de terceros para su introducción en el mercado o puesta en servicio con arreglo a los actos legislativos de armonización de la Unión enumerados en el anexo I" (letra b)).

La exigencia de ambos requisitos (deben reunirse las dos condiciones) provoca que sólo aquellos productos para los que la regulación armonizada de la UE requiera evaluación de conformidad serán considerados como sistemas de alto riesgo. Ello supone que, si dicha regulación armonizada no exige dicha evaluación de conformidad, por considerar que el producto no entraña un riesgo elevado a los efectos de dicha normativa, no puede ser considerado como sistema de IA de alto riesgo.

41 Véase al respecto HERNÁNDEZ PEÑA, *El marco jurídico*..., op. cit., págs. 136-138.

Asimismo, es preciso tener en cuenta lo dispuesto en el art. 74.3, que permite en relación con los actos legislativos de la Sección A del Anexo I, que los Estados opten por designar otra autoridad pertinente como autoridad de vigilancia del mercado.

Finalmente, conviene advertir que este apartado 1 y las obligaciones unidas al mismo se aplicarán a los 36 meses desde la entrada en vigor de la LIA.

IV.1.C.- Sistemas de alto riesgo independientes de un producto: el Anexo III

IV.1.C.1.- Sistemas de IA de alto riesgo del Anexo III

El segundo supuesto de sistemas de IA de alto riesgo es el de aquellos sistemas incluidos en el Anexo III de la LIA. Así lo afirma el apartado 2 del art. 6: "...se considerarán de alto riesgo los sistemas de IA contemplados en el anexo III". Como puede verse, la LIA no se refiere a sectores sino a sistemas de IA[42].

El Anexo III enumera los sistemas de IA de alto riesgo a que se refiere el artículo 6, apartado 2, es decir, los que no están vinculados a productos conforme a las normas normalizadas del Anexo I. Así lo señala el Considerando 52: "En cuanto a los sistemas de IA independientes, a saber, aquellos sistemas de IA de alto riesgo que no son componentes de seguridad de productos, o que son productos en sí mismos, deben clasificarse como de alto riesgo si, a la luz de su finalidad prevista, presentan un alto riesgo de ser perjudiciales para la salud y la seguridad o los derechos fundamentales de las personas, teniendo en cuenta tanto la gravedad del posible perjuicio como la probabilidad de que se produzca, y se utilizan en varios ámbitos predefinidos especificados en el presente Reglamento".

42 HERNÁNDEZ PEÑA considera acertado que la propuesta de la Comisión Europea se apartara "de un enfoque basado exclusivamente en sectores (banca, defensa, salud, etc.), que podría suponer costes elevados y desproporcionados en aplicaciones y sistemas con un riesgo bajo para los derechos fundamentales, la salud o la seguridad de las personas" (*El marco jurídico...*, op. cit., pág. 139).

1. Biometría

El Considerando 54 afirma que los datos biométricos constituyen una categoría de datos personales sensibles, y, en definitiva, remite su aceptación a las disposiciones del RGPD y de la Directiva (UE) 2016/680. De ahí el inciso de que "su uso esté permitido por el Derecho de la Unión o nacional aplicable". A tal efecto la LIA parece entender que, en principio, los sistemas biométricos están autorizados por la normativa de protección de datos (con el cumplimiento de los requisitos en ella establecidos) y solamente califica como de alto riesgo algunos sistemas biométricos específicos por entender que pueden dar lugar a resultados sesgados y tener efectos discriminatorios[43].

Por tanto, en el caso de los sistemas biométricos de IA se imponen tres niveles: 1) Prohibición absoluta de los sistemas de identificación remota en tiempo real en espacios públicos (con la excepción ya explicada para casos de garantía de cumplimiento del Derecho); 2) sistemas de alto riesgo especificados en el punto 1 del Anexo III; y 3) demás sistemas biométricos, que son considerados como de bajo o nulo riesgo.

Los tres sistemas biométricos de alto riesgo son los siguientes:

a) Sistemas de identificación biométrica remota. Aquí se incluyen la identificación biométrica remota que no se desarrolle en espacios de acceso público y la identificación biométrica remota en diferido; y se excluyen tanto la identificación biométrica que no sea remota, por contar con la participación de las personas físicas, como la verificación biométrica cuya única finalidad sea confirmar que una persona física concreta es la persona que afirma ser. El apartado 36 del art. 3 define la verificación biométrica

43 Esta regulación da solución, cuando menos parcial, a la denuncia de COTINO HUESO en el sentido de que la regulación de la LIA y del RGPD se ignoran ("Sistemas de inteligencia artificial...", op. cit., pág. 75). Además, debe tenerse en cuenta las cautelas introducidas en la fase legislativa en la LIA, especialmente, la evaluación de impacto introducida por el art. 27 de la LIA, que habrá de compaginarse con la prevista en el RGPD.

en los siguientes términos: "la verificación automatizada y uno-a-uno, incluida la autenticación, de la identidad de las personas físicas mediante la comparación de sus datos biométricos con los datos biométricos facilitados previamente". Y la verificación biométrica es completamente diferente de la identificación biométrica, que aparece definida en el art. 3.35, más arriba transcrito. Ello provoca que en la biometría haya tres niveles: prohibición, permisión en caso de alto riesgo sometida a requisitos y obligaciones, y permisión por ser de bajo o nulo riesgo, sujeta únicamente a códigos de conducta voluntarios.

b) Sistemas de IA destinados a ser utilizados para la categorización biométrica en función de atributos o características sensibles o protegidos basada en la inferencia de dichos atributos o características. Su límite se encuentra en la prohibición de estos sistemas de categorización cuando se refieran a aquellos atributos personales referidos en la letra g) del art. 5.1 LIA.

c) Sistemas de IA destinados a ser utilizados para el reconocimiento de emociones. Aquí se incluyen los sistemas que se desarrollen en ámbitos diferentes de los lugares de trabajo o centros educativos, en cuyo caso estarían prohibidos.

2. Infraestructuras críticas

El Considerando 55 ofrece una explicación más completa de este punto. En primer término, qué debe entenderse por infraestructuras críticas[44], para lo que es necesario acudir al Anexo de la Directiva (UE) 2022/2557, modificado por el Reglamento Delegado (UE) 2023/2450. La LIA considera de alto riesgo los sistemas vinculados a las infraestructuras digitales críticas, al tráfico rodado y al suministro de agua, gas, calefacción o electricidad.

44 Conforme al art. 2.4 de la Directiva se considera infraestructura crítica: "un elemento, instalación, equipo, red o sistema, o parte de un elemento, instalación, equipo, red o sistema, que es necesario para la prestación de un servicio esencial".

Se trata de componentes de seguridad, no necesarios para el funcionamiento del sistema, pero que pueden dar lugar en caso de fallo o defecto de funcionamiento a riesgos para la salud y seguridad de las personas.

3. Educación y formación profesional

Dentro del ámbito educativo se consideran sistemas de IA de alto riesgo sólo algunos de los que pueden ser utilizados, dado que la gran mayoría fomentarán la formación digital y la adquisición de capacidades digitales.

La característica común a los cuatro supuestos catalogados como de alto riesgo en el ámbito educativo viene explicitada en el Considerando 56: evitar la discriminación de las personas, en especial de las pertenecientes a colectivos vulnerables (mujeres, mayores, discapacitadas, de determinado origen racial o étnico, o personas con una determinada orientación sexual).

Con base en este objetivo de no discriminación se consideren de alto riesgo, en resumen, cuatro casos:

a) Determinación del acceso o admisión a centros educativos[45].

b) Evaluación de resultados de aprendizaje.

c) Evaluación del nivel de educación adecuado a recibir o acceder.

d) Control de exámenes.

4. Empleo, gestión de los trabajadores y acceso al autoempleo

También en este punto, el objetivo de la clasificación de determinados sistemas como de alto riesgo obedece a una finalidad

45 Parece que se ha tenido presente el caso del algoritmo utilizado en el Reino Unido de valoración de estudiantes para acceso a las Universidades y que tuvo que ser retirado a causa de los sesgos en contra de los estudiantes de las escuelas públicas (HERNÁNDEZ PEÑA, *El marco jurídico…*, op. cit., pág. 141).

de evitar la discriminación, sobre todo de determinados colectivos vulnerables (Considerando 57).

Y así se consideran sistemas de IA de alto riesgo los dos siguientes:

a) Acceso al empleo: anuncios, solicitudes de empleo y evaluación de candidatos.

b) Condiciones de trabajo, promoción y despido: cuando se trate de tomar como base comportamientos individuales o características personales para supervisión o evaluación[46].

5. Acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones

Se trata de garantizar el acceso y el disfrute de servicios, tanto privados como públicos, que se consideran esenciales para las personas.

Entre los servicios públicos se encuentran todos aquellos que pueden englobarse bajo el amplio epígrafe de sanidad y asistencia social (asistencia sanitaria, prestaciones de seguridad social contributiva y no contributiva, dependencia, maternidad, enfermedad, dependencia, asistencia social y ayudas a la vivienda). Así también se incluye la solicitud y respuesta a llamadas de emergencia realizadas por personas físicas en servicios como policía, bomberos y servicios de asistencia médica, y en sistemas de triaje de pacientes en el contexto de la asistencia sanitaria de urgencia.

En los servicios privados esenciales, se incluyen los relativos a evaluaciones de solvencia crediticia y para seguros de vida y de salud.

46 Téngase en cuenta la letra d) el art. 64.4 del Texto Refundido del Estatuto de los Trabajadores, introducida por el Real Decreto Ley 9/2021 y confirmada por la Ley 12/2021, de 28 de septiembre, que reconoce como un derecho del comité de empresa: "Ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles".

Procede recordar que el Considerando 58 incluye una importante cautela: "No obstante, el presente Reglamento no debe obstaculizar el desarrollo y el uso de enfoques innovadores en la Administración, que podrían beneficiarse de una mayor utilización de sistemas de IA conformes y seguros, siempre y cuando dichos sistemas no conlleven un alto riesgo para las personas jurídicas y físicas".

6. Garantía del cumplimiento del Derecho

También en ese supuesto se precisa con carácter previo que el uso de sistemas de IA con esta finalidad esté autorizado por el Derecho de la UE o el de los Estados miembros. Además, los sistemas de IA sólo pueden ser utilizados por las autoridades encargadas del cumplimiento del Derecho o en su nombre, bien sean estatales o de la UE.

Cinco son los sistemas de IA considerados como de alto riesgo[47]:

a) Sistemas utilizados para la evaluación del riesgo de que una persona física sea víctima de delitos.

b) Polígrafos o herramientas similares.

c) Sistemas de IA para evaluar la fiabilidad de las pruebas durante la investigación o el enjuiciamiento de delitos.

47 El Considerando 59 explica los motivos de su inclusión como sistemas de alto riesgo: "Dado su papel y su responsabilidad, las actuaciones de las autoridades garantes del cumplimiento del Derecho que implican determinados usos de los sistemas de IA se caracterizan por un importante desequilibrio de poder y pueden dar lugar a la vigilancia, la detención o la privación de libertad de una persona física, así como tener otros efectos negativos sobre los derechos fundamentales consagrados en la Carta. En particular, si el sistema de IA no está entrenado con datos de buena calidad, no cumple los requisitos adecuados en términos de rendimiento, de precisión o de solidez, o no se diseña y prueba debidamente antes de introducirlo en el mercado o ponerlo en servicio, es posible que señale a personas de manera discriminatoria, incorrecta o injusta. Además, podría impedir el ejercicio de importantes derechos procesales fundamentales, como el derecho a la tutela judicial efectiva y a un juez imparcial, así como el derecho a la defensa y a la presunción de inocencia, sobre todo cuando dichos sistemas de IA no sean lo suficientemente transparentes y explicables ni estén suficientemente bien documentados".

d) Sistemas de IA para evaluar la probabilidad de que una persona física cometa un delito o reincida en la comisión de un delito atendiendo no solo a la elaboración de perfiles de personas físicas mencionada en el artículo 3, punto 4, de la Directiva (UE) 2016/680 o para evaluar rasgos y características de la personalidad o comportamientos delictivos pasados de personas físicas o colectivos[48].

e) Sistemas de IA para elaborar perfiles de personas físicas durante la detección, la investigación o el enjuiciamiento de delitos.

El Considerando 59 excluye de estos supuestos los sistemas de IA que se utilicen en la lucha contra el blanqueo de capitales.

7. Migración, asilo y gestión del control fronterizo

De nuevo aquí el punto de partida es el mismo que en los supuestos 1 y 6, que el uso de los sistemas de IA esté permitido. Y al igual que en el cumplimiento del Derecho (punto 6), estos sistemas de IA de alto riesgo sólo pueden ser utilizados por las autoridades públicas competentes o en su nombre, bien sean estatales o de la UE.

Los casos de sistemas de IA de alto riesgo son los siguientes[49]:

a) Polígrafos y herramientas similares.

48 Así se han utilizado sistemas de valoración con este fin, como el Compass en Estados Unidos o el RISCANVI en Cataluña (HERNÁNDEZ PEÑA, *El marco jurídico…*, op. cit., pág. 146).

49 El objetivo de esta inclusión es el siguiente: "Los sistemas de IA empleados en la migración, el asilo y la gestión del control fronterizo afectan a personas que con frecuencia se encuentran en una situación especialmente vulnerable y que dependen del resultado de las actuaciones de las autoridades públicas competentes. Por este motivo, es sumamente importante que los sistemas de IA que se utilicen en estos contextos sean precisos, no discriminatorios y transparentes, a fin de garantizar que se respeten los derechos fundamentales de las personas afectadas y, en particular, su derecho a la libre circulación, a la no discriminación, a la intimidad personal y la protección de los datos personales, a la protección internacional y a una buena administración" (Considerando 60).

b) Sistemas de IA para evaluar un riesgo, por ejemplo, un riesgo para la seguridad, la salud o de migración irregular, que plantee una persona física que tenga la intención de entrar en el territorio de un Estado miembro o haya entrado en él.

c) Sistemas de IA para ayudar a las autoridades públicas competentes a examinar las solicitudes de asilo, visado o permiso de residencia y las reclamaciones conexas con el fin de determinar si las personas físicas solicitantes reúnen los requisitos necesarios para que se conceda su solicitud, con inclusión de la evaluación conexa de la fiabilidad de las pruebas.

d) Sistemas de IA para detectar, reconocer o identificar a personas físicas, con excepción de la verificación de documentos de viaje.

8. Administración de justicia y procesos democráticos

Por un lado, se introducen como sistemas de IA de alto riesgo aquellos relacionados con la administración de justicia y, por tanto, utilizados por una autoridad judicial directa o indirectamente. Son los sistemas que vayan a ser utilizados para ayudar a una autoridad judicial en la investigación e interpretación de hechos y de la ley, así como en la garantía del cumplimiento del Derecho a un conjunto concreto de hechos, o a ser utilizados de forma similar en una resolución alternativa de litigios. Como se indica, son sistemas que ayudan al juez, pero que no le sustituyen, dado que la reserva de humanidad debe ser aquí prevalente. Pero, incluso, en estos supuestos de colaboración, que no de sustitución, también deben cumplirse los requisitos y obligaciones de los sistemas de IA de alto riesgo.

El segundo supuesto hace referencia a los sistemas democráticos, en orden a evitar la influencia en los votantes en los referendos o

elecciones[50]. Por eso se excluyen de esta consideración aquellos sistemas que no afectan directamente a las personas físicas.

IV.1.C.2.- Excepciones a la aplicación del Anexo III: la evaluación del proveedor

Como acaba de exponerse, los sistemas de IA relacionados en el Anexo III son de alto riesgo. Sin embargo, el art. 6.3 permite que determinados sistemas incluidos en dicho Anexo III no sean considerados de alto riesgo.

La razón de esta exclusión se encuentra en el nivel de riesgo, que se define como no "importante"[51]: "No obstante lo dispuesto en el apartado 2, un sistema de IA a que se refiere el anexo III no se considerará de alto riesgo cuando no plantee un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, también al no influir sustancialmente en el resultado de la toma de decisiones" (art. 6.3).

Así pues, la falta de magnitud o gravedad del riesgo y, sobre todo, el respeto a la autonomía de las personas puede provocar que el sistema de IA, a pesar de estar incluido en la relación del Anexo III, pueda no ser considerado de alto riesgo.

Sin embargo, algunos sistemas siempre van a ser de alto riesgo, como por ejemplo aquellos que lleven a cabo la elaboración de perfiles de personas físicas. Se entiende por elaboración de perfiles: "toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento

50 A tal respecto cabe hacer referencia a las experiencias de influencia en procesos electorales de la conocida empresa *Cambridge Analytica*.

51 La noción de riesgo importante aparece también mencionada en el apartado 6, con relación a la posible adición o modificación de las condiciones de excepción a la consideración de un sistema del Anexo III como no de alto riesgo.

profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física" (art. 4.4 RGPD).

Para la aplicación de la excepción antedicha de no consideración como de alto riesgo, se requieren dos requisitos:

1) Que se dé una o varias de las cuatro condiciones[52] señaladas en el citado apartado 3. Por tanto, bastaría que se diera sólo una de ellas.
2) Que el proveedor documente una evaluación previa de la que se concluya que, a pesar de estar incluido el sistema en el Anexo III, no es de alto riesgo por concurrir alguna de las condiciones señaladas.

En definitiva, es el proveedor quien, previa evaluación, declara que su sistema de IA no es de alto riesgo, por darse alguna de las condiciones enumeradas en este artículo.

La problemática derivada del debate de esta disposición se puede ver plasmada en el apartado 6, donde se permite a la Comisión la revisión de estas condiciones en función de la evolución tecnológica y del mercado, aunque sin que pueda reducir el nivel global de protección de las personas. La Comisión puede adoptar tres conductas:

1) Modificar las condiciones.
2) Añadir nuevas condiciones.
3) Suprimir las condiciones existentes. Pero sólo si la supresión es necesaria para mantener el nivel de protección de las personas.

52 El Considerando 53 explica la relación de cada una de esas cuatro condiciones con el concepto general de no influir sustancialmente en la toma de decisiones o no perjudicar dichos intereses sustancialmente.

La excepcionalidad de esta medida de exclusión de sistemas de IA del Anexo III va acompañada de una cautela regulada en el art. 80, referida al supuesto en el que el proveedor califique erróneamente un sistema de IA como excluido. En estos supuestos si la autoridad de vigilancia del mercado constata que el sistema de IA es de alto riesgo exigirá el cumplimiento de los requisitos y obligaciones de la LIA.

IV.1.D.- Modificación del Anexo III

El apartado 1 del art. 7 LIA habilita a la Comisión para modificar los supuestos contemplados en el Anexo III o para añadir otros nuevos. La redacción de este apartado se ajusta más bien a la previsión de la propuesta de la Comisión consistente en la adición, puesto que la posibilidad de modificación se introdujo en el procedimiento legislativo.

Desde dicha perspectiva, la posible adición de nuevos casos de uso de sistemas de alto riesgo del Anexo III requiere el cumplimiento de las dos condiciones expresadas en el citado apartado 1. Merece la pena detenerse en la segunda condición que es la equivalencia de riesgo, que implica la adición de nuevos casos por el hecho de que su riesgo es equivalente o mayor al de los sistemas ya contemplados en el Anexo III. A tal efecto, el apartado 2 del art. 7 refiere nada menos que once criterios, que permitan a la Comisión evaluar esta segunda condición a los efectos de la inclusión de un sistema de IA en el Anexo III, tal como reza el título de este artículo 7.

Por último, el apartado 3 del art. 7 permite a la Comisión suprimir sistemas de IA enumerados en el Anexo III cuando se den las dos condiciones expresadas en dicho precepto y que hacen ver que se ha reducido el nivel de riesgo (el riesgo ya no es considerable) y que no se ha bajado el nivel general de protección de las personas.

IV.2.- REQUISITOS DE LOS SISTEMAS DE IA DE ALTO RIESGO

Una vez definidos los sistemas de alto riesgo, la LIA pasa a establecer cuáles son los requisitos obligatorios para su introducción en el mercado y comercialización.

Los sistemas de alto riesgo del apartado 1 en relación con el Anexo I y del apartado 2 del art. 6 en relación con el Anexo III, no convierten a dichos sistemas en sistemas que pueden ser puestos en uso directamente. Lo expresa con rotundidad el Considerando 63: "El hecho de que un sistema de IA sea clasificado como un sistema de IA de alto riesgo en virtud del presente Reglamento no debe interpretarse como indicador de que su uso sea lícito con arreglo a otros actos del Derecho de la Unión o del Derecho nacional compatible con el Derecho de la Unión, por ejemplo, en materia de protección de los datos personales o la utilización de polígrafos y herramientas similares u otros sistemas para detectar el estado emocional de las personas físicas. Todo uso de ese tipo debe seguir realizándose exclusivamente en consonancia con los requisitos oportunos derivados de la Carta y de los actos aplicables del Derecho derivado de la Unión y del Derecho nacional. No debe entenderse que el presente Reglamento constituye un fundamento jurídico para el tratamiento de datos personales, incluidas las categorías especiales de datos personales, en su caso, salvo que el presente Reglamento disponga específicamente otra cosa".

Es, por ello, que la Sección 2 del Capítulo III establece los requisitos que debe cumplir un sistema de alto riesgo para poder ser introducido en el mercado y utilizado. Así pues, el operador que debe cumplir con estos requisitos es el proveedor, tal como impone el art. 16. a). Para el proveedor ésta es su primera obligación: velar porque sus sistemas de IA de alto riesgo cumplan los requisitos de la Sección 2. Además, tanto el importador como el distribuidor están obligados a verificar que el proveedor ha cumplido con estos requisitos (arts. 23 y 24) e incluso el proveedor inicial puede ser sustituido por nuevos proveedores (art. 25).

La LIA establece varios requisitos que debe cumplir un sistema de alto riesgo para ser introducido en el mercado y su posterior comercialización.

Estos requisitos persiguen mitigar los riesgos y garantizar la fiabilidad de los sistemas de IA. A tal efecto cabe recordar que el Considerando 27 efectúa un resumen de las Directrices éticas para una IA fiable, aprobadas en 2019 por el Grupo Independiente de Expertos de Alto Nivel sobre IA. Los siete requisitos para una IA fiable[53] pasan ahora a convertirse en requisitos obligatorios para los sistemas de IA de alto riesgo.

El art. 8.2 permite un cumplimiento de requisitos efectuado de forma integrada en cuanto a procedimientos y documentos para los sistemas de IA de productos de la Sección A del Anexo I. Como señala el Considerando 64, la LIA completa lo dispuesto en los actos legislativos referidos en dicha Sección A, lo que comportará una aplicación simultánea y complementaria de diversos actos legislativos, cuando menos, el propio del producto y la LIA. En estos casos el proveedor de un producto que incorpore un sistema de IA de alto riesgo puede efectuar la evaluación del producto y simultáneamente la del sistema de IA incorporado al mismo. De forma específica se recoge la integración del sistema de riesgos en el art. 9.10.

Los requisitos de los sistemas de IA de alto riesgo son los siguientes:

1) Implantación de un sistema de gestión de riesgos.
2) Realización de prácticas de gobernanza de datos para el caso de entrenamiento de modelos de IA.
3) Elaboración de la documentación técnica del sistema de IA.

53 Las Directrices para una IA fiable de 2019 señalaban los siguientes siete requisitos clave: 1) acción y supervisión humana; 2) solidez técnica y seguridad; 3) gestión de la privacidad y de los datos; 4) transparencia; 5) diversidad, no discriminación y equidad; 6) bienestar social y ambiental; y 7) rendición de cuentas.

4) Trazabilidad asegurada mediante un registro automático de eventos.
5) Transparencia y comunicación de información a los responsables del despliegue.
6) Vigilancia humana.
7) Nivel adecuado de precisión, solidez y ciberseguridad.

IV.2. A.- Implantación de un sistema de gestión de riesgos

El art. 9.1 LIA recoge este requisito: "Se establecerá, implantará, documentará y mantendrá un sistema de gestión de riesgos en relación con los sistemas de IA de alto riesgo". Este precepto[54] utiliza diversos verbos que suponen cada uno de ellos exigencias para el proveedor, y que van ligados a otros requisitos y obligaciones que la LIA les impone, así como al ciclo de vida.

Además, el art. 17.1 LIA obliga al proveedor a incluir dentro del sistema de gestión de la calidad el sistema de gestión de riesgos (letra g)).

La idea principal es que este requisito constituye no sólo una exigencia previa, sino también una obligación continuada, toda vez que el sistema de gestión de riesgos no termina con su establecimiento e implantación, sino que debe ser mantenido a lo largo de la vida del sistema de IA[55].

En todo caso, el apartado 3 del art. 9 acota los riesgos a que se refiere este precepto: "Los riesgos a que se refiere el presente artículo son únicamente aquellos que pueden mitigarse o eliminarse razonablemente mediante el desarrollo o el diseño del

54 Para un examen completo y amplio del precepto me remito a SCHUETT, J., "Risk management in the Artificial Intelligence Act", en European Journal of Risk Regulation, 2023, págs. 1-19.

55 HERNÁNDEZ PEÑA lo califica de sistema de *compliance* o cumplimiento normativo (*El marco jurídico…*, op. cit., págs. 149-150).

sistema de IA de alto riesgo o el suministro de información técnica adecuada".

Por ello, es relevante la vinculación entre sistema de gestión de riesgos y ciclo de vida del sistema de IA[56]. Mientras un sistema de IA esté en uso debe mantenerse el sistema de gestión de riesgos.

Este requisito está plenamente justificado dado que, como se ha dicho reiteradamente, la regulación de la LIA está basada en el riesgo y por tanto se trata de controlar y minimizar los riesgos desde el inicio hasta el final de la vida del sistema de IA. Es, por ello, que el apartado 2 de este artículo 9 lo califique como "proceso iterativo continuo", sometido a revisiones y actualizaciones, e incluso relate las etapas de gestión de riesgos.

Estas etapas están divididas en función del tipo de riesgo: riesgo evaluable y previsible, y riesgo que podría surgir bien por una estimación de estas eventualidades no conocidas pero que podrían surgir por un uso adecuado o inadecuado del sistema (uso indebido razonablemente previsible) o bien derivados de la ejecución del sistema de vigilancia poscomercialización.

Procede detenerse en el primer caso, referido a los riesgos conocidos y previsibles derivados de una utilización del sistema de IA conforme a su finalidad. Se exige, primero, que se determinen y analicen estos riesgos, con especial atención a si afectan a menores o personas vulnerables (apartado 9). Además, se impone que se diseñen y se adopten medidas adecuadas para la gestión de estos riesgos. Los apartados 4 y 5 se refieren a estas medidas en orden a que minimicen los riesgos o evalúen aquellos que pueden ser considerados admisibles por su carácter residual.

56 PALMA ORTIGOSA, A., diferencia dos fases sustanciales en el ciclo de vida: la fase diseño y la fase de despliegue y toma de decisiones, que analiza posteriormente ("El ciclo de vida de los sistemas de inteligencia artificial. Aproximación técnica de las fases presentes durante el diseño y despliegue de los sistemas algorítmicos", en *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, págs. 37-48).

Para la determinación de estas medidas, los sistemas de IA de alto riesgo deben ser sometidos a pruebas, bien mediante banco de pruebas previo o incluso en condiciones reales, que deben ser realizadas en todo caso de forma previa a su introducción en el mercado y también, si procede, en cualquier momento del proceso de desarrollo (apartados 6-8).

IV.2.B.- Gobernanza de datos

El art. 10 impone el requisito de la gobernanza de datos[57]. Este precepto obliga a diferenciar entre:

1) Sistemas de IA: sólo se les aplica lo dispuesto en los apartados 2 a 5 para los conjuntos de datos de prueba.
2) Modelos de IA: suponen la utilización de conjuntos de datos para entrenamiento del modelo, y son el objeto de regulación de este precepto.
3) Modelos de IA de uso (o propósito) general, regulados de forma singular en el Capítulo V.

Así pues, este artículo se refiere, principalmente, a los modelos de IA, puesto que se aplica en el caso de sistemas de IA solo a los conjuntos de datos de prueba (apartado 6).

El Considerando 67 explica el objetivo de este precepto: "Los datos de alta calidad y el acceso a datos de alta calidad desempeñan un papel esencial a la hora de proporcionar una estructura y garantizar el funcionamiento de muchos sistemas de IA, en especial cuando se emplean técnicas que implican el entrenamiento de modelos, con vistas a garantizar que el sistema de IA de alto riesgo funcione del modo previsto y en condiciones de seguridad y no se convierta en una fuente de algún tipo de discriminación prohibida por el Derecho de la Unión".

57 HERNÁNDEZ PEÑA enfatiza en la importancia de la gobernanza de datos para los sistemas de IA de alto riesgo, dado que "parte importante de los riesgos para los derechos fundamentales de los sistemas de IA derivan de los datos utilizados para entrenar o validar los modelos" (*El marco jurídico…*, op. cit., págs. 150-151).

Se somete a los conjuntos de datos a prácticas de gobernanza y gestión de datos a fin de asegurar la calidad de datos y evitar o eliminar la introducción de sesgos.

En todo caso, la gobernanza de datos debe garantizar la aplicación de los principios de la normativa de protección de datos personales respecto de estos datos: minimización, anonimización o cifrado[58], puesto que, como señala el Considerando 69, "El derecho a la intimidad y a la protección de datos personales debe garantizarse a lo largo de todo el ciclo de vida del sistema de IA".

Téngase también en cuenta que se impone al responsable del despliegue la obligación de asegurarse de que los datos de entrada sean pertinentes y suficientemente representativos (art. 26.2).

Aparecen dos cuestiones relevantes en relación con los datos. La primera es el control para la evitación de la aparición de sesgos en los sistemas de IA (letras f) y g) del art. 10.2), que incluso permite el tratamiento excepcional de datos personales de categoría especial en orden a detectar y corregir sesgos (apartado 5)[59]. La aparición de sesgos se ha mostrado en diversos casos de uso de inteligencia artificial tanto en el sector público como en el sector

58 Así lo afirma el Considerando 69: "El derecho a la intimidad y a la protección de datos personales debe garantizarse a lo largo de todo el ciclo de vida del sistema de IA. A este respecto, los principios de minimización de datos y de protección de datos desde el diseño y por defecto, establecidos en el Derecho de la Unión en materia de protección de datos, son aplicables cuando se tratan datos personales. Las medidas adoptadas por los proveedores para garantizar el cumplimiento de estos principios podrán incluir no solo la anonimización y el cifrado, sino también el uso de una tecnología que permita llevar los algoritmos a los datos y el entrenamiento de los sistemas de IA sin que sea necesaria la transmisión entre las partes ni la copia de los datos brutos o estructurados, sin perjuicio de los requisitos en materia de gobernanza de datos establecidos en el presente Reglamento".

59 El Considerando 67 explica esta preocupación por los sesgos: "Los sesgos, por ejemplo, pueden ser inherentes a los conjuntos de datos subyacentes, especialmente cuando se utilizan datos históricos, o generados cuando los sistemas se despliegan en entornos del mundo real. Los resultados de los sistemas de IA dependen de dichos sesgos inherentes, que tienden a aumentar gradualmente y, por tanto, perpetúan y amplifican la discriminación existente, en particular con respecto a las personas vulnerables pertenecientes a determinados colectivos, en particular colectivos raciales o étnicos".

privado, con resultados inaceptables para determinadas personas o categorías de personas[60]. Se trata de salvaguardar los derechos fundamentales, principalmente, aquí la igualdad frente a la discriminación directa e indirecta por medio de algoritmos[61]. Para ello jugará un papel fundamental la evaluación de impacto relativa a los derechos fundamentales exigida por el art. 27[62].

La segunda cuestión es la de la representatividad de los datos. El art. 10 LIA exige que los datos sean suficientemente representativos (apartado 3) y que carezcan de errores, y que tengan en cuenta el entorno geográfico, contextual, conductual o funcional (apartado 4).

IV.2.C.- Documentación técnica

La documentación técnica del sistema de IA debe ser elaborada antes de su introducción al mercado, y posteriormente actualizada (art. 11). Su finalidad es doble: facilitar a las autoridades

60 BELLOSO MARTÍN indica que "los sesgos no ocurren de manera espontánea. Los algoritmos de IA basados en datos, no producen sesgos pero sí pueden reproducirlos sin la adecuada intervención humana y ello en tres de las fases principales: en la recolección de los datos, porque tales datos recopilados reflejen prejuicios ya existentes; en la preparación de datos de entrenamiento (a la hora de seleccionar y procesar los atributos que le proporcionamos al algoritmo); y en la toma de decisiones (Las propuestas y decisiones que se adoptan a lo largo de todo el ciclo de vida del desarrollo inteligente)" (pág. 48). Y más adelante pone ejemplos de casos de discriminación algorítmica por el uso de sesgos ("La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos?", en *Inteligencia artificial y filosofía del Derecho*, Laborum. Murcia, 2022, págs. págs. 49-51), con especial atención a la discriminación por género (págs. 55-61). Por su parte, SORIANO ARNANZ, A., se refiere a la introducción de sesgos en la base de datos, en la clasificación, los aprendidos tras la puesta en marcha, y los derivados de las decisiones humanas que intervienen en el uso y funcionamiento de los sistemas algorítmicos ("La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicación de inteligencia artificial", en *Nuevas normatividades: inteligencia artificial, derecho y género,* Aranzadi, Cizur Menor, 2021, págs. 65-68).

61 SORIANO ARNANZ, A., llama la atención sobre la necesidad de atender a ambos tipos de discriminación, tanto la directa como la indirecta ("La aplicación del marco jurídico...", op. cit., págs. 73-78.

62 Así lo sostiene SIMÓN CASTELLANO, "Las evaluaciones de impacto algorítmico en los derechos fundamentales: hacia una efectiva minimización de sesgos", en *Algoritmos abiertos y que no discriminen en el sector público*, Tirant lo blanch, Valencia, 2023, pág. 45-46.

competentes y organismos notificados el ejercicio de sus competencias, y permitir la comprobación del cumplimiento de los requisitos exigidos para los sistemas de IA de alto riesgo[63].

Debe indicar que se cumplen los requisitos de la Sección 2, y ser clara y completa. Su contenido mínimo viene fijado en el Anexo IV, que puede ser modificado por la Comisión (apartado 3).

Se prevé una documentación simplificada para las pymes. Y, asimismo, cuando el sistema de IA esté vinculado a un producto en el caso del Anexo I, la documentación será única, para el producto y para el sistema de IA.

El art. 77 LIA se refiere a un supuesto especial de solicitud de esta documentación por parte de las Autoridades competentes en materia de protección de datos, en orden a determinar si el sistema de IA cumple con la protección de los derechos fundamentales. Incluso de entender que no es suficiente dicha documentación podrá solicitar de la autoridad de vigilancia del mercado que el sistema de IA de alto riesgo se someta a pruebas.

IV.2.D.- Trazabilidad y registro

La idea fundamental de ciclo de vida o de proceso iterativo continuo comporta la exigencia de trazabilidad de los sistemas de IA (art. 12). En todo momento debe ser posible detectar y constatar los eventos en la ejecución del sistema de IA[64]. Por ello, el art. 12 impone el requisito de que los sistemas de IA de alto riesgo permitan técnicamente el registro de eventos en todo momento.

63 Así lo indica HERNÁNDEZ PEÑA (*El marco jurídico…*, op. cit., pág. 153).

64 SIMÓN CASTELLANO señala que la trazabilidad se refiere a que "cualquier acción que lleve a cabo un usuario del sistema…quedará registrado y dejará un rastro que, en el futuro y en caso de ser necesario, podrá ser examinado", por lo que estas acciones deben quedar registradas ("Allende una teoría general…", op. cit., págs. 134-135).

Para garantizar una adecuada trazabilidad, el registro tiene que tener unas capacidades de registro de eventos que permitan la detección de riesgos, la vigilancia poscomercial, y la vigilancia de su funcionamiento por parte de los responsables del despliegue. Incluso se imponen unos registros de eventos específicos para el caso de los sistemas biométricos (apartado 3).

IV.2.E.- Transparencia e información

El art. 13 LIA impone la exigencia de transparencia e información a los proveedores respecto de los responsables del despliegue[65]. Hay que advertir que no se trata de una transparencia general[66], sino que, por el contrario, se limita a determinar las relaciones entre el proveedor, en cuanto diseñador del sistema de IA, y los responsables del despliegue, en cuanto personas que utilizan el sistema[67].

A tal efecto los sistemas de IA de alto riesgo deben ir acompañados de instrucciones, que contendrán una información concisa, completa, correcta, clara, pertinente, accesible y comprensible. Incluso se impone un contenido mínimo de dichas instrucciones (apartado 3).

65 SIMÓN CASTELLANO eleva el marco de la transparencia dándole un alcance más general, diferenciando cinco subcategorías: simulabilidad, descomponibilidad, legibilidad, auditabilidad y publicidad activa. Y considera como categoría diferente, aunque próxima, a la explicabilidad que se divide a su vez en tres subcategorías: inteligilidad, comprensibilidad e interpretabilidad ("Allende una teoría general...", op. cit., págs. 127-133).

66 COTINO HUESO califica la transparencia de la propuesta de LIA como "transparencia interna", dado que no se refiere a las personas afectadas ("Transparencia y explicabilidad de la inteligencia artificial. Elementos conceptuales, generales y de género", en *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo blanch, Valencia, 2022, pág. 47).

67 El Considerando 72 señala en su primer párrafo lo siguiente: "A fin de abordar las preocupaciones relacionadas con la opacidad y complejidad de determinados sistemas de IA y ayudar a los responsables del despliegue a cumplir sus obligaciones en virtud del presente Reglamento, debe exigirse transparencia respecto de los sistemas de IA de alto riesgo antes de su introducción en el mercado o su puesta en servicio".

El Considerando 72 detalla el objeto de estas instrucciones: ayudar a utilizar el sistema y tomar decisiones con conocimiento de causa, y también elegir correctamente el sistema que se va a utilizar. Para ello se refiere a que las instrucciones contengan ejemplos prácticos y se redacten en la lengua que decida el Estado miembro. Por su parte, el art. 15.3 LIA exige que indiquen los niveles de precisión y los parámetros de evaluación de los sistemas de IA.

Por tanto, parece que la LIA engloba dentro de este requisito tanto la transparencia como la explicabilidad de los sistemas de IA[68], aunque limitada a las relaciones entre proveedor y responsable del despliegue. Para obligaciones más generales es preciso acudir a lo dispuesto en el art. 50 que es aplicable, también, a los sistemas de alto riesgo.

IV.2.F.- Supervisión humana

El art. 14 LIA impone la vigilancia humana[69] que supervise el funcionamiento del sistema de IA de alto riesgo, plasmando en el plano normativo las previsiones de la Declaración Europea sobre los Derechos y los Principios Digitales para la Década Digital[70]. Dicha supervisión debe estar prevista en el diseño del

68 Diversos autores, entre los que cabe destacar a COTINO HUESO, han insistido en las diferencias conceptuales y prácticas entre transparencia y explicabilidad (COTINO HUESO, L. y CASTELLANOS CLARAMUNT, J., *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo blanch, Valencia, 2022).

69 PONCE SOLÉ, J., considera la supervisión humana como una técnica menos drástica que la reserva de humanidad, puesto que permite el uso de un sistema de IA, aunque sometido a supervisión o vigilancia humana ("Reserva de humanidad y supervisión humana de la Inteligencia artificial", en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, 2022, pág. 64).

70 Esta Declaración en su Capítulo III, apartado 6, sobre Libertad de elección, se refiere a las interacciones con algoritmos y sistemas de inteligencia artificial, y dado que, la inteligencia artificial debe ser un instrumento al servicio de las personas, establece el compromiso de "velar por que los sistemas algorítmicos se basen en conjuntos de datos adecuados para evitar la discriminación y permitir la supervisión humana de todos los resultados que afecten a la seguridad y los derechos fundamentales de las personas" (letra c)). También en el Capítulo

sistema de IA y por tanto éste debe contar con un interfaz humano-máquina[71].

La supervisión se encomienda a una persona física. Como señala el art. 26.2 el responsable del despliegue debe designar para esta tarea personas físicas que tengan la competencia, formación y autoridad necesarias.

Es aquí donde cobran relevancia las instrucciones que el proveedor debe facilitar al responsable del despliegue, y que éste deberá transmitir a la persona supervisora. La vigilancia exige que la persona supervisora entienda el sistema de IA y su funcionamiento, hasta el extremo de que pueda decidir intervenir en dicho funcionamiento e, incluso, interrumpirlo (letra e) del apartado 4).

Las medidas de vigilancia las define el proveedor antes de la introducción del sistema de IA en el mercado o de su puesta en funcionamiento, y pueden venir integradas en el propio sistema y actuar desde el momento inicial o a lo largo de su funcionamiento (apartado 3).

El art. 14.5 recoge un supuesto de supervisión humana reforzada para determinados sistemas de identificación biométrica (Considerando 73), en los que la supervisión debe efectuarse por dos personas que lo verifiquen y confirmen por separado. No obstante, este requisito reforzado no se aplica en determinados casos (garantía del cumplimiento del Derecho, migración, control fronterizo o asilo) si el Derecho nacional o de la UE lo consideran desproporcionado (segundo párrafo del apartado 5).

II sobre Solidaridad e inclusión en el epígrafe 6 relativo a Condiciones de trabajo justas y equitativas, fija el compromiso de: "garantizar, en particular, que las decisiones importantes que afecten a los trabajadores cuenten con supervisión humana y que, en general, se los informe de que están interactuando con sistemas de inteligencia artificial" (letra e)).

71 La interfaz humano-máquina (HMI) debe estar incorporada al sistema y hay diversos modelos que permiten su control y supervisión por las personas físicas.

IV.2.G.- Precisión, solidez y ciberseguridad

El diseño y desarrollo de los sistemas de IA de alto riesgo debe garantizar su precisión, solidez y ciberseguridad a lo largo de todo su ciclo de vida (art. 15.1). Y en las instrucciones de uso deben indicarse los niveles de precisión y los parámetros para su evaluación (apartado 3).

La LIA diferencia entre solidez y resistencia. La solidez es *ad intra* del sistema, de modo que sea resistente a errores, fallos e incoherencias. El Considerando 75 pone como ejemplos de la solidez: "Estas soluciones técnicas pueden incluir, por ejemplo, mecanismos que permitan al sistema interrumpir de forma segura su funcionamiento (planes de prevención contra fallos) en presencia de determinadas anomalías o cuando el funcionamiento tenga lugar fuera de determinados límites predeterminados".

La resistencia es *ad extra* del sistema, de forma que se impida el acceso de terceros no autorizados, o terceros maliciosos (Considerando 76), por lo que se deben tener medidas de ciberseguridad. El Considerando 77 se remite en este punto de forma adicional a lo que disponga el futuro Reglamento de ciberseguridad, que deberá ser tenido en cuenta en el procedimiento de evaluación de conformidad previsto en la LIA (Considerando 78). En este momento se halla vigente la Directiva (UE) 2022/2555, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en la UE.

La UE ha creado la Agencia de la Unión Europea para la Ciberseguridad (ENISA), que se encuentra regulada en el Reglamento (UE) 2019/881, con el fin de alcanzar un nivel elevado de ciberseguridad, ciberresiliencia y confianza dentro de la Unión, en orden a garantizar el correcto funcionamiento del mercado interior.

V.- SISTEMAS DE IA DE RIESGO LIMITADO

El art. 50 impone obligaciones de transparencia para los sistemas de IA de alto riesgo y, también, para determinados sistemas de IA, aunque no sean de alto riesgo. De ello, se ha derivado la configuración de una nueva categoría de sistemas de IA: la de riesgo limitado.

El Considerando 132 lo explica así: "Determinados sistemas de IA destinados a interactuar con personas físicas o a generar contenidos pueden conllevar riesgos específicos de suplantación o engaño, con independencia de si cumplen las condiciones para ser considerados como de alto riesgo o no".

Esto significa que aquellos sistemas de bajo riesgo que incidan en alguna de las situaciones descritas en el art. 50 están sometidos a las obligaciones de información y transparencia, pero únicamente a éstas. Y, por otra parte, el art. 50 constituye una obligación adicional para los proveedores y responsables del despliegue de sistemas de IA de alto riesgo, añadida a las enumeradas en los arts. 16 y 24 LIA, como advierte el apartado 6 del art. 50: "Los apartados 1 a 4 no afectarán a los requisitos y obligaciones establecidos en el capítulo III...".

El art. 50 prevé cuatro supuestos:

1) Sistemas de IA destinados a interactuar directamente con personas físicas.
2) Sistemas de IA, entre los que se incluyen los sistemas de IA de uso general, que generen contenido sintético de audio, imagen, vídeo o texto,
3) Sistemas de reconocimiento de emociones o de un sistema de categorización biométrica. Procede advertir que estos sistemas son de alto riesgo (Anexo III. punto 1).
4) Sistemas de IA que generen o manipulen imágenes o contenidos de audio o vídeo que constituyan una ultrafalsificación o que generen o manipulen texto que se publique

con el fin de informar al público sobre asuntos de interés público.

En todos estos casos se impone la obligación, a los proveedores en los dos primeros supuestos y a los responsables del despliegue en los supuestos tercero y cuarto, de informar a las personas sobre dichas circunstancias. También se contemplan excepciones a las obligaciones de transparencia, especialmente, cuando se trata de detección, prevención, investigación y enjuiciamiento de delitos.

Además de estas obligaciones de información, que debe facilitarse de forma clara y distinguible, se prevé la elaboración de códigos de buenas prácticas e, incluso, la adopción por la Comisión de actos de ejecución para especificar normas comunes para el cumplimiento de estas obligaciones (apartado 7 del art. 50).

VI.- SISTEMAS DE IA DE BAJO O NULO RIESGO

La LIA no regula los sistemas de IA de bajo o nulo riesgo, pero se refiere de forma directa a ellos en sus Considerandos 165 y 166. En el Considerando 165 reitera el sometimiento de los sistemas que no sean de alto riesgo a profundizar en la adopción de una inteligencia artificial ética y fiable en la Unión. Para ello, se opta por un sistema de autoregulación mediante códigos de conducta que vayan incorporando los requisitos y obligaciones contempladas en la LIA también para estos sistemas que no son de alto riesgo. Se trata de que se autoimpongan requisitos adicionales, de los que se ofrecen algunos ejemplos: "los elementos de las Directrices éticas de la Unión para una IA fiable[72], la sostenibilidad medioambiental, medidas de alfabetización en

72 El considerando 27 de la LIA señala que las Directrices éticas para una IA fiable son aplicables a todos los sistemas de IA.

materia de inteligencia artificial, la inclusividad y la diversidad en el diseño y el desarrollo de los sistemas de IA, lo que incluye tener en cuenta a las personas vulnerables y la accesibilidad de las personas con discapacidad, la participación de las partes interesadas".

Por su parte el Considerando 166 se refiere a los sistemas de IA asociados a productos, que quedan fuera de la LIA por no ser sistemas de alto riesgo, para los que se exige que sean seguros, remitiéndose a la aplicación, como red de seguridad, del Reglamento (UE) 2023/988[73].

La implicación más relevante de la LIA respecto de los sistemas de IA que no son de alto riesgo se encuentra establecida en su art. 95, referido a los códigos de conducta. Se anima a los proveedores, y también a los responsables del despliegue, a contar con códigos de conducta, que les conduzcan, de forma voluntaria, hacia el cumplimiento de los requisitos y obligaciones establecidos en la LIA.

Por último, conviene no confundir los sistemas de bajo o nulo riesgo con la declaración de un proveedor de que un sistema de IA incluido en el Anexo III como sistema de alto riesgo no es de alto riesgo. Estos supuestos encuentran diversas previsiones en la LIA. En primer lugar, se establece la obligación de registrarlos en la base de datos de la UE, prevista en el art. 71 (art. 49.2). En segundo lugar, se dispone que, cuando un sistema de IA que haya sido considerado que no es de alto riesgo se convierta en un sistema de IA de alto riesgo por cualquier motivo (desarrollo, uso, funcionamiento, modificaciones, etc.) deberá procederse a calificarlo como sistema de IA de alto riesgo y realizar la evaluación de conformidad (art. 80).

73 Este Reglamento regula la seguridad de los productos, para velar por la protección de los consumidores. Es una norma de aplicación supletoria, para todos aquellos productos que no tengan requisitos específicos de seguridad impuestos por el Derecho de la UE.

VII.- CONCLUSIONES

Del examen que acaba de efectuarse se desprenden algunas conclusiones generales sobre la regulación de la LIA respecto de los sistemas de IA.

En primer término, la LIA efectúa una clasificación de los sistemas de IA en cuatro categorías en función de los riesgos. No obstante, se trata de una categorización llena de excepciones.

No sólo el propio concepto de riesgo (inadmisible, elevado o mínimo) constituye una cuestión con contornos a menudo poco precisos, sino también la LIA encierra demasiada imprecisión y una gran vaguedad.

La LIA, a pesar de su ánimo de ser una ley unificadora de un régimen general aplicable a los sistemas de IA, es una norma que recibirá una aplicación desde diversas perspectivas, principalmente la innovación y la protección de los derechos fundamentales, y por organismos con competencias y funciones diversas tanto a nivel europeo (Comité Europeo de Inteligencia Artificial, Oficina de la IA, Comité Europeo de Protección de Datos) como a nivel estatal (autoridad de vigilancia del mercado, autoridad notificantes, autoridades de protección de datos personales).

Así pues, la LIA precisa de una mayor concreción, sobre todo, para una aplicación práctica segura y acorde a sus disposiciones, que se efectuará tanto por actos normativos o de ejecución como por guías, orientaciones o instrucciones (*soft law*).

VIII.- BIBLIOGRAFÍA

BECK, U., *La sociedad del riesgo. Hacia una nueva modernidad,* Paidós, Barcelona, 2010.

BELLOSO MARTÍN, N., "La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos?", en *Inteligencia artificial y filosofía del Derecho*, Laborum. Murcia, 2022, págs. 45-78.

CHRISTAKIS, T. y KARRATHANASIS, T., "Tools for Navigating the EU AI Act (2) Visualisation Pyramid", AI Regulation Papers 24-03-5, AI-Regulation.com, March 8th, 2024.

COTINO HUESO, L., "Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal", en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, 2022, págs. 68-79.

COTINO HUESO, L., "Transparencia y explicabilidad de la inteligencia artificial. Elementos conceptuales, generales y de género", en *Transparencia y explicabilidad de la inteligencia artificial,* Tirant lo blanch, Valencia, 2022, págs. 25-70.

COTINO HUESO, L. y CASTELLANOS CLARAMUNT, J., *Transparencia y explicabilidad de la inteligencia artificial*, Tirant lo blanch, Valencia, 2022.

COTINO HUESO, L. et al., "Un análisis crítico constructivo de la propuesta de Reglamento de la Unión Europea por la que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act)", en *Diario La Ley, sección Ciberderecho*, 2 de julio de 2021.

COTINO HUESO, L., "Nuevo paradigma en la garantía de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivos de la inteligencia artificial", en *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Mayor, 2022, págs. 69-105.

COTINO HUESO, L., "Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos", en *Derecho Público de la Inteligencia Artificial*, Fundación Manuel Giménez Abad, Zaragoza, 2023, págs. 347-402.

DE LA QUADRA FERNÁNDEZ DEL CASTILLO, T. (2019). "Derechos fundamentales, democracia y mercado en la edad digital". *Derecho digital e Innovación* núm. 1.

DE MIGUEL ASENSIO, P.A., "Propuesta de Reglamento sobre inteligencia artificial", *La Ley Unión Europea*, núm. 92, mayo 2021.

ESTEVE PARDO, J., *Técnica, riesgo y Derecho. Tratamiento del riesgo tecnológico en el Derecho ambiental,* Ariel, Barcelona, 1999.

FERNÁNDEZ HERNÁNDEZ, F., "La futura regulación europea de la inteligencia artificial: objetivos, principios y pautas", en *Claves de inteligencia artificial y derecho*, La Ley, Madrid, 2022, págs. 115-179.

HERNÁNDEZ PEÑA, J.C., *El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza*, Aranzadi, Cizur Menor, 2022.

HERNÁNDEZ PEÑA, J.C., "Organización y gobernanza de la inteligencia artificial: marco general", en *Inteligencia artificial y sector público. Retos, límites y medios*, Tirant lo blanc, Valencia, 2023, pp. 599-630.

HUERGO LORA, A., "Gobernar con algoritmos, gobernar los algoritmos", *El Cronista del Estado Social y Democrático de Derecho,* núm. 100, 2022, págs. 80-89.

MORAL SORIANO, L. "Modelos de gobernanza global de la inteligencia artificial", en *Inteligencia artificial y Derecho. El jurista ante los retos de la era digital,* Aranzadi, Cizur Menor, 2021, págs. 235-258.

PALMA ORTIGOSA, A., "El ciclo de vida de los sistemas de inteligencia artificial. Aproximación técnica de las fases presentes durante el diseño y despliegue de los sistemas algorítmicos", en *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, págs. 29-51.

PONCE SOLÉ, J., "Reserva de humanidad y supervisión humana de la Inteligencia artificial". *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, 2022, págs. 58-67.

SALAZAR GARCÍA, I., "Retos actuales de la ética en la Inteligencia Artificial", en *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Aranzadi, Cizur Menor, 2022, págs. 53-66.

SCHUETT, J., "Risk management in the Artificial Intelligence Act", en European Journal of Risk Regulation, 2023, págs. 1-19.

SIMÓN CASTELLANO, P., "Allende una teoría general de las garantías jurídicas para una inteligencia artificial confiable", en *Derecho Público de la Inteligencia Artificial*, Fundación Manuel Giménez Abad, Zaragoza, 2023, págs. 111-148.

SIMÓN CASTELLANO, P., "Las evaluaciones de impacto algorítmico en los derechos fundamentales: hacia una efectiva minimización de sesgos", en *Algoritmos abiertos y que no discriminen en el sector público*, Tirant lo blanch, Valencia, 2023, págs. 27-56.

SORIANO ARNANZ, A., "La aplicación del marco jurídico europeo en materia de igualdad y no discriminación al uso de aplicación de inteligencia artificial", en *Nuevas normatividades: inteligencia artificial, derecho y género,* Aranzadi, Cizur Menor, 2021, págs. 63-87.

VIDA FERNÁNDEZ, J., "La gobernanza de los riesgos digitales: Desafíos y avances en la regulación de la Inteligencia artificial", *Cuadernos de Derecho Transnacional*, núm. 14-1.

RIESGOS DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL GENERATIVA (*)

RISKS OF GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEMS AND THE EUROPEAN ARTIFICIAL INTELLIGENCE REGULATION

Por Mª JESÚS JIMÉNEZ LINARES
Profesora Titular de Derecho civil de la Universidad de Granada

(*) Este trabajo se recibió el 11 de junio de 2024 y fue aceptado el 30 de julio.

REVISTA DE
PRIVACIDAD Y DERECHO DIGITAL

RESUMEN

La inteligencia artificial generativa permite la creación automática de contenido "original" en diferentes formatos (texto, audio, vídeo e imágenes). Un contenido que cada vez se acerca más al creado por los humanos. La inteligencia artificial generativa ofrece a la sociedad tantos beneficios como posibles riesgos. Se analizarán algunos como la confusión, la desinformación, los riesgos de los derechos de autor, a la seguridad, a la privacidad, la ciberdelincuencia, al derecho al honor, la intimidad y la propia imagen que pueden lesionarse con los deep fakes, la manipulación, los riesgos en el mercado laboral, los psicológicos, los sesgos...El nuevo Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) acoge la inteligencia artificial generativa dentro de la inteligencia de uso general, teniendo en cuenta la presencia, de posibles riesgos sistémicos y la exigibilidad de la ciberseguridad. Se analizará también la conexión de la inteligencia artificial generativa con la prohibida y la de alto riesgo. El Reglamento muestra especial preocupación ante las "ultrasuplantaciones", deep fakes (donde es difícil distinguir entre lo real y lo irreal), y el necesario conocimiento de su origen artificial cuando las personas físicas interactúen con ellas, con la correspondiente obligación de transparencia. Todo ello, sin olvidar, la necesidad de hacer lo ético jurídico y de la alfabetización.

PALABRAS CLAVE: *Inteligencia artificial generativa, riesgos, desinformación, seguridad, privacidad, ciberseguridad, transparencia, deep fakes, ultrasuplantaciones, riesgos sistémicos, ley de inteligencia artificial.*

ABSTRACT

Generative artificial intelligence allows the automatic creation of "original" content in different formats (text, audio, video and images). Content that is getting closer and closer to that created by humans. Generative artificial intelligence offers society as many benefits as possible risks. Some will be analysed such as confusion, disinformation, copyright risks, security, privacy, cybercrime, the right to honour, privacy and self-image that can be injured by deep fakes, manipulation, risks in the labour market, psychological risks, bias... The new Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) welcomes generative artificial intelligence within general purpose intelligence, considering the presence of possible systemic risks and the enforceability of cybersecurity. The connection of generative artificial intelligence with prohibited and high-risk artificial intelligence will also be analysed. The Regulation shows special concern for "ultra-phishing", deep fakes (where it is difficult to distinguish between the real and the unreal), and the necessary knowledge of their artificial origin when natural persons interact with them, with the corresponding obligation of transparency. All this, without forgetting the need for legal ethics and literacy.

KEY WORDS: *Generative artificial intelligence, risks, disinformation, security, privacy, cybersecurity, transparency, deep fakes, ultra spoofing, systemic risks, artificial intelligence law.*

SUMARIO

I.- INTRODUCCIÓN[1]

En 1984, en la película "los gremlins" aparecieron unas "criaturas" muy especiales. Gizmo, un mogwai, ser adorable, cuyo dueño, Mr.Wing, no pensaba, en principio, venderlo a Rand Peltzer (que lo quería para su hijo Billy) porque suponía una gran responsabilidad. Gizmo acabó generando unos seres totalmente destructivos, como Stripe, asesinos sin escrúpulos. Todo ello provocado por un comportamiento descuidado humano, por el incumplimiento de unas normas mínimas con Gizmo: a la criatura no le gustaba la luz brillante, la luz del sol lo mataría y nunca debían darle agua (ni bañarlo) ni darle de comer después de la medianoche[2].Finalmente, ante el caos que se generó en la sociedad, Mr. Wing volvió a recuperar a Gizmo.

Igual que los gremlins una simple IA que nos divierte generando vídeos, fotos, imágenes o clonando voces, puede llegar a convertirse en la causante de amenazas múltiples y poner en jaque la seguridad nacional, si no se ejerce el uso y el control humano adecuado de la misma y se establece su correcta regulación. La gran paradoja de la IA y en este caso concreto de la IA generativa es ser portadora a la vez de grandes beneficios y riesgos como se verá posteriormente.

1 Este trabajo forma parte del Proyecto de investigación "Nuevos avances en la legislación de transparencia en España: mejoras en la definición del marco regulatorio" (PID 2021-124724NB-100), del que es IP la profesora Ana de Marcos Fernández.

2 WIKIPEDIA, https://es.wikipedia.org/wiki/Gremlins (recuperado el 1 de mayo del 2024).

II.- LA IA GENERATIVA: CONCEPTO, BENEFICIOS Y RIESGOS

El 12 de julio del 2024 se publicó 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 llamada Ley de Inteligencia Artificial, a partir de ahora LIA)[3] centrado y enfocado en la gestión de riesgos. Se clasifica la IA en función de su riesgo potencial e impacto, imponiéndose más obligaciones y control cuanto mayor es el riesgo. Busca una IA fiable, ética, digna de confianza, cuyo eje sea el ser humano. Establece un marco jurídico de normas armonizadas que protegen fuertemente los intereses públicos, como la salud y la seguridad y los derechos fundamentales, particularmente, la democracia, el Estado de Derecho y el medio ambiente. En conexión con la LIA, es necesario fomentar un ecosistema de confianza preocupado por la rendición de cuentas (la responsabilidad), los datos, desarrollo de confianza y despegue, notificación de incidentes, pruebas y garantía y seguridad[4]. Se analizarán los riesgos de la IA generativa y su presencia en la LIA.

3 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) nº 300/2008, (UE) nº 167/2013, (UE) nº 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), DOUE, Serie L, 12 de julio del 2024. Una visión general sobre el Reglamento se puede ver en: FERNÁNDEZ HERNÁNDEZ, C., et al. "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", Diario LA LEY, Sección Ciberderecho, nº 85, 28 de junio de 2024 y BARRIO ANDRÉS, M., "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", Diario La Ley, nº 86. Sección Ciberderecho, 30 de julio de 2024 y sobre la conjugación de la innovación y la regulación, GARCÍA MEXÍA, P. "Europa ante el reto de la inteligencia artificial",The objective, 3 de agosto del 2024, (https://theobjective.com/tecnologia/2024-08-03/europa-ante-el-reto-de-la-inteligencia-artificial/) (recuperado el 5 de agosto del 2024).

4 Informe Model AI Governance Framework for Generative AI Fostering a Trusted Ecosystem, de 30 de mayo del 2024 (pág.5). En el Informe se muestra: -La rendición de cuentas: establecer la estructura de incentivos adecuados para que los distintos participantes en el ciclo de vida de desarrollo del sistema de IA sean responsables ante los usuarios finales; -Datos: garantizar la calidad de los datos y abordar los datos de formación potencialmente conflictivos de forma pragmática, y que son fundamentales para el desarrollo de modelos; -Desarrollo de confianza y despliegue: aumentar la transparencia en torno a las medidas

II.1.- EL CONCEPTO DE IA GENERATIVA

El concepto de IA generativa no aparece en la LIA, ni en sus definiciones (art.3 LIA). Si bien, sí define la IA que la engloba, la IA de uso general, así, para la LIA (en su Considerando 99) "los grandes modelos de IA generativa son un ejemplo típico de un modelo de IA de uso general, ya que permiten la generación flexible de contenidos, por ejemplo, en formato de texto, audio, imágenes o vídeo, que pueden adaptarse fácilmente a una amplia gama de tareas diferencias[5]". En el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, se expresa que la Inteligencia Artificial General (IAG)

básicas de seguridad e higiene basadas en las mejores prácticas del sector, en desarrollo, evaluación y divulgación; -Notificación de incidentes: implantación de un sistema de gestión de incidentes para la notificación y corrección oportunas y mejoras continuas, ya que ningún sistema de IA es infalible; -Pruebas y garantía: proporcionar validación externa y confianza añadida mediante pruebas de terceros, y desarrollar normas comunes de pruebas de IA en aras de la coherencia; -Seguridad: hacer frente a los nuevos vectores de amenaza que surgen gracias a los modelos generativos de IA; -Procedencia de los contenidos: transparencia sobre la procedencia de los contenidos como señales útiles para los usuarios finales;- I+D en seguridad y alienación: acelerar la I+D mediante la cooperación mundial entre los Institutos de seguridad de la IA para mejorar la alienación de los modelos con la intención y los valores humanos; -IA para el bien público: la IA responsable incluye el aprovechamiento de la IA en beneficio del público mediante la democratización del acceso, la mejora de la adopción por parte del sector público, la mejora de las cualificaciones de los trabajadores y el desarrollo sostenible de los sistemas de IA.

5 Esta idea ya aparecía en el art.3.5 del Real Decreto 817/2023, de 8 de noviembre, por el que se establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial. Posteriormente la LIA, distingue y nos define el "modelo de IA de uso general" (art.3 definición 63) y el "sistema de IA de uso general" (art.3 definición 66) Siendo el primero: "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" y el segundo como" un sistema de IA basado en un modelo de IA de uso general y que puede servir para diversos fines, tanto para su uso directo como para su integración en otros sistemas de IA". Su capítulo V se dedica a "los modelos de IA de uso general", apareciendo en su sección primera las reglas de clasificación (arts.51 y 52 LIA). En el Considerando 97 se incide en la importancia de la distinción entre modelo y sistema de IA de uso general para garantizar la seguridad jurídica.

"se refiere comúnmente a la IA que posee la capacidad de comprender, aprender y realizar una amplia gama de tareas a un nivel que iguala o supera las capacidades humanas. Contrasta con la IA restringida, que solo puede realizar una tarea específica"[6]. La IA generativa, añade el documento, son modelos generativos que aprenden la distribución subyacente de los datos y pueden generar nuevos contenidos (literatura, audio, vídeos) a partir de esta distribución aprendida. Concepto que queda matizado en el Informe Model AI Governance Framework for Generative AI Fostering a Trusted Ecosystem, de 30 de mayo del 2024 [7] al señalar que «son modelos de IA capaces de generar texto, imágenes u otros tipos de medios. Aprenden los patrones y la estructura de sus datos de entrenamiento de entrada y generan nuevos datos con características similares. Los avances en las redes neuronales profundas basadas en transformadores permiten que la IA generativa acepte como entrada indicaciones en lenguaje natural, incluidos los grandes modelos lingüísticos (LLM) como GPT-4, Gemini, Claude y LlaMA".

II.2.- BENEFICIOS DE LA IA GENERATIVA

La IA, en general, y la generativa en particular, se está integrando prácticamente en todos los ámbitos de la vida,[8] aportando

6 En el Informe se muestra la evolución de sus versiones, desde resolver tareas específicas a los modelos de base y los modelos fundacionales.

7 Nota 3 pág.3

8 La propia LIA lo expresa en su Considerando 4 al mostrar los beneficios de la IA: "La IA es un conjunto de tecnologías en rápida evolución que contribuye a generar beneficios económicos, medioambientales y sociales muy diversos en todos los sectores económicos y las actividades sociales. El uso de la IA puede proporcionar ventajas competitivas esenciales a las empresas y facilitar la obtención de resultados positivos desde el punto de vista social y medioambiental en los ámbitos de la asistencia sanitaria, la agricultura, la seguridad alimentaria, la educación y la formación, los medios de comunicación, el deporte, la cultura, la gestión de infraestructuras, la energía, el transporte y la logística, los servicios públicos, la seguridad, la justicia, la eficiencia de los recursos y la energía, el seguimiento ambiental, la conservación y restauración de la biodiversidad y los ecosistemas, y la mitigación del cambio climático y la adaptación a él, entre otros, al mejorar la predicción, optimizar las operaciones

múltiples beneficios, con un crecimiento exponencial[9]. La transformación de la IA generativa en la forma de generar contenidos va a afectar a todos los aspectos de nuestra forma de vivir, trabajar y jugar. Lo muestra el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023[10], que a su vez, reconoce que se ha descubierto el valor de los Chatbots como valiosos asistentes y el valor de la IA generativa en campañas de marketing de éxito, en el diseño de productos, de moda y de fármacos, en la mejora de los modelos de diagnóstico (sin afectar la privacidad de los pacientes), en la creación del gemelo del paciente, para aplicarse en medicina y precisión de ensayos clínicos. Muestra también su valor en la industria de entretenimiento, en el cine y en las plataformas de búsqueda en línea, con resultado que no serán hipervínculos sino conversacional. En el sector público, afirma el mencionado documento, hace a los ciudadanos más eficientes y accesibles los servicios públicos y puede ayudar en tareas de infraestructuras públicas y en el cambio climático de las ciudades inteligentes. Aporta beneficios como la posibilidad de mejorar la productividad, la accesibilidad y la diversidad de los contenidos mediáticos. Permite a su vez desarrollar nuevas creaciones o versiones mejoradas, agilizar procesos de ejecución, ofrecer nuevas líneas de negocio, etc.[11] La cultura se enriquece con sus creaciones.

y la asignación de los recursos, y personalizar las soluciones digitales que se encuentran a disposición de la población y las organizaciones".

9 Sobre el crecimiento de la misma véase, la encuesta e informe "El estado de la IA a principios de 2024: la adopción de la IA generativa aumenta y comienza a generar valor", McKinsey& Company, 30 de mayo de 2024, (SINCLA, A et al) (https://www.mckinsey.com/locations/south-america/latam/hispanoamerica-en-potencia/el-estado-de-la-ia-a-principios-de-2024-la-adopcion-de-la-ia-generativa-aumenta-y-comienza-a-generar-valor/es-CL) (recuperado el 2 de agosto de 2024).

10 Pág.7

11 Así afirman FRANGANILLO, J., "La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos", methaodos.revista de ciencias sociales (2023) 11(2) m231102a1010.17502/mrcs.v11i2.710, pág.3. y SUÁREZ JAQUET, H. et HINOJAL CUADRADO, E. "El uso del deepfake en producciones audiovisuales: consideraciones jurídicas", coordinador: ORTEGA BURGOS, E., Propiedad intelectual, 2022, Documento TOL9.141.396, pág.1

Los propios artistas, a su vez, utilizan las redes neuronales para realizar sus obras.

En nuestro devenir diario aparece constantemente, por ejemplo, en buscadores y navegadores[12]. Son múltiples las apps que nos ofrecen sus beneficios (por ej. Chat GPT, Lensa, Dalle 2).

El Informe del SEPD "La IA generativa y el EUDPR. Primeras orientaciones del SEPD para garantizar el cumplimiento de la protección de datos al utilizar sistemas de IA", de 3 de junio de 2024, reconoce que "la IA generativa, al igual que otras tecnologías, ofrece soluciones en varios campos destinadas a apoyar y mejorar las capacidades humanas. Sin embargo, también plantean retos con posibles repercusiones en los derechos y libertades fundamentales que corren el riesgo de pasar desapercibidos, pasarse por alto o no ser debidamente considerados y evaluados"[13]. En este panorama de bondad, la IA generativa, conlleva también riesgos para los derechos fundamentales, la sociedad, la economía y la democracia.

II.3.- LOS RIESGOS DE LA IA GENERATIVA

Los contenidos generados por la IA generativa

Se van a distinguir diferentes riesgos provocados por la IA generativa[14]. Antes de analizar los puntos siguientes, hay que tener presente el concepto de datos personales, ya que pueden verse afectados por estas creaciones, como "toda información sobre una persona física identificada o identificable ("el interesado");

12 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.11 añade los paquetes ofimáticos, bases de datos científicas y programas de edición (imagen y vídeo).

13 Pág.6

14 Sobre esta materia en EEUU, véase el informe sobre el Marco de Gestión de Riesgos de la Inteligencia Artificial: Perfil de la Inteligencia Artificial Generativa, del NIST AI 600-I (borrador público inicial), abril 2024, NIST AI 600-I, julio de 2024 (recuperado el 2 de agosto del 2024).

se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona" (art.4.1 Reglamento (UE) 2016/679 de Protección de Datos, concepto al que remite el art.3 definición 50 LIA). Conforme a esta definición, tal y como afirman SUÁREZ JAQUET, H et al. [15] la apariencia, la voz y los gestos entrarían en esta categoría de datos a efectos legales.

La IA generativa puede crear textos a través del uso de los modelos de lenguaje (por ejemplo Generative-Pre-trained Transformer y LLaMa), son tan reales que parecen auténticamente escritos por personas. Crea imágenes originales muy realistas, usando redes generativas antagónicas [16] con indicaciones en lenguaje natural.

La tecnología deepfake utiliza la IA generativa para crear vídeos sintéticos tan reales que pueden hacerte creer que lo irreal es real se califican así de hipertrucados. En la LIA se utiliza el término "ultrasuplantación" y se define como "un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos" (concepto 60, art.3).

15 SUÁREZ JAQUET, H et al. "El uso…op.cit.pág.4. Sobre la evolución del derecho a la protección de datos de carácter personal: el algoritmo transparente y responsabilidad-accountability, véase BENDITO CAÑIZARES, M.T, "Estadio intermedio de reflexión para una futura regulación de la ética en el espacio digital europeo: los principios de transparencia y accountability", *Revista Aranzadi de Derecho y Nuevas Tecnologías* , núm 55/2021, BIB 2021/1465.

16 Sistemas de IA que se entrenan mediante Deep learning partiendo de gran cantidad de datos. Por ejemplo, el uso de Craiyon, DALL-E, Midjourney y Stable Diffusion, recogidos por MEIJOMIL, S., Inboundcycle, "6 generadores de imágenes con IA que no puedes perderte, 8 de noviembre del 2023, (https://www.inboundcycle.com/blog-de-inbound-marketing/generadores-de-imagenes-con-ia) (recuperado el 2 de mayo del 2024).

El término deepfake tiene su origen en la combinación de los términos deep learning y fake para referirse a un contenido falso generado con técnicas de aprendizaje profundo ("deep learning") [17]. Definido el deepfake por SUÁREZ JAQUET, H et al. Como "una técnica de IA que permite editar vídeos falsos de personas que aparentemente son reales, utilizando para ello algoritmos de aprendizaje no supervisados, conocidos en español como RGAs, y vídeos o imágenes ya existentes. El resultado final de dicha técnica es un vídeo muy realista, aunque ficticio"[18]. Los vídeos generados por esta técnica pueden "revivir" a personas fallecidas[19], hacer decir cosas que nunca se dijeron, hacer cosas que nunca hicieron o transformar totalmente la apariencia de los protagonistas. La manipulación de los vídeos, anteriormente se ceñía a trabajos manuales tediosos para modificar el contenido original (recortar, editar) pero no a las palabras o apariencia de los

17 El Estudio, la política europea frente a los deepfakes, de julio del 2021 lo recoge y señala su origen en Reddit (págs.2 y ss.). Lo muestra como un subconjunto de una categoría más amplia de «medios sintéticos» generados por IA, que no sólo incluye vídeo y audio, sino también fotos y texto. Sobre la diferencia entre deep fake y medios sintéticos, (págs.2 y ss.), Deepfake y tecnologías de medios sintéticos (págs.7 y ss): fotografía y videografía (pág.7), técnicas específicas de deepfake gráfico (págs.8 y ss.), de clonación de voz (págs.12 y 13) y síntesis de texto (13), nuevas tendencias y futuro y evolución de riesgos (págs.13 y ss.). Dicho Estudio desarrolla el panorama normativo europeo (y las lagunas) sobre los deepfakes (págs.37 y ss.) y las dimensiones de las medidas políticas para mitigar el impacto negativo de los deepfakes (dimensión tecnológica, de la creación, de la circulación, del objetivo y de la audiencia) (págs.58 y ss.).

18 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág. 4 e 1 y 2. FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.13, afirma que se nutre de datos e información derivada del comportamiento humano y, aplicada a determinados sectores y servicios, es capaz de imitarlo. El deepfake usa algoritmos de IA denominados Redes Generativas Antagónicas propios del "Deep learning. Añade en su pág.2 SUÁREZ JAQUET, H. et al. en qué consisten las RGAs. El Estudio, la política europea frente a los deepfakes, de julio del 2021, los define como medios sonoros o visuales manipulados o sintéticos que parecen auténticos, en los que aparece(n) una(s) persona(s) que parece(n) decir o hacer algo que nunca dijo (dijeron) o hizo (hicieron), producidos mediante inteligencia artificial o aprendizaje automático (págs.XIII y 2).

19 Por ejemplo, recientemente se ha vuelto a dar vida a Marilyn Monroe y Sean Connery en el thriller de espías, DUCK, de Rachel Maclean presentado en el Festival Internacional de Cine de Rotterdam 2024 y a Marilyn en solitario en un vídeo para ayudar a la gente a dormir. "Marilyn conoce a James Bond en la película Deepfake del artista", El informe de Marilyn, 5 de marzo de 2024, "https://themarilynreport-com.translate.goog/2024/03/05/marilyn-meets-james-bond-in-artists-deepfake-movie/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc (recuperado el 2 de agosto del 2024).

"protagonistas"[20]. El poder recrear o revivir personajes de forma realista se originó en el cine con la técnica CGI (computer generated imagery) [21], muy utilizada además en publicidad; por ejemplo, "ha revivido" a Lola Flores, en la campaña publicitaria "Con mucho acento" (2021) de Cruzcampo por la que la agencia Ogilvy recibió el Gran Premio a la Eficacia y dos premios EFI de Oro.[22]

Por otra parte, la IA generativa puede crear vídeos sintéticos partiendo de unas simples instrucciones o indicaciones de texto, técnicamente menos evolucionados, sin perjuicio de una futura evolución más realista[23]. Otro contenido que genera esta IA es la voz. La voz es un elemento característico de la persona, como vimos, puede calificarse de dato personal, y por tanto el uso que se haga de ella por terceras personas debe ser responsable.

La síntesis de voz es una tecnología que permite convertir texto en una voz muy parecida a la humana que ya se estaba utilizando (asistentes de voz, voz en la navegación GPS...) aunque ha avanzado, ampliándose las posibilidades de su uso comercial (audiolibros, proyectos publicitarios, etc.), llegándose a utilizar voces sintéticas de personas fallecidas[24].

20 FRANGANILLO, J., "La inteligencia artificial...op.cit. pág.6

21 IBIDEM., pág.1, muestra cómo se sustituía digitalmente el rostro de un artista fallecido durante el rodaje por el de un doble y otra técnica: el "efecto Forrest Gump", originada en dicha película mezclando elementos históricos y actuales en las imágenes. SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.2. muestran películas, en esa línea como, por ejemplo, Parque Jurásico y la Trilogía del Señor de los Anillos.

22 BORRACHERO GARRO, A. "Los retos de la tecnología en la publicidad: la campaña de Lola Flores", coordinador: ORTEGA BURGOS, E., *Propiedad intelectual*, 2022, Documento TOL9.141.406, págs.13 y ss., sobre la campaña de Lola Flores. Por ejemplo rejuvenece a personajes de ficción, con la tecnología digital rejuvenecedora ("de-aging technology") como el interpretado por Harrinson Ford en Indiana Jones y el dial del destino, y en películas como El curioso caso de Benjamin Button, también se utiliza para suplantar personas en programas satíricos de entretenimiento ("Entrevista por la cara"), hacer vídeos divertidos con rostros de estrellas (ej. Nicolas Cage) y suplanta a personajes de películas (FRANGANILLO, J., "La inteligencia artificial...op.cit.pág13 y SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.3).

23 Como muestra FRANGANILLO, J., "La inteligencia artificial...op.cit. pág.10.

24 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.3, muestra como la Compañía: "Flawless AI" Software"TruySync" mejora el doblaje de las películas. Lo que también lleva

Desde el punto de las creaciones sintéticas musicales, puede tocar todos los puntos expuestos, ya que está por un lado la composición de letras de canciones, la música de las mismas, la posible utilización de voces sintéticas (de autores conocidos o no) y la realización de vídeos musicales[25]. Todo ello con los correspondientes problemas que pueden surgir en materia de propiedad intelectual. También se pueden crear, entre otros contenidos, códigos.

Los riesgos de desinformación y confusión

El que parezcan creaciones humanas los textos, imágenes[26], vídeos y voces originados por la IA generativa puede provocar, consciente o inconscientemente, desinformación, confusión, engaño, desconfianza y contenidos de mala o baja calidad, afectando, por ejemplo, al derecho a una información veraz[27].

posibles problemas legales ya que la interpretación está alterada al sustituir el movimiento de sus labios por los del actor de doblaje siendo necesario el consentimiento expreso del actor. Por otra parte, FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.10, señala que se clonan voces de famosas para usarlas, con ética y con licencia, en audiolibros, productos audiovisuales y entornos inmersivos, voces en off o de doblaje (pionera la empresa Veritone que ofrece voces de celebridades, actores, actrices, deportistas y otras personas influyentes para cualquier proyecto sonoro o multimedia recibiendo royalties por el uso comercial de su voz). La usan medios de comunicación para otros fines por ej. para narrar la noticia con la voz de un periodista o presentador famoso.

25 DAVID, DemoCreator, "Los 10 mejores generadores de música IA gratis en 2024", 13 de marzo de 2024 (https://dc.wondershare.es/ai-voice/top-free-ai-music-generators.html) (recuperado el 7 de mayo del 2024). En ellas se puede dar un texto para que genere música, voces y vídeos musicales. En la lista que expone, señala (reconociendo además, en algunas, su posible comercialización) expresamente las que están libres de derechos de autor : Stability.ai, Beatoven.ai, Loudly, Soundful, Mubert. Añade otras sin aludir a los derechos de autor: Boomy, Soundraw.io, MusicaStar.AI., Riffusion, Suno AI. Posteriormente habla de la creación de vídeo musical con Wondershare Demoheator. Incide mucho en general,en la facilidad de uso, basta con unos clics.

26 La IA generativa se ha usado para crear "falsas" fotografías en prensa como ocurrió, entre otras, con las que se hicieron virales, entre marzo y abril del 2023, del expresidente Donald Trump forcejeando con la policía, tras la actuación de Midjourney otras de las escenas falsas eran del Papa Francisco con un abrigo de plumas o el abrazo de Yolanda Díaz y Pablo Iglesias. Esto provoca confusión, desconfianza, engaño y desinformación (Véase FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.5 y SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.2)

27 Sin embargo, la creación de imágenes conceptuales es menos problemática. FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.12

Los deepfakes, sobre todo, generan "confusión" por el grado de sofisticación del falseamiento de la realidad en los vídeos. Se califica así este contenido de ultrafalso o hipertrucado. Reconoce FRANGANILLO que hoy la IA generativa es válida para situaciones que admiten cierto margen de error, cierta superficialidad argumental e incluso algún disparate, pero no lo es para cuestiones críticas (un trabajo científico, un consejo legal o financiero o una consulta médica). Produce "una engañosa ilusión de pensamiento racional", no entiende en un sentido humano nada de lo que escribe. Puede que su contenido sea incorrecto, ya que no dispone de un modelo de verdad y no siempre se apoya en fuentes fiables o evidencias robustas[28].

Un uso que, independientemente de los problemas legales que pueda generar, puede hacerse de forma maliciosa, por ejemplo, para abrir la puerta a los ciberdelitos a través de la suplantación de las personas. Debe tenerse presente como expone el Estudio, la política europea frente a los deepfakes, de julio del 2021 que estas actuaciones además de daños psicológicos y sociales conllevan daños financieros y perjuicios económicos (extorsión, robo de identidad, fraude, manipulación de precios de acciones, daños de marca y de reputación)[29].

Las creaciones sintéticas de IA generativa (imagen, vídeos, textos y voz) pueden como se ha visto, consciente o inconscientemente generar confusión y desinformación en el ciudadano al creer que son reales, por la perfección técnica alcanzada. Es difícil determinar su falsedad.

Son claros riesgos de estas tecnologías, como muestra BORRACHERO GARRO [30]por ejemplo, desde el punto de vista penal: la mencionada suplantación de personalidad (imagen y voz (deepvoice): para llamadas delictivas), enviar a la población mensajes

28 IBIDEM; pág.12

29 Págs.30 y 31.

30 BORRACHERO GARRO, A. "Los retos...op. cit. págs.17 y 18.

erróneos (fake news), imitar gestos o captar movimientos o señas de identidad con el fin de usarlos para dañar la reputación de una persona relevante (el deepfake es habitual en grabaciones pornográficas o campañas electorales con fines políticos). Es claro el uso de las mismas para generar contenido pornográfico, superponiendo la imagen de artistas famosas en un material pornográfico preexistente[31]. En el Estudio, la política europea frente a los deepfakes, de julio del 2021, se propone ampliar el marco jurídico actual en materia de delitos, se afirma que "teniendo en cuenta el daño que pueden causar los usos malintencionados de deepfakes, una evaluación de la solidez de las normas y reglamentos existentes a nivel de los Estados miembros podría ser útil para valorar si es necesario/ deseable añadir y especificar los delitos penales existentes. En Alemania, por ejemplo,

31 SUÁREZ JAQUET, H et al. "El uso…op.cit.pág.2. Como reitera FRANGANILLO, J. "La inteligencia artificial…op. cit. pág.9, "La amenaza es real: el 96%d de los vídeos deepfake publicados en línea en 2019 eran pornográficos y no consentidos, siendo las mujeres el colectivo más afectado (Ajder et al, 2019). Y en 2020 se identificaron más de 85.000 vídeos dañinos contra la reputación de figuras públicas, creados a un ritmo que se duplicaba cada seis meses (Ajder, 2020). Los algoritmos de aprendizaje profundo se entrenan con infinidad de imágenes que brinda internet, pero la tecnología avanza tan rápido que cada vez necesita menos datos de entrada para lograr un nivel similar de realismo (Giansiracusa, 2021)". Por su parte, VALERO, A., "Deepfakes Porn y violencia contra las mujeres", Fundación Cañada Blanch, 4 de junio de 2024, https://www.fundacioncanadablanch.org/noticias/deepfakes-porn-y-violencia-contra-las-mujeres/ (recuperado el 2 de agosto del 2024), muestra que el Informe State of Deepfakes 2023, de la empresa Home Security Heroes, afirma "que el 98% de los deepfakes que hay en Internet son pornográficos; que 7 de cada 10 sitios web pornográficos alojan deepfake porn y que El 99% de las personas que aparecen en las deepfake pornográficas son mujeres". En Almendralejo (Badajoz) unos menores compañeros de instituto o amigos, manipularon y difundieron imágenes, de un grupo de 20 chicas menores de edad, desnudas elaboradas con Inteligencia Artificial, (VIGARIO, D.,"Un año de libertad vigilada para los 15 jóvenes que manipularon y difundieron imágenes con IA de menores desnudas en Almendralejo", El mundo, 9 de julio de 2024 (recuperado en 1 de agosto del 2024)). Sobre este ejemplo y otros, en la misma línea en México, EEUU, véase LUCIO LÓPEZ, L.A., "Deep fake porn, la inteligencia artificial da nueva cara al ciberacoso escolar", Ciem, 2024. Recientemente aparecieron imágenes manipuladas con IA en X, deTaylor Swift desnuda (DURAN, I., "Taylor Swift y sus desnudos hechos con IA: CEO de Microsoft dice "hay que actuar ya" ante los deepfakes, Infobae, 29 de enero del 2024, https://www.infobae.com/tecno/2024/01/27/taylor-swift-y-sus-desnudos-hechos-ia-ceo-de-microsoft-dice-hay-que-actuar-ya-ante-los-deepfakes/ (rescatado el 3 de agosto del 2024)). Sobre este punto, ÁLVAREZ, P. y EGUILUZ, J. " El Reglamento de IA ante los deepfakes de desnudos", 2 de octubre del 2023, Cuatrecasas (https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/el-reglamento-de-ia-ante-los-deepfakes-de-desnudos) (recuperado el 1 de mayo del 2024).

está prohibida la distribución de un deepfake que viole los derechos de la persona (como la pornografía deepfake), pero no su producción"[32].

Riesgo de ataque a los derechos al honor y la propia imagen

De lo expuesto, se observa cómo se ven afectados, derechos fundamentales de la persona como el derecho al honor, la intimidad y la propia imagen (art.18 CE y LO 1/1982, De 5 de mayo, sobre protección al derecho al honor, a la intimidad personal y a la propia imagen). Nuestra privacidad, nuestros datos personales se ven atacados[33] para utilizarlos, de forma maliciosa o no, incluso delictiva. Ponen en tu boca palabras que nunca dijiste y tu imagen realiza hechos que nunca hiciste, creando falsas o perversas realidades, tengas o no una proyección pública (periodista, político, famosos...) afecta a tu reputación y credibilidad. Pueden ser el hilo transmisor de noticias falsas manipulando la opinión pública, provocando la desinformación.

Tenemos un derecho personalísimo a la propia imagen, que incluye la apariencia física, la voz, o la semejanza o parecido físico, (en definitiva "datos personales"[34]), además del derecho al honor y la intimidad. Datos que se están utilizando. La única forma de admitir el uso por terceros de estos datos personales es con el necesario consentimiento de los dueños de la imagen o la voz, o de sus herederos para que no se produzca una intromisión ilegítima al derecho al honor, la intimidad y la propia imagen[35]. El

32 Estudio, la política europea frente a los deepfakes. Panel para el futuro de la Ciencia y la Tecnología dirigido por Philip Boucher, EPRS/ Servicio de Estudios del Parlamento Europeo, Unidad de Prospectiva Científica (STOA), PE 690.039-julio de 2021, pág.61. Seguido por ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA ante...op.cit.

33 SUÁREZ JAQUET, H. et al. "El uso...op. cit. pág.1

34 IBIDEM; pág.5.

35 Con respecto al derecho fundamental a la propia imagen como expresa BORRACHERO GARRO, A. ("Los retos...op. cit. págs.15 y 16): Se produce un desdoblamiento del derecho fundamental a la propia imagen. El derecho fundamental a la propia imagen

uso de la voz, imágenes y vídeos de personas fallecidas genera además un debate ético, aunque jurídicamente se puede legalizar dicho uso si lo consienten los herederos, pero ¿hasta qué punto lo habrían querido ellas?[36]

Por otra parte, es necesario que expresamente se identifique cuando una IA generativa es la autora de la creación, siendo una creación sintética y no humana. Se evita así la confusión, como ha sucedido, por ejemplo, en concursos de arte y fotografía, en los que se desconocía el origen sintético de las creaciones premiadas[37](por ejemplo, en materia de doblaje se pide que se distinga al oírla - que tenga "acento de IA"-, otra forma sería insertar una huella digital indeleble indicándolo). Es esencial también,

---

se extingue con la muerte pero subsiste la protección a la memoria del fallecido (su tutela corresponde a las personas que establece el art.4.2 LO 1/1982)(SUÁREZ JAQUET, H et al. "El uso…op.cit.pág.5.). El aspecto patrimonial del derecho de imagen, es un bien jurídico diferente que no está incluido en el contenido esencial del derecho fundamental y puede protegerse a nivel de legalidad ordinaria (aunque dicha Ley no contempla expresamente la transmisión patrimonial mortis causa puede tener cabida en nuestro ordenamiento jurídico), para evitar conflictos lo mejor es que "se recabe el consentimiento de los titulares del derecho de imagen de la persona fallecida refiriéndonos a la vertiente patrimonial a la vez que se solicita una renuncia adicional a emprender acciones legales por la defensa del derecho fundamental". Muestran SUÁREZ JAQUET, H et al. ("El uso…op.cit.págs. 4 y 5) como se considera "intromisión ilegítima" la "captación , reproducción o publicación por fotografía, filme o cualquier otro procedimiento, de la imagen de una persona en lugares o momentos de su vida privada o fuera de ellos" (art.7.5 LO 1/1982) y el art.7.6 LO 1/1982 extiende la consideración de intromisión ilegítima" a la utilización del nombre, de la voz o de la imagen de una persona para fines publicitarios, comerciales o de naturaleza análoga". Expresan que el derecho a la propia imagen no es absoluto, puede entrar en conflicto con la libertad de expresión del creador (constitucional también), que protege al mismo para difundir sus ideas, pensamientos y opiniones de forma libre y sin censuras, pero en el caso del deepfake no vale ya que hay "riesgo de confusión", imagen irreal pero que quiere que se perciba como real. Sí se considera, por ejemplo, legítima según el art. 8.2 b) LO/1982" la utilización de la caricatura de personajes públicos, de acuerdo con el uso social". El derecho a la propia imagen y al honor prevalece sobre el de libertad de expresión cuando hay descrédito o trato vejatorio de la persona.

36 Véase BORRACHERO GARRO, A. "Los retos…op.cit. pág.19; FRANGANILLO, J., "La inteligencia artificial…op.cit.pág.13.

37 Por ejemplo, el premio Sony Wordl Photography Awards, lo ganó Boris Eldagsen, lo rechazo posteriormente al reconocer que no era una fotografía (MORAN, I., Photolari, 2 de mayo, 2024, "20.000 euros por la imagen IA que engañó al jurado de los Sony Wordl Photography Awards el año pasado" ((https://www.photolari.com/20-000-euros-por-la-imagen-ia-que-engano-al-jurado-de-los-sony-world-photography-awards-el-ano-pasado/) (recuperado el 20 de mayo del 2024).

establecer mecanismos de verificación y transparencia que permitan identificar la fuente y la autenticidad de las grabaciones sonoras[38]. No debemos olvidar que cualquiera actualmente puede acceder a esta IA generativa fácilmente, sin conocimientos y sin gasto o con un pequeño gasto (lo cierto es que esto favorece la democratización cuando su uso es correcto). Se requiere un uso responsable.

Las personas tienden a sobreestimar su capacidad de detectar engaños o manipulaciones, lo que les lleva a carecer de una actitud crítica. Por ejemplo, se incrementan las fake news a las que el público les concede absoluta credibilidad[39]. Por ello, además de la formación del público ante estas situaciones, la ingeniería debe intensificar los avances técnicos, la creación de herramientas proactivas de detección y buscar modos de informarle de que el contenido es sintético y potencialmente malicioso[40].

El estudio, la política europea frente a los deepfakes, de julio del 2021[41] destaca entre sus preocupaciones, la detección (manual o automática) y la prevención de deepfakes: "Debido al papel central que desempeñan las plataformas en línea y otros intermediarios en la difusión de deepfakes, la Comisión plantea obligar a dichas plataformas e intermediarios a disponer de un software de detección de deepfakes como requisito previo para un posible

38 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.13.

39 IBIDEM., pág..13. Véase Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, pág.16

40 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.13

41 En el Estudio, la política europea frente a los deepfakes de julio de 2021, sobre la detección (manual o automática) y prevención (ataques contra los algoritmos de deepfake, refuerzo de marcadores de autenticidad de los materiales audiovisuales y ayudas técnicas para que la gente los detecte más facilmente) de deep fakes se resalta que no es difícil evadir la detección (p. III y ss, y 17 y ss.). Sobre los sistemas de "marcado" ESPUGA TORNÉ, G. "Cómo identificar contenido generado por IA" Linkedin, junio, 2024 (recuperado el 10 de julio). Herramientas que permiten detectar textos generados por Inteligencia Artificial (Metadatos, Marcas de agua, Fingerprinting, herramientas de detección). HOLCOMBE, J., "Las 9 Mejores Herramientas de Detección de Contenidos con IA que tienes que conocer", Kinsta, 5 de abril del 2023 https://kinsta.com/es/blog/deteccion-de-contenidos-ia/ (recuperada el 6 de julio del 2024).

etiquetado. Una alternativa para la detección es el uso de filtros de carga –por ejemplo, que los rostros aparezcan difuminados hasta que las personas retratadas den su consentimiento–. Aunque esta opción también conlleva inconvenientes como los límites de la técnica o los peligros de censurar expresiones artísticas, vulnerando la libertad de expresión"[42]. Se había planteado, la futura regulación de las IA generativas en la reunión del G7 de 19-21 de mayo de 2023[43], viendo los riesgos de las mismas y ante la posibilidad de que la falta de regulación permitiera que las grandes compañías empezaran a autorregular el entrenamiento de sus IA generativas (sin olvidar la participación Italiana que prohibió temporalmente el uso de Chat GPT).

No obstante, la tecnología deepfake también se ha utilizado para concienciar al público de sus riesgos, que es necesario contrastar la información que aparece en internet y en los medios de comunicación. Ser diligentes e identificar las fuentes fiables de las que no lo son tal y como afirman SUÁREZ JAQUET, H et al [44].

42 Pág.62.

43 GARAY, J. ,Wired, 19 de mayo del 2023,"El G 7 promete regular las IA generativas antes de que termine el 2023"(https://es.wired.com/articulos/g7-promete-regular-las-ia-generativas-antes-de-que-termine-el-2023) (recuperado el 30 de mayo del 2024) , muestra esos efectos negativos: Los delitos que la utilizan y las consecuencias sociales, son evidentes y el mal uso de las mismas: por ejemplo deepfakes sexuales no consensuados, estafas que usan la voz de los familiares de las víctimas, imágenes que nunca existieron que provocan desinformación y la filtración de conversaciones privadas con chatbots. Exponen los miembros del G7 que debe regularse la IA y reconocen los riesgos y problemas que hay con las IA generativas. Las naciones más avanzadas del mundo no esperarán para que las grandes compañías (Open AI, Meta (LlaMA) Y Google (PaLM) establezcan reglas para el entrenamiento de sus IA generativas. Se establece "el proceso de la IA de Hiroshima".

44 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.2, el ejemplo al que se refieren es el del cómico Jordan Peele que en 2018 creó un vídeo en el que aparecía el expresidente de los Estados Unidos Barack Obama señalando cómo esta tecnología nos podía hacer creer cosas que nunca diríamos, al tiempo que estaba insultando a Donald Trump.

Riesgo de infracción o vulneración de los derechos de autor y la moderación de contenidos

La IA generativa, crea contenidos partiendo de datos ya existentes (entrenando con ellos). Para entrenar los sistemas de aprendizaje profundo se utiliza el raspado de datos web scraping, se recopila automáticamente un volumen grande de información accesible públicamente. Se cuestiona la licitud y legalidad de nutrirse con obras que normalmente tienen derecho de autor. Los propios creadores han contribuido al sistema al dar acceso a través de internet a sus obras. En la Unión Europea, el Reglamento General de Protección de Datos no establece que ese método sea ilegal, si bien restringe lo que se puede hacer con los datos extraídos. Una minería de textos avalada legalmente cuando su finalidad es la investigación científica no comercial[45].

Determinar la originalidad de las creaciones de la IA generativa, es complicado, no sabemos a ciencia cierta cuando se violan los derechos de autor de obras anteriores, por plagio. Otro problema es determinar quién es el "creador", el copyright, como señala MARTÍNEZ ESPÍN[46], el autor se plantea si las creaciones generadas por sistemas de Inteligencia Artificial son obras protegidas por derechos de autor. Afirma que el derecho de autor protege cualquier creación de la mente. Pero ¿mente humana o máquina? Preguntándose a su vez el autor si la creación de un sistema de IA ¿es una obra? ¿quién es el autor? ¿el sistema de IA, el programador o el usuario?

45 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.5. Sobre la minería de datos y textos véase MUÑOZ VELA, J. M, "Inteligencia artificial generativa. Desafíos para la propiedad intelectual", *Revista de Derecho UNED*, núm.33, 2024, Premio de artículos jurídicos "García Goyena", 22ª convocatoria (curso 2022-2023), Facultad de Derecho. UNED, págs.35 y ss.

46 MARTÍNEZ ESPÍN, P.,La propuesta de marco regulador de los sistemas de inteligencia artificial en el mercado de la UE ", Revista CESCO de Derecho de Consumo, nº46/2023, pág.3, (doi.org/10.18239/RDC_2023.46.3322).

En el mundo de la música, se han producido graves problemas desde el punto de la originalidad de las canciones creadas a través de la IA generativa, no solo por la utilización de parte de canciones existentes o música anterior para poder crearla, sino también, por la utilización de voz sintética de algunos autores sin que ellos lo sepan y hayan consentido.

FRANGANILLO[47] expresa que se podría proteger la propiedad intelectual señalando explícitamente la IA las fuentes que ha utilizado y que han entrenado al sistema. Manifiesta el autor que esta situación reclama una regulación y establecer formas de compensación para garantizar la viabilidad de la industria de medios y el ecosistema cultural del que se aprovechan hoy gratuitamente los modelos generativos. Las empresas de síntesis de voz han creado normas éticas y se aseguran de que la persona preste el consentimiento para que se clone su voz, controlando su uso y recibiendo una retribución. Debe, así, contarse con el consentimiento del dueño de la voz (o la imagen y resto de datos personales). Recientemente la actriz Scarlett Johanson paralizó el uso no consentido de una voz que "replicaba con gran exactitud" la suya cuando previamente había rechazado la propuesta de que se utilizara su voz en la versión más avanzada de ChatGPT (ChatGPT 40)[48].

MUÑOZ VELA[49] reconoce que los sistemas inteligentes generativos tienen capacidad creadora pero no creativa y que la IA actual carece de autonomía efectiva. Manifiesta, que dichos sistemas

47 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.15

48 MCMAHON, L., BBC News, 20 de mayo de 2024, "Cuando la escuché me quedé en shock": por qué el programa de IA ChatGPT dejará de usar la voz que se parece a la de Scarlett Johanson, https://www.bbc.com/mundo/articles/cprrzn8g2wqo.) (recuperado el 30 de mayo del 2024) en el artículo se muestra como inicialmente se le hizo una oferta para que prestara su voz a la nueva versión de Chat GPT. 40, se negó, y se utilizó una voz tan similar a la suya, la voz "Sky", teniendo en cuenta, el avance de las funciones de voz, su avance conversacional de Chat GPT 40,que la actriz ha conseguido la paralización de su uso.

49 MUÑOZ VELA, J. M, "Inteligencia artificial generativa. Desafíos...op.cit. págs.65 y ss., pág.67.

no pueden garantizar en el momento presente la originalidad y singularidad de sus resultados. Añade, que los marcos reguladores de la propiedad intelectual requieren la autoría humana y la originalidad del resultado para su protección (derechos de autor y conexos), negándola a las creaciones absolutamente artificiales, sin intervención humana o con intervención no relevante (en estos casos la IA no tiene la condición de autor ni su protección). Se plantea, si estos resultados tendrían protección a través del derecho de autor o del de la propiedad u otros. Expresa el autor que los datos y contenidos de los que se nutren los sistemas inteligentes (inputs), sí están protegidos, su uso requiere la autorización expresa del titular (salvo excepciones o limitaciones legalmente establecidas). Reconoce también la protección a través del derecho de autor de los prompts suministrados al sistema para crear el resultado (ouput), "si la secuencia de instrucciones evidencia un esfuerzo y complejidad cualitativa y cuantitativa que pueda determinar un resultado único y a las que se pueda asociar la autoría humana y la originalidad, considero que podrían resultar protegibles como PI, incluso como obra literaria/técnica desde su creación".

El Informe del SEPD 2024, en relación al consentimiento expresa que "El tratamiento de datos personales en el contexto de los sistemas de IA generativa requiere una base jurídica acorde con el Reglamento. Si el tratamiento de datos se basa en una obligación legal o en el ejercicio de la autoridad pública, dicha base jurídica debe establecerse de forma clara y precisa en la legislación de la UE. El uso del consentimiento como base jurídica requiere una cuidadosa consideración para garantizar que cumple los requisitos del Reglamento, a fin de que sea válido"[50].

50 Pág.17.En el Considerando 105 LIA se muestra que cuando se utilizan en técnicas de prospección de textos y datos, contenidos protegidos por el derecho de autor, se requiere la autorización del titular, "salvo que se apliquen las excepciones y limitaciones pertinentes en materia de derechos de autor. La Directiva (UE) 2019/790 introdujo excepciones y limitaciones que permiten reproducciones y extracciones de obras y otras prestaciones con fines de prospección de textos y en determinadas circunstancias".

Debe tenerse presente la responsabilidad de los prestadores de servicios, como establecen SUÁREZ JAQUET, H et al.[51], en virtud del art.73 del Real Decreto-ley 24/2021, de 2 de noviembre, se transpone al ordenamiento español las Directivas sobre derechos de autor y derechos afines en el mercado digital, en el uso de contenidos protegidos por parte de prestadores de servicios para compartir contenidos en línea (Instagram, Facebook, Twitter). Ellos tienen que obtener las correspondientes autorizaciones de los titulares de los derechos, si no se las otorgan responderán de los actos no autorizados de comunicación al público, salvo que demostraran que hicieron sus mayores esfuerzos por obtenerla y para garantizar la indisponibilidad de la obra y prestaciones. Por ello, como afirman los autores, si un deepfak no ha obtenido las autorizaciones pertinentes y esta circunstancia se verifica por los prestadores de servicios, podría considerarse como contenido que vulnera los derechos de autor y se debería retirar de la plataforma, para no asumir responsabilidad.

Recientemente Meta, en Instagram, por ejemplo, ha lanzado un formulario a sus usuarios para que expresen que no consienten que se utilicen sus datos (que están en la aplicación), para el entrenamiento de la IA, pero no ha comunicado expresamente esta opción. Dada la transcendencia, debería haberse informado de una manera más directa e individual, en la que quede un conocimiento claro de lo que implica aportar dichos datos personales al entrenamiento de la IA. Imágenes, vídeos, etc. subidos a la plataforma que pueden usarse y aparecer de una forma u otra donde menos te lo esperas. Siendo una autorización consciente y necesaria. Tendría que ser requisito de acceso a la plataforma el responder a dicha solicitud, se forzaría así obtener el consentimiento o la negativa de forma segura. Esta actuación se ha parado finalmente por Meta. La autoridad de protección de datos de Irlanda, Data Protection Commission

51 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.7

(DPC) en nombre de la Unión Europea no la ha permitido al afectar a la politica de privacidad. Por otra parte, la Fiscalía española abrió recientemente diligencias por esta actuación que ya había cesado. La actuación y normativa europea ha llevado a Meta a afirmar que no ofrecerá sus nuevos modelos de IA generativa en Europa[52].

GOLDMAN SACHS [53], destaca otra problemática de la IA generativa conectada con este punto, la moderación de contenidos. Muestra como las plataformas de redes sociales y sitios web, donde se publican contenidos por los usuarios no se hacen legalmente responsables de ellos (estableciendo una especie de escudo legal) y sí a los usuarios.

Riesgo del mercado laboral

La IA generativa sigue la estela de la IA en general, que transformará, en todos los sentidos, el ámbito laboral (con la creación, cualificación y desaparición de empleos. Nadie "tiene segura la silla", ni los CEOs[54]), e incluso afectará a los sistemas de seguridad

52 Sobre este punto, RAA J., "Meta detiene su proyecto para entrenar a la IA con publicaciones de Facebook e Instagram en Europa", Tecnología, El País, 14 de junio de 2024 (recuperado el 1 de julio del 2024) ; AFP, "La fiscalía investiga si Meta vulnera la protección de datos de sus usuarios", El Mundo, Empresas, 4 de julio de 2024 (recuperado el 7 de julio del 2024) y PASCUAL, M.G.,Meta no ofrecerá sus nuevos modelos de IA generativa en Europa por su "impredicible entorno regulatorio", El Pais, Tecnología, 18 de julio de 2024 (recuperado el 19 de julio del 2024).

53 GOLDMAN SACHS, Principales riesgos que entraña la inteligencia artificial generativa: enumeración, fundspeople, (https://fundspeople.com/es/principales-riesgos-que-entrana-la-inteligencia-artificial-generativa/), 26 de abril de 2023 (rescatada en 20 de abril del 2024). Goldman Sachs ha publicado un white titulado Generative AI- Part I: Laying Out the investment Framewor, en dicho trabajo académico los autores analizan los riesgos asociados a la IA generativa, rama que se centra en la generación de contenido original a partir de datos existentes. (5 riesgos: moderación de contenidos, desinformación, infracción de los derechos de autor, privacidad, cuestiones éticas).

54 Los chatbots están siendo utilizados para tomar decisiones de directivos que podrían ver peligrar sus puestos de trabajo, VIDAL. M, en LinkedIn recoge una noticia, el New York Times explora la tendencia de incorporar la IA como directora de algunas empresas, reconoce el potencial de la IA para reemplazar a los CEOs. (Fuente: https://www.nytimes.com/2024/05/28/technology/ai-chief-executives.html) (recuperado el 7 de junio de 2024).

social. Los gobiernos deberán ayudar a capacitar digitalmente a los trabajadores[55].

La utilización de la IA por su automatización y aumento de productividad puede hacer innecesarios ciertos trabajos, destruyendo así empleos. En materia cultural, de medios de comunicación y audiovisual, como muestra FRANGANILLO, están especialmente en peligro los diseñadores, los ilustradores, los fotógrafos, los dobladores y los guionistas [56]. En estos casos, nos encontramos con tecnología al alcance de cualquiera sin conocimiento de diseño gráfico ni de otro tipo al efecto (fotografía, etc.,). Puede crear contenido sintético (aun sin vulnerar la imagen de una persona), afectando al trabajo de los mencionados profesionales, que han visto devaluada su habilidad para presentar trabajos de calidad[57]. Podemos añadir que la creación de códigos con la IA generativa puede afectar al trabajo de los creadores de páginas web y a los programadores.

El mencionado autor muestra como la comercialización de estas creaciones generan ganancias y que, por ello, algunos bancos de imágenes han reescrito sus directrices para impedir la venta de materiales creados por la IA generativa. Termina reconociendo que la IA, para entrenar sus modelos, se abastece de trabajos

También para preparar las entrevistas de trabajo (ANDRÉS, R., "La última tendencia para bordar las entrevistas de trabajo: entrenar con ChatGPT como reclutador", Xataca, 6 de marzo del 2024 (recuperado el 28 de julio del 2024)).

55 Un estudio detallado de la evolución de la IA en el ámbito laboral lo ofrece MUÑOZ VELA, J.M, *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*, Thomson Reuters- Aranzadi, Pamplona, 2022, págs.36 y ss., aportando las teorías que ven una mejora con la misma y otras más catastrofistas. Por otra parte, La LIA recoge la importancia del empleo en los Considerandos 2,9,57 y 58 y en el Anexo III apartado 4.

56 FRANGANILLO, J. "La inteligencia artificial...op. cit. págs.13 y 17, recoge también la huelga del Sindicato de guionistas de Estados Unidos, que exige, entre otras demandas, que la IA no sustituya labores creativas, ni escriba ni reescriba material literario, ni se entrene con obras de guionistas y que el 36% de los trabajadores de la industria estadounidense del entretenimiento temen el impacto de la IA generativa en sus empleos, sobre todo por la vulneración de la propiedad intelectual (pág.17).

57 Véase IBIDEM.,. pág.17.

ajenos, afectando al derecho a la propiedad intelectual. Un "deber legal ético" no resuelto que exigiría recibir una compensación por su uso. El problema surge si sustituye el servicio de los profesionales. La Asociación de Medios de Información exige una nueva tasa por el aprovechamiento que la IA hace de sus contenidos sin reconocimiento ni retribución[58].

Riesgo para el medio ambiente

El derecho al medio ambiente es esencial. Se afirma en el Considerando 48 LIA, que hay que tener en cuenta, cuando se evalúe la gravedad del perjuicio que puede ocasionar un sistema de IA, el derecho fundamental a un nivel elevado de protección del medio ambiente consagrado en la Carta y aplicado en las políticas de la Unión.

La IA generativa, al igual que el resto de IAs y de tecnologías como Blockchain, tienen un gran consumo energético.

Nos muestra FIGUEROA como la IA generativa en sus creaciones consume mucha energía y contamina, al emitir gran cantidad de carbono. Por ejemplo, la acción que más consume es la creación de imágenes, comparada con la de texto[59], una imagen equivale a la carga de un celular. Lo afirma, como muestra FIGUEROA, un Estudio del startup IA Hugging Face y la Universidad Carnegie Mellon de Estados Unidos, dirigido por Sasha Luccioni. Descubrió

58 DEL CASTILLO, C., "Los creadores del canon AEDE quieren una "tasa ChatGPT" para la inteligencia artificial, el Diario.es, 3 de mayo del 2023, (https://www.eldiario.es/tecnologia/creadores-canon-aede-quieren-tasa-chatgpt-inteligencia-artificial_1_10171676.html), (recuperado el 20 de mayo del 2024).

59 Siguiendo a FIGUEROA, J.C., "Crear una sola imagen con inteligencia artificial consume tanta energía como cargar tu teléfono", Hipertextual, Tecnología, 12 de diciembre de 2023 (https://hipertextual.com/2023/12/crear-imagen-con-inteligencia-artificial-consume-esta-energia) (recuperado el 20 de abril del 2024), el autor muestra que el estudio de la startup IA Hugging Face y la Universidad Carnegie Mellon de Estados Unidos, afirma que: ejecutar 1.000 acciones de generación de texto, en una herramienta como ChatGPT, solo consume tanta energía como el 16% de la batería de tu celular. El autor a lo largo del artículo ofrece datos más detallados.

el estudio, a su vez, que el uso de grandes modelos generativos, como ChatGPT o Bard, consumía mucha más energía que los modelos de inteligencia artificial más pequeños diseñados para tareas específicas. El motivo es porque intentan hacer muchas cosas a la vez: generar, clasificar y resumir texto, en lugar de una sola tarea. Además, sus emisiones diarias exceden con creces las generadas durante el entrenamiento de modelos grandes. El impacto ambiental ofrece distintos frentes, por ejemplo, para enfriar los servidores de productos con ChatGPT son impresionantes las cantidades de agua necesarias. La industria busca cómo mitigar el impacto (principalmente por motivos económicos para reducir gastos, aunque conllevará una mejora en la calidad de vida) desarrollando estrategias, concretamente a través de la energía nuclear[60].

Riesgo por los sesgos (la necesaria alfabetización y democratización)

Otro de los riesgos a los que nos expone la IA generativa son los sesgos. La LIA no recoge la definición de sesgo, pero si hace alusiones expresas a los mismos[61].

Como expone FRANGANILLO[62] en conexión con la IA generativa (enfocada por él principalmente en el mundo audiovisual), los algoritmos de generación de imágenes, entrenados con datos sin filtrar, pueden reproducir estereotipos raciales (étnicos), culturales y de género incluidos en los datos, provocando sesgos. Los algoritmos no son neutrales y conviene corregir sus desviaciones para garantizar el principio de justicia y una actuación ética.

60 IBIDEM., recoge que Microsoft en mayo llegó a un acuerdo de compra de energía con Helion, un startup de fusión nuclear, para comprarle electricidad a partir de 2028 y también Sam Altman, director ejecutivo de OpenAI, quien también es uno de los inversores más importantes en Helion.

61 Concretamente en los Considerandos 27, 61,67, 70,110, arts.10.2 f) y g) y 5 a), e) y f); 14.4.b), 70.1, y Anexo XI, sección primera 2.c) LIA.

62 FRANGANILLO, J." La inteligencia artificial...op. cit. pág.13.

El sistema se debe calibrar y reentrenar. El autor muestra otros dos sesgos: un sesgo cognitivo, al tenerse la percepción pública distorsionada (por el cine y la literatura) por la creencia de que lo que dice la IA es cierto y otro sesgo porque la IA generativa suele seleccionar de la información accesible en internet la de fuente inglesa considerándola de mayor calidad, frente a la diversidad cultural. Se está intentando solucionar complementando las respuestas con el resultado de búsqueda en tiempo real.

Reconoce GOLDMAN SACHS [63], a su vez, que la precisión se puede mejorar si vuelven a entrenar con información mejorada pero el sesgo es más difícil al ser los seres humanos lo que entrenan estos modelos de IA. Un mal entrenamiento de los datos puede originar datos inexactos y sesgados.

Sobre los sesgos, el Informe del SEPD 2024, manifiesta el carácter prioritario de su minimización, al afirmar que "La aplicación de procedimientos y mejoras prácticas para minimizar y mitigar los sesgos debería ser una prioridad en todas las fases del ciclo de vida de los sistemas generativos de IA, para garantizar un procesamiento justo y evitar prácticas discriminatorias. Para ello, es necesario supervisar y comprender cómo funcionan los algoritmos y los datos utilizados para entrenar el modelo"[64]. Se habla de minimización y mitigación ya que son actualmente imposibles de erradicar de forma absoluta por la intervención del ser humano.

En la LIA se ve la preocupación por los sesgos que puedan afectar a la salud, la seguridad de las personas y negativamente a los derechos fundamentales o dar lugar a algún tipo de discriminación prohibida por el Derecho de la Unión. Especialmente ocurre cuando las salidas de datos influyan en las informaciones de entrada de futuras operaciones (art.10.2 f) LIA). Alude la LIA

63 GOLDMAN SACHS, "Principales...op.cit.

64 Pág.31

a las medidas adecuadas para detectarlos y prevenirlos[65] en los arts.10.2 g) y 10.5 a), e) y f) en conexión con los sistemas de IA de alto riesgo. Se muestra en el art.14.4 b) LIA un tipo de sesgo "el sesgo de automatización": "la tendencia a confiar automáticamente o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo" y se pide ser conscientes del mismo. También se conectan los sesgos con los sistemas generales con riesgos sistémicos[66].

Los sesgos nos llevan a la manipulación y la desinformación de nuevo[67].

El uso correcto de la IA generativa pasa por entenderla, comprenderla por todos los agentes de la cadena que entran en contacto con la misma. Por ello, los gobiernos deben proporcionar programas de concienciación para que el público comprenda el alcance de la IA generativa y aprenda a defenderse de las nuevas falsificaciones. La educación es esencial para preparar la ciudadanía ante la proliferación de contenidos artificiales, aprendiendo sobre la IA y capacitando a la sociedad para adaptarse a sus efectos[68]. La capacitación en el conocimiento de la IA generativa puede además facilitar la apertura a la nueva forma de concebir el empleo que está aportando la IA.

En la reciente Convención Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho de 17 de mayo del 2024[69] , su art.20 sobre alfabetización y competencias digitales, promueve ambas, dispone que "cada Parte fomentará y promoverá la alfabetización digital

65 Véanse los Considerandos 67 y 70 de LIA.

66 El Considerando 110 LIA, en relación a los sistemas generales con riesgos sistémicos reconoce que los tienen: "los modelos pueden dar lugar a sesgos dañinos y discriminación que entrañan riesgos para las personas".

67 Sobre los sesgos véase MUÑOZ VELA, J.M, *Retos*...op.cit.págs.56 y ss.

68 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.15.

69 13ª Sesión del Comité de Ministros (Estrasburgo, 17 de mayo de 2024), Comisión de Inteligencia Artificial (CAI), CM (2024)52-final.

y las competencias digitales adecuadas para todos los segmentos de la población, incluidas las competencias especializadas específicas para los responsables de la identificación, evaluación, prevención y mitigación de los riesgos planteados por los sistemas de inteligencia artificial".

La LIA muestra la necesaria alfabetización. Recoge expresamente que debe dotar a los proveedores, responsables del despliegue y personas afectadas de los conceptos necesarios para tomar decisiones con conocimiento de causa en relación con los sistemas de IA (Considerando 20).[70]Se define en el art.3 número 56 la "alfabetización en materia de IA" como "las capacidades, los conocimientos y la comprensión que permiten a los proveedores, responsables del despliegue y demás personas afectadas, teniendo en cuenta sus respectivos derechos y obligaciones en el contexto del presente Reglamento, llevar a cabo un despliegue informado de los sistemas de IA y tomar conciencia de las oportunidades y los riesgos que plantea la IA, así como de los perjuicios que puede causar" y el art.4 establece sobre la alfabetización en materia de IA que "Los proveedores y responsables del despliegue de sistemas de IA adoptarán medidas para garantizar que, en la mayor medida posible, su personal y demás personas que se encarguen en su nombre del funcionamiento y la utilización de sistemas de IA tengan un nivel suficiente de alfabetización en materia de IA, teniendo en cuenta sus conocimientos técnicos, su experiencia, su educación y su formación, así como el contexto previsto de uso de los sistemas de IA y las personas o los colectivos de personas en que se van a utilizar dichos sistemas".

70 Expone el Considerando 20 que "Con el fin de obtener los mayores beneficios de los sistemas de IA, protegiendo al mismo tiempo los derechos fundamentales, la salud y la seguridad, y de posibilitar el control democrático, la alfabetización en materia de IA debe dotar a los proveedores, responsables del despliegue y personas afectadas de los conceptos necesarios para tomar decisiones con conocimiento de causa en relación con los sistemas de IA" desarrolla posteriormente los conceptos y la actuación en el Considerando, conceptos que afectan a todos los agentes pertinentes de la cadena de valor de la IA.

Existe una conexión con la democratización, porque, aunque la IA está llegando rápidamente a todos los públicos, no se puede afirmar que se esté democratizando totalmente, ya que esos públicos no sólo han de poder usarla, sino que también deben poder entenderla y por eso aboga por una educación digital de conocimiento y de pensamiento crítico[71]. La democratización implica velar tanto por el acceso como por la educación, la ética y la transparencia. Hay que impedir que los intereses privados perjudiquen este proceso[72].

El riesgo de uso malicioso

La naturaleza humana puede dar a las nuevas tecnologías un uso reprobable, imprudente o malicioso. Por ello, las organizaciones que desarrollan aplicaciones de ella deben ser transparentes, éticas y responsables con esta potente herramienta y deben mantener una estrecha vigilancia para mitigar o compensar posibles efectos negativos[73].

El riesgo de la desinformación

La desinformación puede originarse en la IA generativa: por modelos mal entrenados que dan lugar a datos inexactos o sesgados[74], por el uso indebido de la misma y por ataques maliciosos

71 En parte del Considerando 56 LIA se expone que "El despliegue de sistemas de IA en el ámbito educativo es importante para fomentar una educación y formación digitales de alta calidad y para que todos los estudiantes y profesores puedan adquirir y compartir las capacidades y competencias digitales necesarias, incluidos la alfabetización mediática, y el pensamiento crítico, para participar activamente en la economía, la sociedad y los procesos democráticos." Se refieren también a la alfabetización los Considerandos 91, 165 y los arts.66 c) y 95.2 c) LIA.

72 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.14.

73 IBIDEM., pág.14.

74 GOLDMAN SACHS, "Principales...op.cit.

(deepfakes o fakes news) con información engañosa, lo que contribuiría a la difusión de noticias falsas y desinformación[75].

La toxicidad

Los modelos de IA generativa pueden propagar, al reflejar el lenguaje de la red, contenidos tóxicos (por ejemplo, blasfemias y contenido sexual explícito) [76].

El riesgo en relación a la seguridad

Para ANNEMANS [77]son posibles problemas y riesgos en relación a la seguridad de la IA generativa: el desbordamiento de datos, la fuga de la propiedad intelectual y la confidencialidad, el entrenamiento de datos, el almacenamiento de datos, el cumplimiento, los datos sintéticos, las fugas accidentales, el uso indebido de la IA y los ataques maliciosos. La IA generativa necesita una gran cantidad de datos (puede incorporar datos de cualquier tipo -incluidos con información confidencial y/o privada-), que se van incrementando para conseguir un mayor aprendizaje y mejora del modelo. En el entrenamiento de datos (entrenamiento de algoritmos), podrían desvelarse involuntariamente datos

75 El Reglamento (UE) 2024/1083 del Parlamento Europeo y del Consejo, de 11 de abril de 2024, por el que se establece un marco común para los servicios de medios de comunicación en el mercado interior y se modifica la Directiva 2010/13/UE (Reglamento Europeo sobre la Libertad de los Medios de Comunicación), DOUE, nº 1083, de 17 de abril de 2024 recoge la lucha contra la desinformación en los Considerandos 4,6,14,51,53 y 56 y el art.19.1.c). Por su parte en el Estudio, la política europea frente a los deepfakes de julio de 2021, se define la desinformación como "difusión consciente, normalmente encubierta, de información engañosa con el objetivo de perjudicar el debate público, los procesos democráticos, la economía abierta o la seguridad nacional" (pág.XIV) y reconoce modalidades de la misma (pág.23). MUÑOZ VELA, J.M, *Retos...* op.cit.págs.93 y ss., sobre el concepto de desinformación.

76 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, págs.10 y 11.

77 ANNEMANS,R.,"Seguridad de la IA generativa: 8 riesgos que debes conocer". GlobalSing by GMO, 4 de diciembre del 2023 https://www.globalsign.com/es/blog/8-riesgos-de-seguridad-de-la-inteligencia-artificial-generativa (recuperado el 10 de mayo del 2024).

confidenciales, afectando a privacidad. Datos, que pueden almacenarse por terceros, que en caso de ser confidenciales pueden usarse indebidamente o filtrase si no están suficientemente protegidos (por ejemplo con elementos de cifrado y controles de acceso). No se puede olvidar el necesario cumplimiento de las correspondientes normas de datos. Por otra parte, los datos sintéticos, a través de sus patrones o detalles podrían llevarnos a la identidad o características sensibles. Pueden darse fugas accidentales, por ejemplo, en los modelos con base en textos o imágenes que de forma involuntaria incluyan información de datos de entrenamiento que no deberían revelarse (información personal) o datos comerciales confidenciales. No podemos olvidar el gran valor de la información.

El riesgo para la privacidad y confidencialidad

Uno de los riesgos para la privacidad como reconoce el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, es que los modelos "memoricen" al por mayor un registro de datos específico y cuando se les consulta lo repliquen, sobre todo cuando es información sensible (por ejemplo datos médicos). Esto también afecta a los derechos de autor y confidencialidad: por ejemplo, un empleado de Samsung que de manera involuntaria filtra información sensible cuando pega el código fuente, confidencial y protegido por derechos de autor, en ChatGPT comprobando si hay errores y optimizando el código[78].

Sobre la aplicación de la normativa de datos del Reglamento General de Protección de Datos, recientemente se han dado unas orientaciones del SEPD en el Informe, "La IA generativa y el EUDPR. Primeras orientaciones del SEPD para garantizar el cumplimiento de la protección de datos al utilizar sistemas de IA", de 3

78 Pág.10 del Documento.

de junio de 2024. Señalan estas orientaciones (que se entienden sin perjuicio de la LIA[79]), entre otras cosas, a que el uso masivo de datos disminuya: "El uso de grandes cantidades de datos para entrenar un sistema de IA generativa no implica necesariamente una mayor eficacia o mejores resultados. El diseño cuidadoso de conjuntos de datos bien estructurados, que se utilicen en sistemas que prioricen la calidad sobre la cantidad, siguiendo un proceso de entrenamiento debidamente supervisado y sometido a un seguimiento periódico, es esencial para lograr los resultados esperados, no solo en términos de minimización de datos, sino también cuando se trata de la calidad del resultado y la seguridad de los datos"[80]. También se reconoce la dificultad de erradicar los datos inexactos[81].

79 Son algunas de sus conclusiones que, "La supervisión periódica y la aplicación de controles en todas las etapas pueden ayudar a verificar que no hay tratamiento de datos personales, en los casos en que el modelo no está destinado a ello" (pág.8). "Es responsabilidad de la EUI gestionar adecuadamente los riesgos relacionados con el uso de sistemas de IA generativa. Los riesgos para la protección de datos deben identificarse y abordarse a lo largo de todo el ciclo de vida del sistema de IA generativa. Esto incluye una supervisión periódica y sistemática para determinar a medida que evoluciona el sistema, si los riesgos ya identificados están empeorando o si están apareciendo nuevos riesgos. La comprensión de los riesgos relacionados con el uso de IA generativa aún está en curso, por lo que es necesario mantener un enfoque vigilante con respecto a riesgos emergentes no identificados. Si se identifican riesgos que no pueden mitigarse por medios razonables, es el momento de consultar al SEPD" (págs.12 y ss.)." Las instituciones de la UE deben proporcionar a las personas físicas toda la información exigida en el Reglamento cuando utilicen sistemas de IA generativa que traten datos personales. La información facilitada a las personas deberá actualizarse cuando sea necesario para mantenerlas debidamente informadas y en control de sus propios datos" (pág.25)." Cuando se prevean sistemas de IA generativa para apoyar los procedimientos de toma de decisiones, las instituciones de la UE deberán considerar detenidamente la posibilidad de ponerlos en funcionamiento si su uso plantea dudas sobre su legalidad o su potencial de constituir decisiones injustas, poco éticas o discriminatorias" (pág.27)." La falta de información sobre los riesgos de seguridad ligados al uso de sistemas de IA generativa y su posible evolución obliga a las IUE a extremar la precaución y a realizar una planificación detallada de todos los aspectos relacionados con la seguridad informática, incluyendo la monitorización continua y el soporte técnico especializado. Las IUE deben ser conscientes de los riesgos derivados de los ataques de terceros malintencionados y de las herramientas disponibles para mitigarlos" (pág.35).

80 Pág.20, añade que "a medida que las tecnologías de IA avanzan rápidamente, las instituciones de la UE deben considerar cuidadosamente cuándo y cómo utilizar la IA generativa de manera responsable y beneficiosa para el bien público. Todas las etapas del ciclo de vida de una solución de IA generativa deben operar de conformidad con los marcos jurídicos aplicables, incluido el Reglamento, cuando el sistema implique el tratamiento de datos personales "(pág.7).

81 Así afirma el Informe del SEPD "La IA generativa y el EUDPR. Primeras orientaciones del SEPD para garantizar el cumplimiento de la protección de datos al utilizar sistemas de

El riesgo para la ciberseguridad

Para MUÑOZ VELA la seguridad lógica o informática tiene impacto en el mundo físico para personas, instalaciones (especialmente alto en las infraestructuras críticas y servicio esenciales), empresas y Gobiernos. Señala el autor el crecimiento exponencial de los ciberataques cuantitativa y cualitativamente, contra la reputación, la democracia, la influencia política y el ciberterrorismo[82]. Destacando en el ámbito de la IA generativa la posible aplicación de las técnicas de Deep Fakes para realizar el ciberataque del spear phishing o phishing selectivo y concretamente el "whaling" o "fraude del CEO"[83]. Ciberataques con co-

IA", de 3 de junio de 2024 que "A pesar de los esfuerzos por garantizar la exactitud de los datos, los sistemas generativos de IA siguen siendo propensos a resultados inexactos que pueden repercutir en los derechos y libertades fundamentales de las personas. Aunque los proveedores están implantando sistemas de información avanzados para garantizar que los modelos utilicen y generen datos precisos, las IUE deben evaluar cuidadosamente la precisión de los datos a lo largo de todo el ciclo de vida de los sistemas de IA generativa y plantearse el uso de dichos sistemas si no se puede mantener la precisión" (pág.22). Véase por otra parte, "El Comisionado de Hamburgo para la protección de Datos y la Libertad de Información pública en documento en el que se analiza los grandes Modelos de Lenguaje desde el punto de vista de la protección de datos, Boletín de julio del 2024", Lks, noticias, (https://www.lksnext.com/es/noticias_boletin/el-comisionado-de-hamburgo-para-la-proteccion-de-datos-y-la-libertad-de-informacion-publica-un-documento-en-el-que-analiza-los-grandes-modelos-de-lenguaje-desde-el-punto-de-vista-de-la-proteccion-de-d/) (rescuperado el 2 de agosto del 2024).

82 MUÑOZ VELA, J.M, *Retos*...op.cit.pág152 especifica, ataques distribuidos de denegación de servicios (DDos), mediante redes botnet (Como "Mirai atack"), ataques de ransomware, ataques incesantes de ciberespionaje mediante virus, especialmente para acceder y sustraer información confidencial , como "Moonlight Maza", "Titan Rain", "Duq"," Flame", "Red October" o "Gauss", entre muchos otros (ej. Cisco informó en 2018 que se bloquearon 7 billones de amenazas en nombre de sus clientes").

83 IBIDEM., págs.152 y ss., señala el autor que está este ciberataque dirigido a altos cargos para obtener información confidencial de una organización o dinero y, el "spear phishing basado en inteligencia artificial". Como expone el autor los delincuentes, que incluso forman parte de organizaciones criminales (Crimen as a Service), se benefician de estos sistemas para rastrear las redes (datos rastreables que se comparten por internet: dirección electrónica, imágenes y la voz) y encontrar información útil sobre la persona a suplantar, analizarla e imitar su lenguaje, su estilo de comunicación o su timbre de voz (por ejemplo dirigen una orden a un subordinado de una entidad, suplantando a la persona y falsificando las comunicaciones y su contenido). Por ejemplo, una empleada de una firma que tiene su sede en Hong Kong (en algunos medios hablan de empleado), transfirió 25 millones de dólares a unos estafadores tras recibir esa instrucción por su director financiero en una videollamada con otros profesionales. Realmente no era una llamada real sino una

rreos electrónicos de phishing y código malicioso que incluso poniendo filtros para impedirlos se los pueden saltar[84].

El uso de la IA como instrumento para la comisión de actos ilícitos o delictivos, seguirá creciendo, estudiando el comportamiento del sistema y humano ("la parte débil"). Es necesario por ello, reforzar la seguridad de la IA generativa y controlarla ya que puede ser instrumento para el ciberataque (dada la información que guarda)[85] y al mismo tiempo la IA se convierte en instrumento al servicio de la ciberseguridad[86]. Pudiendo desempeñar en materia de ciberseguridad un papel destacado[87].

réplica creada por los estafadores (LALCHAND, S. et al. (Centro de Servicios Financieros de Deloitte), "Se espera que la IA generativa aumente el riesgo de deepfakes y otros fraudes de la banca", Deloitte, Servicios Financieros, 29 de mayo del 2024, https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html)) (recuperado el 2 de agosto del 2024).

84 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, pág.16.

85 Es necesario actualizar la legislación (SÁNCHEZ, L., "Escrivá advierte que la estrategia de IA necesita de entornos seguros y anuncia una nueva ley integral de ciberseguridad", Economist & Jurist, 7 de junio del 2024 (recuperado en 20 de julio del 2024)).

86 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.149, afirma que "La IA puede servir para garantizar o mejorar la seguridad, especialmente mediante medidas y controles preventivos , de efectivos y reactivos de gestión, contención y mitigación y para atentar con la misma con absoluta precisión, personalización y efectividad.".

87 "La inteligencia artificial generativa puede desempeñar un papel valioso en el campo de la ciberseguridad de varias maneras interesantes: 1. Simulación de amenazas: La IA generativa puede crear escenarios de ciberataques simulados que ayudan a entrenar a los sistemas de defensa cibernética y a los profesionales en la identificación y mitigación de riesgos. Esto permite a las organizaciones prepararse mejor para ataques reales y entender mejor las tácticas que los atacantes podrían utilizar. 2. Generación de Datos de Prueba:Puede generar grandes cantidades de datos de red realistas pero ficticios que pueden ser usados para entrenar modelos de detección de intrusos sin comprometer datos reales sensibles. Esto es especialmente útil para empresas que necesitan cumplir con regulaciones estrictas de privacidad.3.Análisis de Comportamiento Anómalo: La IA generativa puede ayudar a modelar lo que se considera "normal" en un red y luego generar comportamientos que se desvíen de esta norma, lo que es crucial para sistemas de detección de anomalías.4.Fortalecimiento de la Autenticación: En la autenticación, técnicas generativas pueden ser utilizadas para crear métodos de verificación biométrica más complejos y seguros, como la generación de voz o imagen que ayuden a mejorar los sistemas de reconocimiento facial o de voz.5.Mejora de los Sistemas de Respuesta e Incidentes: Al integrar IA generativa, los sistemas de respuesta a incidentes pueden simular diversas estrategias de respuesta a ataques, permitiendo a las organizaciones evaluar y optimizar sus tácticas y protocolos de respuesta antes de que un

Riesgos psicológicos

En el Considerando 5 de la LIA se alude expresamente a los riesgos psicológicos, afirma que "dependiendo de las circunstancias relativas a su aplicación, utilización y nivel de desarrollo tecnológico concretos, la IA puede generar riesgos y menoscabar los intereses públicos y los derechos fundamentales que protege el Derecho de la Unión. Dicho menoscabo puede ser tangible o intangible e incluye los perjuicios físicos, psíquicos, sociales o económicos". La IA generativa puede tener un impacto psicológico y emocional. Por ejemplo, el posible sufrimiento, malestar de familiares y allegados, al escuchar una voz clonada y ver la reproducción de la imagen de una persona fallecida o desaparecida[88]. Impacto que podría generar daños psicológicos tanto si se usa de forma normal (es algo impactante) como de manera imprudente o manipulada. Como afirma FRANGANILLO[89] toda manipulación tiene un impacto psicológico, por eso son peligrosos los deefakes y las alucinaciones o fallos. No sabemos el alcance de las consecuencias. Se ha llegado a producir un suicidio inducido por inteligencia artificial, un chatbot ("Eliza") basado

incidente real ocurra" RESPUESTA de Chat GPT a prompt: Puede la IA generativa servir de ciberseguridad, que le planteé el 10 de mayo del 2024 a las 15,26. Por su parte, Microsoft Security establece medidas de protección frente a las amenazas de ciberseguridad con IA generativa en su guía: IA generativa, la ventaja de los defensores, dirigida a los directivos de seguridad de la información (CISO) y a los profesionales en ciberseguridad que buscan colaborar con directivos de toda su organización y asegurarse de que la empresa aprovecha la IA generativa, LinkedIn, Microsoft (recuperado en 6 de mayo de 2024).

88 En el artículo "Las empresas de inteligencia artificial ofrecen ya el servicio de recrear a seres queridos fallecidos e interactuar con ellos",20 minutos,20 bits, 3 de diciembre del 2023,(https://www.20minutos.es/tecnologia/empresas-inteligencia-artificial-ofrecen-servicio-recrear-seres-queridos-fallecidos-interactuar-con-ellos-5195903/) (recuperado el 1 de junio del 2024) se muestra como se ofrece este servicio de recrear seres queridos fallecidos e interactuar con ellos, sus beneficios y problemas (entre ellos, que se quiera maximizar las ganancias por parte de la empresa con ofertas "adictivas" como cobrar cada vez que lo utilicen en vez de usar tarifa fija) y el debate ético que se plantea. Muestra como ya hay gente preparando su avatar posterior, preparan su "yo" virtual para después de la muerte. Véase MUÑOZ VELA, J.M, *Retos*...op.cit.pág.76.

89 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.14.

en tecnología GPT-J[90]. El Estudio, la política europea frente a los deepfakes, de julio del 2021 reconoce entre los daños psicológicos: la extorsión, la sextorsión, la difamación, la intimidación, el acoso escolar y socavar la confianza[91]. Otro ejemplo, sería como dice MUÑOZ VELA[92], el previsible fuerte impacto psicológico del desempleo temporal, y sobre todo, la exclusión permanentemente del mundo laboral, ante el cambio que está provocando y provocará la IA en el mercado laboral.

Riesgos por errores y alucinaciones

Los errores y alucinaciones[93] (cuando son vividos y adoptan una antropomorfización) son riesgos de la IA generativa. Chat GPT, comete errores, tiene mayores dificultades para tareas como la lógica, las matemáticas y el sentido común. Crea respuestas erróneas muy convincentes, "fiables", por ejemplo, a preguntas médicas. Ha creado falsas historias de acoso sexual y código de software susceptible de vulnerabilidades. No perdiendo de vista que cualquier vulnerabilidad de un modelo base corre el riesgo de heredarse en los modelos derivados de él.[94]

90 SOTO ARAMENDARIZ, S "Primer suicidio inducido por inteligencia artificial: algo que temer, 4 de abril de 2023 (https://observatorioblockchain.com/ia/primer-suicidio-inducido-por-inteligencia-artificial-algo-que-temer) (recuperado el 28 de mayo del 2024), aunque la IA también ayuda a detectar los posibles intentos de suicidio y evitarlos (PUFPAFF, M. ¿ La tecnología como fuerza para el bien? ¿Cómo se está utilizando la inteligencia artificial para prevenir los suicidios en China?, Razón y fe, Tomo 282, nº1447, 2020, págs.205 y ss.).

91 Págs.30., reconoce también el Informe daños financieros y sociales (págs.30 y ss).

92 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.44.

93 En el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023 se considera más correcto el uso del término "confabulación" o pastiche (pág.9). Considera preocupantes estos problemas en modelos de cimentación por su diseño para un uso amplio y general.

94 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023 , pág.9.

Riesgos en relación a las cuestiones éticas. Códigos de conducta y alienación de valores

La IA generativa puede mejorar la productividad y ofrecer beneficios, pero como se ha visto, implica riesgos éticos y sociales (¿Hasta qué punto es ético "revivir" a las personas fallecidas?, la manipulación del público generando confusión y desinformación y el contenido sesgado o engañoso, con los deepfakes, por ejemplo) que afectan a aspectos esenciales como el empleo, la democracia, la cultura, el sistema económico, la privacidad, la propiedad intelectual, la intimidad, la no discriminación (creada o acentuada por la IA), etc. Los contenidos son tan realistas que pueden confundir socialmente, por ello como establece FRANGANILLO[95], es exigible un código de conducta, una concienciación sobre un uso ético y responsable, para el bien común, en iniciativas con impacto social, como el proyecto OpenAI, el observatorio OdiseaIA o el programa AI for Social Good, de Google. Un uso transparente y responsable de la tecnología que respete los valores fundamentales de la sociedad.

La seguridad de la IA generativa se asocia a menudo, según el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023[96] con el concepto de alineación de valores (alineados con los valores y objetivos humanos para evitar que hagan daño a sus creadores humanos). Se formulan instrucciones para que la IA generativa alcance determinados "objetivos" que pueden estar mal especificados o representados. Explica el Informe que una función objetivo para los asistentes de IA debe ser que sea "útil" o "inofensivo", conceptos difíciles de definir y especificar y de determinar su compensación. Pone el ejemplo, de que la insistencia en evitar el "daño" puede llevar a respuestas "seguras" que quizás no sean

95 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.11

96 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, pág.19

valiosas para el usuario, pero, por otra parte, dar más importancia a ser útil puede hacer que el sistema ofrezca respuestas tóxicas que causen daños. Reconoce que se puede mitigar este problema, basándose en el Aprendizaje por Refuerzo a través de la Retroalimentación Humana (RLHF).

III.- RIESGOS DE LA IA GENERATIVA QUE HAY QUE TENER EN CUENTA EN RELACIÓN A LA LIA

Como expone JARQUES[97], la idea principal de la LIA es regular la IA en función de su capacidad de causar daño a la sociedad con un enfoque "basado en el riesgo", a mayor riesgo más estrictas son las reglas. La normativa establece obligaciones para la IA en función de los riesgos potenciales y su nivel de impacto, dividiéndose los sistemas de IA en: Riesgo inaceptable (prohibidos), alto (requiere estrictas medidas), limitado y mínimo [98].

97 JARQUES, A., "El futuro Reglamento Europeo de Inteligencia Artificial", *Actualidad Jurídica Aranzadi,* nº1003, 2023, Editorial Aranzadi, (BIB 2024/278) (rescatada en 27 de mayo del 2024). Afirma que los dos últimos tienen regulaciones y medidas menos estrictas, sus obligaciones de transparencia son más leves (señala la autora la obligación de divulgar que el contenido se generó mediante IA).

98 La idea podría seguir siendo válida tras la redacción definitiva del Reglamento, ya que continúa la IA prohibida y la de alto riesgo. Se puede considerar con menor riesgo ("limitado") a la IA de riesgos sistémicos (siempre que no esté incluída en la prohibida y la de alto riesgo), y también la IA presenta en determinadas ocasiones un riesgo mínimo fuera de los tres casos mencionados. En esta línea establece el Considerando 26 LIA que"con el fin de establecer un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA, es preciso aplicar un enfoque basado en los riesgos claramente definido, que adapte el tipo y contenido de las normas a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA de que se trate. Por consiguiente, es necesario prohibir determinadas prácticas de IA que no son aceptables, definir los requisitos que deben cumplir los sistemas de IA de alto riesgo y las obligaciones aplicables a los operadores pertinentes, así como imponer obligaciones de transparencia a determinados sistemas de IA".

III.1.- LA IA GENERATIVA COMO IA PROHIBIDA

Un punto de partida en común, para toda clase de sistema de IA, es el establecimiento de prácticas prohibidas, aplicables a la IA de uso general, y por tanto a los grandes modelos de IA generativa. Se están de esta forma limitando a priori posibles riesgos inaceptables del uso de una IA generativa, evitándolos de forma taxativa.

En el artículo 5 LIA se recogen las prácticas de IA que quedan prohibidas y también las excepciones a las mismas. Se prohíbe la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA considerado prohibido por dicha norma.

Dentro de estas prohibiciones, se observa, que algunas pueden afectar directamente a la IA generativa y otras, entiendo, podrían hacerlo, solo si se usan para el entrenamiento de otras tecnologías prohibidas, o en apoyo a las mismas.

En primer lugar, podrían considerarse como prácticas prohibidas que afectan directamente a la IA generativa, principalmente:

1º- La "que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas" (art.5.1 a) LIA). En el artículo inicialmente, la necesidad de alteración sustancial del comportamiento parece exigible a ambos supuestos, pero podría interpretarse que el primer supuesto "que se sirva de técnicas subliminales que transciendan la conciencia de una persona" debe ser un supuesto de prohibición

en sí mismo (después de engañosas no lleva coma lo que podríamos interpretar que el resto de matiz del artículo se refiere al segundo supuesto: técnicas deliberadamente manipuladoras o engañosas) sin necesidad de alterar sustancialmente el comportamiento a la hora de tomar la decisión. PLAZA PENADÉS[99] ya consideró y comparto, que realmente se debe prohibir toda IA que se sirva de técnicas subliminales que transciendan la conciencia de una persona, independientemente de que sea determinante o no en la decisión, por "el principio de que "los derechos fundamentales están por encima de la tecnología" y "porque los sistemas IA que se sirvan de técnicas subliminales que trasciendan la conciencia de una persona suponen una intromisión ilegítima al honor (en sentido inmanente, la concepción que una persona tiene de sí misma) y a su más estricta y pura intimidad". El autor manifiesta que esa primera prohibición podía haber sido un "neuroderecho" y redactarse así:"Quedan prohibidos los sistemas IA que tengan como finalidad acceder y trascender la conciencia de una persona, aunque sea simplemente para acceder a dicha conciencia. Revestirá especial gravedad si además pretenden alterar o alteran la voluntad cognitiva de las personas, salvo que sea por razones médicas y se realice con la supervisión de un médico especialista responsable".

Es evidente que a través de las imágenes, audios, videos y textos sintéticos creados por la IA generativa, como se expuso, se puede de forma subliminal transcender la conciencia

99 PLAZA PENADÉS, J., Dossier, "Las claves de la futura Ley de Inteligencia Artificial Europea", Aranzadi La Ley, Navarra, mayo, 2023, pág.15. Para el autor debe prohibirse la IA que afecte a los derechos fundamentales básicos de las personas: "Por tanto, podemos colegir que la principal prohibición es que un sistema IA infrinja o no garantice la observancia o cumplimiento de los derechos fundamentales básicos de las personas, lo que formulo siempre bajo el principio "los derechos fundamentales prevalecen sobre la tecnología", que debemos de aplicar a todo el desarrollo de Internet en lo que se refiere al necesario respeto de los derechos de honor, intimidad, propia imagen, protección de datos…. Principio que obviamente también resulta extensible y aplicable a todos los sistemas de Inteligencia Artificial" (pág.14).

de una persona, o deliberada, manipular y engañar con el objetivo o el efecto de alterar de manera sustancial (es determinante en su actuación) el comportamiento de una persona o colectivo de personas que impide tomar una decisión informada y le lleva a tomar otra que provoca o puede provocar perjuicios considerables a la persona o colectivo de personas que la adoptan. En el momento que das una información, imagen o vídeo falsos o manipulados (deepfakes, fake news,deep voice), cuando no se puede distinguir fácilmente la realidad de la ficción, el uso en este sentido está prohibido.

2º- La "que explote alguna de las vulnerabilidades de una persona física o un determinado colectivo de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona o a otra" (art.5.1 b) LIA).

Los contenidos creados por la IA generativa podrían utilizarse para explotar vulnerabilidades de una persona o colectivo específico de personas (por edad, discapacidad, situación social o económica específica), a través de textos, vídeos o imágenes falseadas dirigidas a ella o a un colectivo, haciéndole creer lo que no es, pero estas personas no son capaces de discernir, creando confusión, desinformación y manipulación. Todo ello, con el objetivo o efecto de alterar sustancialmente (es determinante en su actuación) su comportamiento esa persona u otra que pertenece al colectivo, provocando o siendo razonable que provoque perjuicios considerables a las mismas.

En estas prohibiciones como se ha visto, la ética está en la esencia de las mismas: la protección de las personas vulnerables y la

no manipulación del comportamiento de las personas que deben tomar decisiones libre y conscientemente[100].

En relación al resto de supuestos de prácticas prohibidas (con el fin de evaluar o clasificar a personas físicas o a grupos de personas, para realizar evaluaciones de riesgo de personas, base de reconocimiento facial, inferir emociones e IA de categorización biométrica), normalmente, el Sistema de IA principal no es una IA generativa. Por ejemplo, en materia de reconocimiento facial y de inferir emociones, habitualmente la base son otros sistemas, pero se puede utilizar la IA generativa para el entrenamiento de estos generando imágenes con rostros, o con rostros que tengan diferentes emociones. Podrían así ser útiles para crear datos sintéticos con los que entrenar estos sistemas de reconocimiento facial y de emociones, especialmente en situaciones donde los datos reales son limitados o cuando se necesita preservar la privacidad. En resumen, mientras que la IA generativa puede apoyar indirectamente las tareas de reconocimiento de emociones a través de la generación de datos para el entrenamiento, el trabajo directo de analizar y clasificar emociones se realiza mediante técnicas de IA no generativas especializadas en reconocimiento de patrones y aprendizaje automático. Entiendo que, si estas están prohibidas, también lo está el uso de una inteligencia generativa a su servicio.

La Comisión establecerá directrices sobre la aplicación del Reglamento sobre las prácticas prohibidas a que se refiere el art. 5 (art.96.1 b) LIA). Lo establecido en el art.5 LIA no afectará a las prohibiciones aplicables cuando una práctica de IA infrinja otras disposiciones de Derecho de la Unión (art.5.8 LIA).

La posible inclusión directa de los deepfakes como IA prohibida o, como se verá IA de alto riesgo, aparecía ya en el Estudio, la política europea frente a los deepfakes, de julio del 2021 que planteaba: "Prohibir determinadas aplicaciones: Considerando el

100 En el Considerando 29 de la LIA se recogen expresamente estas técnicas de manipulación.

potencial impacto negativo de aplicaciones específicas de deepfakes (p. ej., la pornografía deepfake no consentida), las obligaciones de transparencia por sí solas parecen insuficientes para hacer frente a esos efectos negativos. Por lo tanto, la Comisión proponía como opción prohibir determinados tipos de aplicaciones y usos concretos de esta tecnología"[101].

III.2.- LA IA GENERATIVA COMO IA GENERAL CON RIESGOS SISTÉMICOS

La IA generativa se incluye en los modelos de IA de uso general, modelos que en determinados casos pueden clasificarse como modelo de IA de uso general con riesgo sistémico. Aparece así uno de los riesgos reconocidos en la LIA el riesgo sistémico, definido en su artículo 3 número 65, como "un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general (aquellas que igualan o superan las capacidades mostradas por los modelos de IA de uso general más avanzados-definición 64 art.3 LIA-), que tienen unas repercusiones considerables en el mercado de la Unión debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor".

Dicho riesgo se da si el modelo de IA de uso general reúne alguna de las siguientes condiciones (art.51 .1 LIA) :a) tiene capacidades de gran impacto evaluadas a partir de herramientas y metodologías técnicas adecuadas, como indicadores y parámetros de referencia; b) con arreglo a una decisión de la Comisión, adoptada de oficio a raíz de una alerta cualificada del grupo de expertos científicos, tiene capacidades o un impacto equivalente a los establecidos en la letra a), teniendo en cuenta los criterios establecidos en el anexo XIII. En dicho Anexo XIII aparecen

101 Pág..61. ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA...op.cit.

criterios de carácter técnico, objetivo (para la clasificación de los modelos de IA de uso general con riesgo sistémico a que se refiere el art. 51[102]).

Además de la presencia de posibles capacidades de gran impacto, es indudable que la IA generativa puede tener efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto[103]. Esto ha quedado

102 Señala el art.51.2 y 3 LIA respectivamente que "se presumirá que un modelo de IA de uso general tiene capacidades de gran impacto con arreglo al apartado 1, letra a), cuando la cantidad acumulada de cálculo utilizada para su entrenamiento, medida en operaciones de coma flotante, sea superior a 10 25" y que "La Comisión adoptará actos delegados de conformidad con el artículo 97 para modificar los umbrales a que se refieren los apartados 1 y 2 del presente artículo, así como para complementar los parámetros de referencia e indicadores en función de los avances tecnológicos, como las mejoras algorítmicas o la mayor eficiencia del hardware, cuando sea necesario, para que los umbrales reflejen el estado actual de la técnica". Añade su art.52.6 LIA que "La Comisión velará por que se publique una lista de modelos de IA de uso general con riesgo sistémico, que mantendrá actualizada, sin perjuicio de la necesidad de respetar y proteger los derechos de propiedad intelectual e industrial y la información empresarial confidencial o los secretos comerciales de conformidad con el Derecho de la Unión y nacional".

103 Es esencial en este punto el Considerando 110 de LIA que recoge de forma detallada dichos riesgos "Los modelos de IA de uso general pueden plantear riesgos sistémicos, por ejemplo, cualquier efecto negativo real o razonablemente previsible en relación con accidentes graves, perturbaciones de sectores críticos y consecuencias graves para la salud y la seguridad públicas, cualquier efecto negativo real o razonablemente previsible sobre los procesos democráticos y la seguridad pública y económica o la difusión de contenidos ilícitos, falsos o discriminatorios. Debe entenderse que los riesgos sistémicos aumentan con las capacidades y el alcance de los modelos, pueden surgir durante todo el ciclo de vida del modelo y se ven influidos por las condiciones de uso indebido, la fiabilidad del modelo, la equidad y la seguridad del modelo, el nivel de autonomía del modelo, su acceso a herramientas, modalidades novedosas o combinadas, las estrategias de divulgación y distribución, la posibilidad de eliminar las salvaguardias y otros factores. En particular, los enfoques internacionales han establecido hasta la fecha la necesidad de prestar atención a los riesgos derivados de posibles usos indebidos intencionados o de problemas en materia de control relacionados con la armonización con la intención humana no deseados, a los riesgos químicos, biológicos, radiológicos y nucleares, como las maneras en que las barreras a la entrada pueden reducirse, también para el desarrollo, el diseño, la adquisición o el uso de armas, a las cibercapacidades ofensivas, como las maneras en que pueden propiciarse el descubrimiento, la explotación o el uso operativo de vulnerabilidades, a los efectos de la interacción y el uso de herramientas, incluida, por ejemplo, la capacidad de controlar sistemas físicos e interferir en el funcionamiento de infraestructuras críticas, a los riesgos derivados del hecho que los modelos hagan copias de sí mismos o se «autorrepliquen» o entrenen a otros modelos, a las maneras en que los modelos pueden dar lugar a sesgos dañinos y discriminación que entrañan riesgos para las personas, las comunidades o las sociedades, a la

expuesto anteriormente al analizar ampliamente sus riesgos los cuales afectan o pueden afectar a todas las esferas antes mencionadas. Estos riesgos pueden propagarse a gran escala a lo largo de toda la cadena de valor. También queda reflejada esa posibilidad en el Considerando 136 de la LIA al hablar de las obligaciones impuestas a los proveedores y a los responsables del despliegue de determinados sistemas de IA (detectar, divulgar que el resultado es artificial y etiquetarlo), se refiere expresamente a los riesgos sistémicos que pueden surgir de la divulgación de contenidos generados o manipulados de manera artificial.

III.3.- LA IA GENERATIVA COMO IA DE ALTO RIESGO

Una IA generativa puede ser o convertirse en una IA de alto riesgo sujeta a la más estricta regulación a la que está sometida dicha IA en la LIA.

Los sistemas de IA de alto riesgo[104] (art.6.1 a) y b) LIA) son los que están destinados a utilizarse como componentes de seguridad de un producto (que entre en los ámbitos de aplicación de los actos legislativos de armonización de la Unión de su anexo I) o son uno de esos productos sujetos ambos a una evaluación de la conformidad de terceros para su introducción en el mercado

facilitación de la desinformación o el menoscabo de la intimidad, que suponen una amenaza para los valores democráticos y los derechos humanos, al riesgo de que un acontecimiento concreto dé lugar a una reacción en cadena con efectos negativos considerables que podrían afectar incluso a una ciudad entera, un ámbito de actividad entero o una comunidad entera". Posteriormente, en el Considerando 111 de la LIA se muestra el establecimiento de una metodología para la clasificación de los modelos de IA de uso general como modelos de IA de uso general con riesgos sistémicos y en el 112, el procedimiento para la clasificación de un modelo de IA de uso general con riesgos sistémicos.

104 En La LIA estos sistemas se regulan en el capítulo III: en la sección 1 (arts.6.-7) se dispone su clasificación. En la sección 2 se recogen los requisitos de los Sistemas de las IA de alto riesgo (arts.8-15 LIA). En la sección 3 se recogen las obligaciones de los proveedores y responsables del despliegue de sistemas de IA de alto riesgo y otras partes (arts.16-27 LIA). En la sección 4, recoge las autoridades notificantes y los organismos notificados (arts.28-39) En la sección 5, Normas, evaluación de la conformidad, certificados, registros (arts.40-49 LIA).

o puesta en servicio, con arreglo a los mencionados actos legislativos de armonización de la Unión Europea, enumerados en el anexo I. También son sistemas de IA de alto riesgo, los sistemas que formen parte de cualquiera de los ámbitos recogidos en el Anexo III (art.6.2 LIA) pero deben plantear un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, e influir sustancialmente en el resultado de la toma de las decisiones (a sensu contrario del art.6. 3 LIA que no considera a una IA de alto riesgo cuando no plantee un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, también al no influir sustancialmente en el resultado de la toma de decisiones y cumpla cualquiera de las condiciones que establece el artículo, aunque siempre se considerarán de alto riesgo cuando el sistema de la IA elabore perfiles de personas físicas).

Las materias recogidas en el Anexo III son: biometría, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable (sistemas de identificación y categorización biométrica y reconocimiento de emociones en los términos establecidos); infraestructuras críticas: sistemas de IA destinados a ser utilizados como componentes de seguridad ; educación y formación profesional; empleo, gestión de los trabajadores y acceso al autoempleo; acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones; garantía del cumplimiento del Derecho, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable; migración, asilo y gestión del control fronterizo, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable y administración de justicia y procesos democráticos[105].

105 Sintetiza PLAZA PENADÉS, J., Dossier, "Las claves de...op. cit. págs.35 y 36: "Las que se utilicen para: -La identificación biométrica de personas y extraer conclusiones sobre sus características personales. Sin incluir los de verificación biométrica, como la autenticación. No son de alto riesgo los utilizados en ciberseguridad y para la protección

Por otra parte, el propio texto de la LIA nos detalla esos derechos fundamentales que no pueden sufrir un riesgo importante. Entre los derechos fundamentales protegidos por la Carta de los Derechos Fundamentales de la Unión Europea, a los efectos de clasificar un sistema como de alto riesgo, el Considerando 48 LIA incluye: el derecho a la dignidad humana, el respeto de la vida privada y familiar, 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 no discriminación, el derecho a la educación, la protección de los consumidores, los derechos de los trabajadores, los derechos de las personas discapacitadas, la igualdad entre hombres y mujeres, los derechos de propiedad intelectual, el derecho a la tutela judicial efectiva y a un juez imparcial, los derechos de la defensa y la presunción de inocencia, el derecho a una buena administración y los derechos específicos

---

de datos personales.-La educación o la formación profesional: determinan el acceso o admisión, distribución o evaluación de las personas a partir de pruebas realizadas.- El empleo, la gestión de los trabajadores y el acceso al autoempleo, en decisiones relativas a la iniciación, la promoción y la rescisión de contratos y la asignación personalizada de tareas y evaluación de personas en relaciones laborales.- Evaluar la calificación crediticia o solvencia de personas físicas, ya que deciden sobre el acceso a recursos financieros o servicios esenciales (vivienda, electricidad y servicios de telecomunicaciones). No son de alto riesgo los previstos para detectar fraudes en la oferta de servicios financieros.-Decidir si las autoridades deben denegar, reducir, revocar o reclamar ayudas y servicios.- Tomar decisiones o influir en la elegibilidad de las personas físicas para acogerse al seguro de enfermedad y de vida.- Evaluar y clasificar llamadas de emergencia de personas físicas o el envío o establecimiento de prioridades en el envío de servicios de primera intervención- La toma de decisiones de las autoridades policiales y judiciales en la prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de las sanciones.- Evaluar la fiabilidad de las pruebas en un proceso penal, elaborar perfiles, investigación o enjuiciamiento de infracciones penales o de ejecución de las sanciones.-Gestionar la migración, el asilo y el control fronterizo, ayudar a examinar y evaluar la veracidad de las pruebas, el seguimiento, la vigilancia o el tratamiento de datos personales en gestión de fronteras.- La administración de justicia y los procesos democráticos para ayudar a las autoridades judiciales u órganos administrativos a investigar e interpretar los hechos y el Derecho y a aplicar la ley. La utilización de estas herramientas no debe sustituir la toma de decisiones de los jueces ni su independencia, puesto que debe seguir siendo una actividad y una decisión humana; salvo, actividades administrativas accesorias (anonimización o seudonimización documentos o datos; comunicación entre personal; tareas administrativas, o la asignación de recursos).-Influir en el resultado de una elección o un referendo o en el comportamiento electoral en ejercicio del derecho a voto, a excepción de los sistemas para organizar, optimizar y estructurar campañas políticas desde el punto de vista administrativo y logístico".

de los menores. Añade que al evaluar la gravedad del perjuicio que puede ocasionar un sistema de IA, también en la salud y la seguridad de las personas, se debe tener en cuenta, además, el derecho fundamental a un nivel elevado de protección del medio ambiente[106].

La Comisión, previa consulta al Consejo Europeo de Inteligencia Artificial, dará directrices de aplicación práctica, en consonancia con el art.96 LIA y una lista exhaustiva de ejemplos prácticos de casos de uso de sistemas de IA que sean de alto riesgo y que no sean de alto riesgo (arts 6.5 LIA[107]). Hay que estar pues a la espera de dichas directrices y de la lista exhaustiva que dará luz a esta problemática. La lista de posibles IA de alto riesgo está abierta a futuras incorporaciones. La Comisión también puede adoptar actos delegados al objeto de aumentar o modificar la lista del Anexo III (art.97 en relación con el art.7LIA).

En el Anexo III LIA en su punto 8, en materia de administración de justicia y procesos democráticos, se recoge expresamente entre los sistemas considerados de alto riesgo: "b) Sistemas de IA destinados a ser utilizados para influir en el resultado de una elección o referéndum o en el comportamiento electoral de personas físicas que ejerzan su derecho de voto en elecciones o referendos. Quedan excluidos los sistemas de IA a cuyos resultados de salida no estén directamente expuestas las personas físicas, como las herramientas utilizadas para organizar, optimizar o estructurar campañas políticas desde un punto de vista

106 PUERTO MENDOZA, A., Derecho *digital*. *Fundamentos básicos*, Ediciones CEF.,2019 (págs.179 y ss.,200 y ss.) recoge junto al derecho al honor, la intimidad familiar y la propia imagen (art.18 CE y 12 de la Declaración universal de los Derechos Humanos (1948)) , el derecho al nombre (se menciona el anonimato y el problema de la suplantación y usurpación de la personalidad), el derecho a la libertad de expresión y de información (art.20 CE) y otros derechos fundamentales e instrumentales de los anteriores: el derecho a la inviolabilidad del domicilio, al secreto de las comunicaciones y a la autodeterminación informativa o derecho de protección de datos personales. Muestra un derecho constitucional de nueva generación: el derecho a la protección del propio entorno digital (págs.205 y ss.). Sobre los derechos digitales en general, págs.207 y ss.

107 A más tardar, como expresa el artículo el 2 de febrero del 2026.

administrativo o logístico"[108]. La IA generativa sería capaz de integrarse fácilmente en este supuesto, ya que podría a través de imágenes, voces o vídeos sintéticos ser utilizada para influir en el resultado de una elección o referéndum o en el comportamiento electoral de personas físicas que ejerzan su derecho de voto en elecciones o referendos. Por ejemplo, ha sido clara su finalidad de manipulación e influencia en las elecciones de Nigeria, Eslovenia, EEUU y México confundiendo al electorado[109].

108 En la STC núm. 76/2019 de 22 mayo (RTC 2019\76) se declaró inconstitucional y nulo el apartado 1 del art. 58 bis de la Ley Orgánica 5/1985, de 19 de junio, del régimen electoral general que permitía a los partidos políticos recopilar datos relativos a las opiniones políticas ya que como mostraba el motivo e) del recurso, las modernas técnicas de análisis de la conducta sobre la base del tratamiento masivo de datos y la inteligencia artificial permiten procedimientos complejos orientados a modificar, forzar o desviar la voluntad de los electores y sin que estos sean conscientes de ello.

109 En México, las redes sociales y las apps de mensajería instantánea (WhatsApp y Telegram) son los principales escenarios para la difusión de falsa información creando confusión entre la población. Por ejemplo, en un deep fake aparecía Claudia Sheinbaum, la candidata de la alianza "Seguimos haciendo historia", invitando a los electores a escribir en la boleta el nombre del presidente Andrés Manuel López Obrador, para que, supuestamente, continuara en la Presidencia. Por otra parte, Jorge Álvarez Máynez apareció en un vídeo en el que supuestamente estaba pidiendo a sus electores declinar su voto en favor de la candidata de la alianza Fuerza y corazón por México, Xóchitl Gálvez (CALDERÓN C.,"Elecciones México 2024: Deep fakes y fake news ganan en las votaciones", El financiero, 4 de junio del 2024,https://www.elfinanciero.com.mx/elecciones-mexico-2024/2024/06/04/deep-fakes-y-fake-news-marcaron-el-escenario-electoral/ (recuperado el 3 de agosto del 2024)). En EEUU son numerosos los ejemplos: algunos demócratas de New Hampshire recibieron llamadas automáticas que se habían generado por todos los Estados Unidos con la voz de Joe Biden instándoles a que se quedaran en casa en vez de ir a votar en las elecciones primarias del Estado(LINDSAY, J. M., "Elecciones 2024: La amenaza falsa a las elecciones del 2024", Council on Foreign Relations, 2 de febrero, 2024, https://www.cfr.org/blog/election-2024-deepfake-threat-2024-election (recuperado el 3 de agosto del 2024)). Utilizando imágenes de la campaña de Kamala Harris, se manipuló la voz en un vídeo haciendo declaraciones controvertidas que nunca hizo, aunque posteriormente el vídeo se etiquetó como una parodia, el daño ya estaba hecho (BANDARA, P., "Elon Musk compartió un vídeo engañoso de inteligencia artificial de la vicepresidenta Kamala Harris en X", PetaPixel, 29 de julio del 2024 (https://petapixel-com.translate.goog/2024/07/29/elon-musk-shared-misleading-ai-video-of-kamala-harris-on-x-twitter-deepfake-election-presidential-usa/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc) (recuperado el 3 de agosto del 2024). Por otra parte, aparecía un vídeo de Taylor Swift manipulado, en una cuenta pro-Trump X que tiene más de un millón de seguidores (TENBARGE, K.,"Taylor Swift, deepfakes en X describe falsamente que apoya a Trump", NBC NEWS, 8 de febrero del 2024, https://www.nbcnews.com/tech/internet/taylor-swift-deepfake-x-falsely-depict-supporting-trump-grammys-flag-rcna137620 (recuperado el 3 de agosto del 2024)). Las elecciones en Nigeria de febrero de 2023 han sido objeto de numerosas desinformaciones, se difundió un clip de audio manipulado que implicaba falsamente a un

Debería considerarse, en esta situación, IA de alto riesgo y estar sujeta a los controles y obligaciones de transparencia de este tipo de IA.

Un sector doctrinal considera que la IA generativa y concretamente el deepfake debería haberse calificado directamente de alto riesgo[110]. Es más que evidente, como se expuso anteriormente, que pone en riesgo directo a numerosos de los derechos fundamentales reconocidos en el Texto de la LIA, y genera confusión y desinformación que pueden afectar a la capacidad de decisión en una elección o referéndum.

candidato presidencial en planes para manipular los resultados electorales. Así lo muestra ABDULKABBER, T.,"Funcionarios y famosos: así circuló la información falsa en las elecciones nigerianas", ijnet (red internacional de periodistas), Lucha contra la desinformación, 13 de abril del 2023 ,https://ijnet.org/es/story/funcionarios-y-famosos-as%C3%AD-circul%C3%B3-la-informaci%C3%B3n-falsa-en-las-elecciones-nigerianas) (recuperado el 3 de agosto del 2024), que afirma que se han registrado una serie de noticias falsas diseñadas para influir en la decisión de los votantes de todo el país. En las elecciones en Eslovenia se desacreditó al candidato, al difundirse unas grabaciones de audio falsificadas (unos días antes de las elecciones celebradas el 30 de septiembre del 2023). En una grabación, difundida en Facebook, se escuchaban dos voces: supuestamente, la de Michal Šimečka, líder del partido liberal Eslovaquia Progresista, y la de Monika Tódová, del diario *Denník N*. discutiendo la forma de amañar el proceso electoral, en parte comprando votos de la minoría romaní o gitana, marginada del país (y también se planteaba la subida del precio de la cerveza). En la otra grabación de audio de Šimečka, notificada por Meta al equipo de Demagog, difundida en Instagram, Šimečka aparece proponiendo duplicar el precio de la cerveza si vencía, siendo falsa. Estas grabaciones afectaron los resultados electorales y son un claro ejemplo del impacto de los deepfakes en la política (MEAKER, M., "Deepfakes en elecciones de Eslovaquia rearfiman que IA es un peligro para la democracia", Wired, negocios, 3 de octubre del 2023 https://es.wired.com/articulos/deepfakes-en-elecciones-de-eslovaquia-reafirman-que-ia-es-peligro-para-democracia) (recuperado el día 2 de agosto del 2024)).

110 GAMERO CASADO, E. "El enfoque europeo de la Inteligencia Artificial", *Revista de Derecho Administrativo*, nº 20, 2021, pág.279, considera que ya que no han sido prohibidos deberían estar sujetos a los sistemas de Inteligencia Artificial de Alto riesgo así afirma el autor al referirse a la lista de Sistemas de Inteligencia Artificial prohibidos "debe repararse en que esta lista, en realidad, es corta. Y en particular, que no rechaza la implantación de sistemas de la IA cuyos resultados resultan todavía incomprensibles para la mente humana, como el deep learning, las redes neuronales o los algoritmos de caja negra. Tales sistemas podrán utilizarse, sometiéndose, en su caso, a las reglas que analizamos a continuación ", esto último se refiere a las de los sistemas de IA de alto riesgo. Seguido por CASTILLO RAMOS-BOSSINI, S.E., "Regulación europea de la inteligencia artificial", *Nuevas fórmulas de prestación de servicios en la era digital*, dirección Juan Francisco Pérez Gálvez, Dyinson, 2023, pág.326.

En esta línea, estaba el Estudio, la política europea frente a los deepfakes, de julio del 2021, al exponer que[111] "teniendo en cuenta la larga lista de riesgos e impactos adversos asociados a los deepfakes, la Comisión arguye su clara afectación a los derechos fundamentales, la salud y a la seguridad, y hace un llamamiento para que las tecnologías de IA que creen deepfakes se categoricen como de alto riesgo. Categorizar estos sistemas como de alto riesgo, implicaría la aplicación de mayores obligaciones legales al proveedor de la tecnología, incluyendo la realización de evaluaciones de riesgos, la provisión de documentación, la supervisión humana y la garantía de la calidad de los conjuntos de datos de entrenamiento". Se ha mostrado a lo largo de este texto como la IA generativa puede afectar directamente y de forma grave a los derechos fundamentales, más allá de las decisiones electorales, pudiendo ser considerada de alto riesgo en otros supuestos y concretamente en materia de deepfake, en todo caso (si no se incluye en los casos de IA prohibida como se expuso). En dicho Estudio se recogen los daños psicológicos, financieros y sociales que generan los deepfakes incluyendo en esto últimos: la manipulación de los medios de comunicación, los daños a la estabilidad económica, al sistema judicial, al sistema científico, a la democracia, a las relaciones internacionales, la erosión de la confianza, la manipulación de las elecciones y los daños a la seguridad nacional[112]. Como confirma MUÑOZ VELA los deep fakes los han concebido distintos gobiernos como fuente de problemas de seguridad nacional en los últimos años[113].

Por otra parte, en los sistemas de IA relativos a ámbitos predefinidos especificados en el Reglamento, pueden darse casos específicos, que no entrañen un riesgo considerable de causar un perjuicio a los intereses jurídicos que se encuentran amparados en dichos ámbitos, al no influir sustancialmente en el resultado

111 ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA…op.cit., lo resumen.

112 Pág.IV.

113 MUÑOZ VELA, J.M, *Retos*…op.cit.pág.96.

de la toma de decisiones se entiende que el sistema de IA no afecta al fondo, ni por consiguiente al resultado de la toma de decisiones, ya sea humana o automatizada[114].

IV.- ESPECIAL REFERENCIA A LAS OBLIGACIONES DE CIBERSEGURIDAD Y TRANSPARENCIA Y LA ÉTICA

En conexión con las obligaciones de la IA generativa en la LIA, voy a destacar algunos puntos.

Ciberseguridad

En la Inteligencia Artificial de uso general de riesgo sistémico, en la que encuadramos a la IA generativa como se ha expuesto, la ciberseguridad hace acto de presencia. El art.55 LIA (sobre las obligaciones de los proveedores de modelos de IA de uso general con riesgo sistémico), establece en su punto 1 d) que entre las obligaciones de proveedores de modelos de IA de uso general con riesgo sistémico (además de las mencionadas en los arts. 53 y 54 LIA) está la de velar porque se establezca un nivel adecuado de protección de la ciberseguridad para el modelo de IA de uso general con riesgo sistémico y la infraestructura física del modelo[115].

114 El art.6.3 LIA muestra que un sistema así podría incluir situaciones en las que se cumplen cualquiera de las siguientes condiciones:" a) que el sistema de IA esté destinado a realizar una tarea de procedimiento limitada; b) que la tarea realizada por el sistema de IA esté destinada a mejorar el resultado de una actividad humana previa realizada; c) que el sistema de IA esté destinado a detectar patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores y no esté destinado a sustituir la valoración humana previamente realizada sin una revisión humana adecuada, ni influir en ella o d) que el sistema de IA esté destinado a realizar una tarea para una evaluación que sea pertinente a efectos de lo casos de uso enumerados en el anexo III.". Véase también el Considerando 53 LIA.

115 En el Considerando 115 se refleja esta idea.

También se alude a la ciberseguridad en la IA de propósito general con riesgo sistémico en el Considerando 114 al exigir a los proveedores de esta IA garantizar un nivel adecuado de protección en materia de ciberseguridad, independientemente de si los modelos se ofrecen como independientes o integrados en Sistemas de IA o en productos. En su Considerando 115, después de enumerar los posibles riesgos cibernéticos asociados al uso malintencionado o ataques[116], ofrece algunas medidas para fortalecer la ciberseguridad, reconociendo que debe tenerse en cuenta que "esa protección podría facilitarse asegurando los pesos, los algoritmos, los servidores y los conjuntos de datos del modelo, por ejemplo, mediante medidas de seguridad operativa para la seguridad de la información, medidas específicas en materia de ciberseguridad, soluciones técnicas adecuadas y establecidas y controles de acceso cibernéticos y físicos, en función de las circunstancias pertinentes y los riesgos existentes".

Se produce así un avance ya que en la Propuesta de Reglamento de 21 de abril de 2021 de la LIA la exigencia de ciberseguridad se centraba en los Sistemas de Inteligencia Artificial de Alto Riesgo[117]. Se recoge la exigencia de ciberseguridad en la LIA para los sistemas de alto riesgo principalmente en su art.15 (que regula requisitos de precisión, solidez y ciberseguridad). El art.15 exige la resistencia de los sistemas de IA de alto riesgo, en su apartado 5, ante intentos de alteración por parte de terceros no autorizados. El art.15.1 establece que "los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme en esos sentidos durante todo su ciclo de vida". En su párrafo final el art.15 dispone que las soluciones técnicas encaminadas a garantizar la ciberseguridad de los sistemas de

116 Asociados "al uso malintencionado o a los ataques debe tener debidamente en cuenta las fugas accidentales de modelos, las divulgaciones no autorizadas, la elusión de las medidas de seguridad y la defensa contra los ciberataques, el acceso no autorizado o el robo de modelos".

117 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.164

IA de alto riesgo serán adecuadas a las circunstancias y los riesgos pertinentes.

Entre las soluciones técnicas destinadas a subsanar vulnerabilidades específicas de la IA figurarán, como refleja el art.15.5 LIA, "según corresponda, medidas para prevenir, detectar, combatir, resolver y controlar los ataques que traten de manipular el conjunto de datos de entrenamiento («envenenamiento de datos»), o los componentes entrenados previamente utilizados en el entrenamiento («envenenamiento de modelos»), la información de entrada diseñada para hacer que el modelo de IA cometa un error(« ejemplos adversarios» o «evasión de modelos»), los ataques a la confidencialidad o los defectos en el modelo "[118]. Hay que recordar que el Estudio, la política europea frente a los deepfakes, de julio del 2021 mostraba su preocupación por la detección de deepfakes.

Pese a que en el texto final de la LIA se amplía la exigencia de la obligación de garantizar la ciberseguridad más allá de la IA de alto riesgo, avanzando así en relación a la Propuesta del 2021 (no hay que olvidar que la IA generativa puede ser IA de alto riesgo), sigue siendo insuficiente ya que la obligación de la ciberseguridad debería incluirse en una disposición común (al igual que las exigencias éticas) y ser exigible con unos mínimos a todas las IA, no solo a las IA de uso general con riesgo sistémico y a la IA de alto riesgo. Ya afirmaba MUÑOZ VELA[119] que esto "evidencia una omisión de un requerimiento ético y jurídico que debería ser esencial para cualquier sistema inteligente (por ejemplo un chatbot), sea considerado de nivel medio o bajo o simplemente todavía no clasificado" y por tanto, a cualquier IA generativa ya que el acceso a los datos "de la más inocente" de las IA puede utilizarse para generar confusión, podría afectar a la privacidad

118 También se alude a la ciberseguridad en la IA de alto riesgo, entre otros, en los Considerandos 54,55,74,76,77,78,122,131, art.13.3b) ii, Anexo IV punto 2 h) y art.42.2 LIA. También se alude a la ciberseguridad en los arts.31.2, 58.2 i), 66 h),70.3 y 78.2 LIA

119 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.105

de los datos e incluso provocar riesgos mayores, por ello, todas necesitan un control mínimo. Es así reclamable la incorporación de unas normas mínimas de ciberseguridad para todo tipo de IA y evidentemente, ciberseguridad que puede y debe ser intensificada en los casos de alto riesgo y de riesgo sistémico. Las obligaciones de seguridad y concretamente de ciberseguridad están conectadas con la protección de los datos.

Tal y como expone la LIA en su Considerando 76: "La ciberseguridad es fundamental para garantizar que los sistemas de IA resistan a las actuaciones de terceros maliciosos que, aprovechando las vulnerabilidades del sistema, traten de alterar su uso, comportamiento o funcionamiento o de poner en peligro sus propiedades de seguridad". Añade que los ciberataques contra sistemas de IA pueden dirigirse contra activos específicos de la IA, como los conjuntos de datos de entrenamiento o los modelos entrenados, o aprovechar las vulnerabilidades de los activos digitales del sistema de IA o la infraestructura de TIC subyacente. Aunque el Considerando (y el art.15 LIA) se refiere a la IA de alto riesgo[120], esa afirmación podría aplicarse a cualquier IA, pues puede ser objeto de estos ciberataques expuestos.

Las obligaciones de transparencia

la IA generativa en su camino hacia la perfección de sus capacidades, va mejorando los "errores" y si, por ejemplo, en los vídeos sintéticos iniciales se podía percibir que estábamos en presencia de una inteligencia artificial, por ejemplo, en la mirada vacía del "efecto del valle inquietante", ahora es difícil de descubrir. La misma dificultad está al identificar si la voz sintética es humana

120 Añadía: "Por lo tanto, para garantizar un nivel de ciberseguridad adecuado a los riesgos, los proveedores de sistemas de IA de alto riesgo deben adoptar medidas adecuadas, como los controles de seguridad, teniendo también en cuenta, cuando proceda, la infraestructura de TIC subyacente".

o artificial[121]. En su afán por acercarse, lo más posible al ser humano, aumentan las capacidades de la IA y con ello la posible dificultad de diferenciar lo que es real de lo que no lo es. Actualmente, ChatGPT ha realizado progresos con su último modelo ChatGPT40 que incorpora, además de múltiples funciones, mejoras conversacionales (voces cada vez más "humanas") y la reproducción de sentimientos. El sistema no está reconociendo emociones, sino que intenta reproducirlas, buscando superar uno de los escollos que le separaba del ser humano, al decir que las máquinas no tienen sentimientos, cosa que ellas mismas afirman cuando se les pregunta por algo relacionado con estos temas. Pese a todas las mejoras, aún hay fallas en el sistema, por ejemplo, en la demostración la IA confundió al presentador sonriente con una superficie de madera y también, comenzó a resolver una ecuación que aún no se le había mostrado, por ello aún queda camino que andar, hay fallos, alucinaciones que convierten a los chatbots en potencialmente inseguros y poco fiables[122]. Lo que es innegable es que cada vez va a ser más difícil diferenciar que es real y que es artificial, contenido generado o manipulado con la IA generativa, siendo necesaria la identificación del origen de la creación.

Por otra parte, el riesgo de la oscuridad de la IA en general, y concretamente de la generativa, conlleva una obligación de transparencia, cuyo cumplimiento en principio no conlleva la licitud del uso[123]. Las obligaciones de transparencia de los proveedores y

121 FRANGANILLO, J. "La inteligencia artificial...op. cit. págs.8 y 9. Siguiendo a NIGHTINGALE, S.J, y FARID, H., "Los rostros sintetizados por IA son indistinguibles de los rostros reales y más confiables" Proc Natl Acad Sci US A. 2022 22 de febrero; 119(8): e2120481119. Publicado en línea el 14 de febrero de 2022. doi: 10.1073/pnas.2120481119 (recuperado el 28 de mayo).

122 Tal y como muestra el artículo de KLEINMAN, Z., El tiempo, novedades tecnológicas, "Las seis nuevas funciones de la última versión de ChatGPT: es capaz de coquetear y detectar emociones" ,15 de mayo del 2024, (https://www.eltiempo.com/tecnosfera/novedades-tecnologia/6-nuevas-funciones-de-la-ultima-version-de-chatgpt-que-es-capaz-de-coquetear-y-detectar-emociones-y-las-fallas-que-cometio-3342771) (recuperado el 30 de mayo del 2024).

123 Así expone el Considerando 137 LIA que "el cumplimiento de las obligaciones de transparencia aplicables a los sistemas de IA que entran en el ámbito de aplicación del

responsables de sistemas de IA generativa se reflejan en capítulo IV de la LIA que recoge las obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA.

La exigencia de transparencia fundamental es que las personas que interactúan directamente con una IA estén informadas de que están actuando con un sistema de IA y no con un ser humano, así se dispone que "1. Los proveedores garantizarán que los sistemas de IA destinados a interactuar directamente con personas físicas se diseñen y desarrollen de forma que las personas físicas de que se trate estén informadas de que están interactuando con un sistema de IA" (art.50 LIA).

Si bien establece una excepción "cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización". Pienso que realmente, la excepción supone una carga excesiva sobre el usuario, la persona que interactúa con la IA, que podría evitarse obligando en todo momento a la identificación o, en su caso, con carácter excepcional a la persona que por sus conocimientos profesionales deberían reconocerlo (Lex artis). Lógicamente, cuando la IA adquiere la forma de un avatar que es claramente identificable y no es idéntico o prácticamente idéntico a una figura humana sería fácil diferenciarlo, con lo que no se genera confusión, pero ¿hasta qué punto es evidente en otras situaciones y se puede exigir el conocimiento de la interacción con la IA? ¿qué grado de diligencia sería exigible?

presente Reglamento no debe interpretarse como un indicador de que la utilización del sistema de IA o de sus resultados de salida es lícito en virtud del presente Reglamento o de otras disposiciones del Derecho de la Unión y de los Estados miembros, y debe entenderse sin perjuicio de otras obligaciones de transparencia aplicables a los responsables del despliegue de sistemas de IA establecidas en el Derecho de la Unión o nacional".

En la misma línea, de considerar la excepción antes expuesta (y propuesta) a los profesionales, está la siguiente excepción legal[124] que excluye de la obligación cuando son los profesionales los que utilizan el sistema respecto a delitos, con la excepción de que el sistema esté a disposición del público para denunciar el delito penal, ya que implica una interactuación con persona física que puede desconocer que está interactuando con la IA.

El segundo punto del art.50 LIA, directamente se refiere a la IA generativa, al afirmar "los proveedores de sistemas de IA, entre los que se incluyen los sistemas de IA de uso general, que generen contenido sintético de audio, imagen, vídeo o texto, velarán por que los resultados de salida del sistema de IA estén marcados en un formato legible por máquina y que sea posible detectar que han sido generados o manipulados de manera artificial". Se centra en cómo tienen que ser las soluciones técnicas[125] y excluye de la obligación a "los sistemas de IA que desempeñen una función de apoyo a la edición estándar o no alteren sustancialmente los datos de entrada facilitados por el responsable del despliegue o su semántica, o cuando estén autorizados por ley para detectar, prevenir, investigar o enjuiciar delitos".

En su punto cuarto, el artículo distingue entre los responsables del despliegue de un sistema de IA que genere o manipule imágenes o contenidos de audio o vídeo que constituyan una ultrasuplantación[126] y, por otra parte, texto que se publique con el fin de informar al público sobre asuntos de interés público.

124 "Esta obligación no se aplicará a los sistemas de IA autorizados por ley para detectar, prevenir, investigar o enjuiciar delitos, con sujeción a las garantías adecuadas para los derechos y libertades de terceros, salvo que estos sistemas estén a disposición del público para denunciar un delito penal".

125 "Los proveedores velarán por que sus soluciones técnicas sean eficaces, interoperables, sólidas y fiables en la medida en que sea técnicamente viable, teniendo en cuenta las particularidades y limitaciones de los diversos tipos de contenido, los costes de aplicación y el estado actual de la técnica generalmente reconocido, según se refleje en las normas técnicas pertinentes".

126 Art.3 concepto 60 "ultrasuplantación: un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares, entidades

En el primer caso (imágenes, audio y video) de ultrasuplantación, "harán público que estos contenidos o imágenes han sido generados o manipulados de manera artificial. Esta obligación no se aplicará cuando la ley autorice su uso para detectar, prevenir, investigar o enjuiciar delitos". Esta obligación de transparencia aparece matizada, en su Considerando 134, nos muestra que esta obligación además de con las soluciones técnicas utilizadas, se consigue a través del etiquetado de los resultados de salida generados por la IA. Se aclara en el artículo, que cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, estas obligaciones de transparencia se limitarán "a la obligación de hacer pública la existencia de dicho contenido generado o manipulado artificialmente de una manera adecuada que no dificulte la exhibición o el disfrute de la obra".

En relación al punto cuarto, sobre la intervención de la IA generativa en una creación artística, señalar que los artistas cada vez más están usando las redes neuronales para sus obras. Esto no impide la valoración de la misma en sí, su apreciación, pero siempre partiendo de la base de que somos conocedores del origen o de la intervención de la IA generativa en su creación. Ha de recordarse que hay obras sintéticas que han sido premiadas desconociendo su origen. Debe respetarse el derecho a la creación artística. La obligación de transparencia, que nos libre de la confusión y la desinformación, el saber que la obra ha sido generada o manipulada artificialmente, debe buscar el camino adecuado para no afectar a la obra en sí y la posibilidad de su disfrute. El Ministerio de Cultura ha dejado claro que las obras exclusivamente generadas con IA no podrán ser premiadas[127].

o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos".

127 "Las obras creadas "exclusivamente" con Inteligencia Artificial no podrán ganar un Premio Nacional de Cultura" El Periódico de España, Cultura, noticias efe, 19 de febrero de 2024 (https://www.epe.es/es/cultura/20240219/obras-creadas-exclusivamente-inteligencia-artificial-codigo-buenas-practicas-ministerio-cultura-98374508) (recuperado el 13 de mayo

En el segundo caso expuesto en la norma, si se refiere a texto que se publique con el fin de informar al público sobre asuntos de interés público, los responsables "divulgarán que el texto se ha generado o manipulado de manera artificial. Esta obligación no se aplicará cuando el uso esté autorizado por ley para detectar, prevenir, investigar o enjuiciar delitos, o cuando el contenido generado por IA haya sido sometido a un proceso de revisión humana o de control editorial y cuando una persona física o jurídica tenga la responsabilidad editorial por la publicación del contenido"[128]. Se prevé una obligación de divulgación similar a la anterior. Debe recordarse que la generación automática de textos etc. sin control puede provocar la desinformación. Es por ello necesario el control humano, pero ese control humano no debe excluir en ningún caso la identificación de la actuación de la Inteligencia Artificial, en nuestro caso generativa. Aunque tengamos el control de terceras personas o la responsabilidad reconocida a una persona por la publicación del contenido, hay que actuar con prevención no asegurando sin más la responsabilidad concreta "humana" sino evitando el problema (que se produzca el daño) con la referencia, en todo caso, a la intervención de la IA en el texto.

Esta obligación de transparencia del uso de la IA generativa o de sus resultados (en ambos supuestos) no obstaculiza el derecho a la libertad de expresión ni al de creatividad artística y científica, garantizados por la Carta de los Derechos Fundamentales de

del 2024). El Ministerio de Cultura ha elaborado una guía de buenas prácticas relativas a la IA para regular su uso y se aplicarán en: la contratación de actividades y servicios creativos, los Premios Nacionales y las subvenciones.

128 En esta línea de control, el Reglamento (UE) 2024/1083 del Parlamento Europeo y del Consejo, de 11 de abril de 2024, por el que se establece un marco común para los servicios de medios de comunicación en el mercado interior y se modifica la Directiva 2010/13/UE (Reglamento Europeo sobre la Libertad de los Medios de Comunicación), DOUE, nº 1083, de 17 de abril de 2024 establece en su art.18.1 que "Los prestadores de plataformas en línea de muy gran tamaño facilitarán una funcionalidad que permita a los destinatarios de sus servicios: e) declarar que no ofrecen contenidos generados por sistemas de inteligencia artificial sin someterlos a revisión humana o control editorial".

la Unión Europea (arts.11 y 13), particularmente cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, con sujeción a unas garantías adecuadas para los derechos y libertades de terceros, como expone el Considerando 134.

En el apartado 5 del art.50 LIA se afirma que la información de los apartados anteriores "se facilitará a las personas físicas de que se trate de manera clara y distinguible a más tardar con ocasión de la primera interacción o exposición". La inmediatez del aviso (que estamos ante una actuación de la IA) evita el problema, desde el primer acercamiento a la IA y debe conocerse[129]. Como muestra el artículo, desde el punto de vista formal la información debe ser clara y distinguible y ofrecerse, como muy tarde desde la primera interacción o exposición. Añade el artículo que "la información se ajustará a los requisitos de accesibilidad aplicables"[130].

En relación a la obligación de detectar y divulgar que los resultados son originados por la IA generativa, en el Considerando 136, se expresa "que las obligaciones impuestas a los proveedores y a los responsables del despliegue de determinados sistemas de IA en el presente Reglamento destinadas a permitir que se detecte y divulgue que los resultados de salida de dichos sistemas han sido generados o manipulados de manera artificial resultan especialmente pertinentes para facilitar la aplicación efectiva del Reglamento (UE) 2022/2065" del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales).

129 Entiendo que al hacer referencia al primer acercamiento carece de sentido "a más tardar" con la ocasión de la primera interacción o exposición, podría ser más conveniente "desde" la primera interacción o exposición.

130 Manifiesta el art.50 en su punto 6 que, estas obligaciones (de los apartados 1 a 4) no afectarán a los requisitos y obligaciones establecidos en el capítulo III de la LIA y "se entenderán sin perjuicio de otras obligaciones de transparencia establecidas en el Derecho nacional o el de la Unión para los responsables del despliegue de sistemas de IA".

Se aplica particularmente, según el mencionado Considerando, "en lo referente a las obligaciones de los prestadores de plataformas en línea de muy gran tamaño o de motores de búsqueda en línea de muy gran tamaño para detectar y mitigar los riesgos sistémicos que pueden surgir de la divulgación de contenidos que hayan sido generados o manipulados de manera artificial, en particular el riesgo de los efectos negativos reales o previsibles sobre los procesos democráticos, el discurso cívico y los procesos electorales, como a través de la desinformación. La exigencia de etiquetar los contenidos generados por sistemas de IA con arreglo al presente Reglamento se entiende sin perjuicio de la obligación prevista en el artículo 16, apartado 6, del mencionado Reglamento (UE) 2022/2065[131]".

Un claro ejemplo de la información a las personas físicas del uso de la IA generativa y su detección y etiquetado lo tenemos en Instagram (Meta). Ahora cuando se sube una imagen a las historias, se hace una referencia a que se etiquete el uso de IA (pone etiquetar IA) y establece unas normas sobre el etiquetado de contenido generado con IA en Instagram. Se pide una actuación por parte del usuario y se reconoce la actividad de detección que realiza Instagram (explica por qué debes etiquetar el contenido generado con IA en Instagram, cómo funciona el etiquetado de IA en Instagram, cuándo es obligatoria la etiqueta "creado con IA"[132] y cómo etiquetar contenido generado con IA).

131 Añade, "para los prestadores de servicios de alojamiento de datos de tratar las notificaciones que reciban sobre contenidos ilícitos en virtud del artículo 16, apartado 1, de dicho Reglamento, y no debe influir en la evaluación y la decisión sobre el carácter ilícito del contenido de que se trate. Dicha evaluación debe realizarse únicamente con referencia a las normas que rigen la legalidad del contenido".

132 "Meta exige que etiquetes el contenido que compartas y que contenga vídeos fotorrealistas o audios que parezcan realistas generados o alterados digitalmente, incluso con IA. Esto significa que, si este tipo de contenido se crea o se modifica con una herramienta de IA o creación digital, debes etiquetarlo antes de compartirlo. Meta no te exige que etiquetes imágenes que se hayan creado o alterado con IA. Sin embargo, estas imágenes recibirán la etiqueta si nuestros sistemas detectan que se generaron o modificaron mediante IA. Nota: Podría haber sanciones si no etiquetas el contenido según proceda. Aquí te mostramos algunos ejemplos de contenido creado digitalmente que debe etiquetarse: Un vídeo de un

"La oficina de IA fomentará y facilitará la elaboración de códigos de buenas prácticas a escala de la Unión para promover la aplicación efectiva de las obligaciones relativas a la detección y el etiquetado de contenidos generados o manipulados de manera artificial" (apartado 7 art.50 LIA)[133]. En el cumplimiento de las obligaciones de transparencia juegan un papel esencial los códigos de buenas prácticas. "Sin perjuicio del carácter obligatorio y de la plena aplicabilidad de las obligaciones de transparencia", señala el Considerando 135, "la Comisión podrá también fomentar y facilitar la elaboración de códigos de buenas prácticas a escala de la Unión, a fin de facilitar la aplicación eficaz de las obligaciones en materia de detección y etiquetado" mencionadas. Añade el Considerando que "también para apoyar disposiciones prácticas para que, según proceda, los mecanismos de detección sean accesibles y facilitar la cooperación con otros agentes de la cadena de valor, difundiendo los contenidos o comprobando su autenticidad y procedencia, a fin de que el público pueda distinguir efectivamente los contenidos generados por IA"[134].

grupo de personas caminando por un mercado al aire libre que parezca realista, Un archivo de audio de dos personas hablando, una canción creada por voces generadas por IA, un reel narrado con una voz en off realista generada con IA. Aquí te mostramos algunos ejemplos de contenido creado digitalmente que no es necesario que etiquetes: un vídeo de un paisaje al aire libre creado con un estilo similar al de los dibujos animados, un vídeo en el que se ha cambiado ligeramente el tamaño y se ha recortado mínimamente. Puedes obtener más información sobre este requisito en el Centro de transparencia" (Información obtenida en Instagram).

133 Se añade que "La Comisión podrá adoptar actos de ejecución a fin de aprobar dichos códigos de buenas prácticas, de conformidad con el procedimiento establecido en el artículo 56, apartado 6. Si considera que el código no es adecuado, la Comisión podrá adaptar un acto de ejecución que especifique normas comunes para el cumplimiento de las citadas obligaciones de conformidad con el procedimiento de examen establecido en el artículo 98, apartado 2". El Considerando 117 trata sobre la importancia de los códigos de buenas costumbres en el cumplimiento de las obligaciones de los proveedores de modelos de IA de uso general.

134 Por su parte, el Considerando 107 LIA establece para aumentar la transparencia en relación con los datos utilizados en el entrenamiento previo y el de los modelos de IA de uso general, respetando los derechos de autor, que los proveedores de esos modelos elaborarán y pondrán a disposición del público un resumen detallado de los contenidos usados para el entrenamiento.

La ética

La LIA busca una IA fiable, de confianza, afirma en su Considerando 27 que el "enfoque basado en el riesgo es la base de un conjunto proporcionado y eficaz de normas vinculantes", reconociéndose la importancia de recordar las Directrices éticas para una IA fiable. La fiabilidad nos la da el conocimiento, la alfabetización, las medidas de transparencia e información, entre otras cosas, y principalmente las Directrices éticas. En esta línea, menciona las Directrices éticas para una IA fiable, de 2019, elaboradas por el Grupo Independiente de Expertos de Alto Nivel sobre IA creado por la Comisión, en las que se desarrollan siete principios éticos no vinculantes para la IA que tienen por objeto contribuir a garantizar la fiabilidad y el fundamento ético de la IA. Dichos principios son: acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas. Continúa reconociendo que "sin perjuicio de los requisitos jurídicamente vinculantes del presente Reglamento y de cualquier otro acto aplicable del Derecho de la Unión, esas directrices contribuyen al diseño de una IA coherente, fiable y centrada en el ser humano, en consonancia con la Carta y con los valores en los que se fundamenta la Unión". De acuerdo con las directrices del Grupo de Expertos, "por "acción y supervisión humanas" se entiende que los sistemas de IA se desarrollan y utilizan como herramienta al servicio de las personas, que respeta la dignidad humana y la autonomía personal, y que funciona de manera que pueda ser controlada y vigilada adecuadamente por seres humanos".

En el texto se incorporan así, expresamente, las Directrices éticas básicas (acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas) pero se pierde la oportunidad de convertirlas directamente en deberes jurídicos exigibles a todo tipo

de IA, además teniendo en cuenta que ya fueron incorporadas como tales y desapareció posteriormente el artículo que las contemplaba (4 bis)[135]. Esto sería, independientemente de que estas directrices sean la base de la regulación del Reglamento, con las correspondientes matizaciones o añadidos en función del mayor o menor riesgo del sistema de IA. Se exigirían, entonces a las IA generativas (incide el Considerando en el carácter no vinculante de las mismas). Destacar, la acción y supervisión humanas, lo que aleja de las IA autónomas que tengan la última decisión. Quedan así, reflejadas expresamente en dicho Considerando, sin incorporarlas expresamente como normas de carácter general aplicable a todas las IA en el articulado del Reglamento[136].

Convertir las exigencias éticas en jurídicas marcaría las diferencias con otros países, facilitaría el punto de partida de actuación en materia de IA con ellos. Mostraría a la ciudadanía el auténtico valor del tratamiento ético, e iría calando en la misma. En la reciente Convención Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho de 17 de mayo del 2024[137], se ha vuelto a perder otra

135 Fueron considerados principios generales aplicables a todos los sistemas de IA en las enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 en su Enmienda 213 en la que se incluyó una propuesta de Reglamento con un artículo 4 bis que los recogía normativamente como principios generales aplicables a todos los sistemas de IA (Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), Procedimiento legislativo ordinario: primera lectura , Enmienda 213 , Propuesta de Reglamento, artículo 4 bis (nuevo)). Señalando en su apartado 1 que " Todos los operadores que entren en el ámbito de aplicación del presente Reglamento se esforzarán al máximo por desarrollar y utilizar los sistemas de IA o modelos fundacionales con arreglo a los siguientes principios generales que establecen un marco de alto nivel para promover un enfoque europeo coherente centrado en el ser humano con respecto a una inteligencia artificial ética y fiable, que esté plenamente en consonancia con la Carta, así como con los valores en los que se fundamenta la Unión".

136 MUÑOZ VELA, J.M, *Retos*...op.cit.págs.100 y 101, considera un riesgo la ausencia de normas éticas vinculantes , estando a favor de las mismas.

137 13ª Sesión del Comité de Ministros (Estrasburgo, 17 de mayo de 2024), Comisión de Inteligencia Artificial (CAI), CM (2024)52-final.

oportunidad de dar valor jurídico exigible a las exigencias éticas esenciales en el modo de entender cualquier tipo de IA , recogiéndolas expresamente en el articulado como normas éticas.

No queda más que afirmar que la IA generativa, al servicio del hombre, debe seguir la senda de la constante evolución, guiada de mano humana sustentada en una ética exigible, para que el camino de rosas no se convierta en camino de espinas, para que Gizmo no genere seres como Stripe que puedan poner en jaque a la humanidad.

V.- CONCLUSIONES

1. La LIA no define directamente la IA generativa, la engloba en la IA de uso general. La transformación en la forma de generar contenidos de la IA generativa nos proporciona múltiples beneficios en todas las esferas de nuestra vida. La inteligencia artificial generativa permite la creación automática de contenido "original" en diferentes formatos (texto, audio, vídeo e imágenes). Un contenido difícil de diferenciar del creado por humanos. Ofrece tantos beneficios como posibles riesgos para la sociedad, democracia y derechos fundamentales. Son contenidos tan "reales" que pueden llevar a la confusión y a la desinformación, desconociendo que es real y que no. Provocan la manipulación y actuaciones delictivas a través, principalmente de fake news, deepvoice y deepfakes. Estos últimos son especialmente peligrosos, los vídeos hipertrucados, las ultrasuplantaciones, siendo una de las mayores preocupaciones de los gobiernos.
2. Son numerosos los riesgos que puede conllevar la IA generativa. Los riesgos de desinformación, confusión, ataque al derecho al honor, la intimidad y la propia imagen (provocados

principalmente por los deepfakes) que tiene que convivir con el derecho a la libertad de información y expresión. El riesgo de infracción y vulneración de los derechos de autor, ante creaciones que se nutren de otras de acceso público por internet sin retribución inicial y con posibles problemas de plagio y copyright, en las que serán necesarios los correspondientes consentimientos. El riesgo en el mercado laboral, con la supresión de puestos de trabajo, si bien aparecen nuevos trabajos y cualificación laboral e incremento de la productividad. Los riesgos del medio ambiente ante el gasto energético que supone y la busca de salidas como la energía nuclear. Los de sesgos al poder reproducir estereotipos (y producirse discriminaciones) con datos sin filtrar, cuyo entrenamiento tiene que calibrarse y reentrenarse, e incluso minimizarse el uso de datos. Es necesaria la alfabetización, es decir la comprensión y capacitación digital en materia de IA generativa y la democratización en su acceso y comprensión. Riesgos por su uso malicioso, lo que va a exigir un control para mitigar los efectos negativos y la desinformación generada por la confusión o por modelos mal entrenados. Riesgo de toxicidad al reproducir el lenguaje de la red. Riesgos por errores y alucinaciones, ya que la IA generativa crea respuestas lógicas que pueden ser falsas. Riesgo para la seguridad, principalmente de los datos personales, ante posibles fugas o uso indebido o datos inexactos, y además el riesgo para la privacidad, de la memorización y réplica posterior de datos por la IA y el posible peligro para la confidencialidad. La ciberseguridad es esencial ante ataques como, por ejemplo, el del "fraude del CEO". Riesgos psicológicos ante fenómenos como "revivir" a personas fallecidas, la posible exclusión laboral o sextorsion. Riesgo en relación a cuestiones éticas y la exigibilidad de Códigos de Conducta, y en relación a la formulación de las instrucciones en la alienación de valores.

3. La IA generativa podría incluirse directamente en la IA prohibida en los supuestos 5.1 a) y 5.1b) LIA. En el momento que das una información, imagen o vídeo falsos o manipulados (deepfakes, fake news, deep voice) que no permite distinguir fácilmente la realidad de la ficción (sobre todo con los deepfakes), de forma subliminal o deliberada se puede manipular y engañar, con el objetivo o efecto de alterar de manera sustancial, el comportamiento de una persona o colectivo. Lo que impide tomar una decisión informada y le o les lleve a tomar otra que le o les provoca o puede provocar perjuicios considerables. Si bien, interpreto en el caso de hacerse de forma subliminal que debe estar prohibida siempre. Igual ocurre cuando se exploten vulnerabilidades para alterar sustancialmente el comportamiento, haciendo creer lo que no es, y estas personas no son capaces de discernir, creando confusión, desinformación y manipulación. En los otros supuestos de Inteligencias Artificiales prohibidas, la prohibición sería indirecta, podría la IA generativa servirles, por ejemplo, creándoles datos sintéticos, en estos casos su uso, también debería estar prohibido.
4. La IA generativa puede ser IA general con riesgos sistémicos. Además de la presencia de posibles capacidades de gran impacto, es indudable que la IA generativa puede tener efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto. Queda corroborado al analizar previamente sus riesgos. Estos riesgos pueden propagarse a gran escala a lo largo de toda la cadena de valor. A su vez, queda reflejada esa posibilidad en el Considerando 136 de la LIA. La IA generativa podría incluirse directamente en el supuesto de IA del alto riesgo en los Sistemas de IA destinados a influir en el resultado de una elección o referéndum, por lo antes expuesto (la posible manipulación, confusión y desinformación). Debería incluir-

se directamente el deepfake como IA de alto riesgo, al poner en riesgo numerosos derechos fundamentales y poder generar confusión y desinformación que afecte a la capacidad de decisión en una elección o referéndum (y en cualquier decisión). En esta línea está el Estudio, la política europea frente a los deepfakes, de julio del 2021.

5. Se hace especial referencia a las obligaciones de ciberseguridad y transparencia. La ciberseguridad es tan importante que, pese a abrirse el abanico de su exigencia además de a la IA de alto riesgo (art.15 LIA) a la de uso general con riesgo sistémico (art.55 .1 d) LIA), a diferencia de redacciones anteriores de la LIA (2021), realmente debería ser una exigencia para todas las Inteligencias Artificiales. La IA "más inocente" puede abrir la puerta a la confusión, afectar a la privacidad de datos y provocar riesgos mayores. Obligación de transparencia, ante la oscuridad de la IA generativa, reflejada en el art.50 LIA. Es fundamental que las personas físicas sepan que están interactuando con una inteligencia artificial siendo criticable su excepción, en relación al posible conocimiento por parte de la persona física. Un aspecto destacable es la importancia de la ética. La LIA recoge en su Considerando 27,7 Directrices éticas básicas (acción y supervisión humanas; solidez técnica y seguridad, gestión de la privacidad y de los datos, transparencia, diversidad, no discriminación y equidad, bienestar social y ambiental y rendición de cuentas) perdiendo la oportunidad de convertirlas en jurídicas, en normas comunes a todo tipo de IA (incluida la generativa), independientemente de que queden reflejadas en su articulado.

VI.- BIBLIOGRAFÍA

ABDULKABBER, T., "Funcionarios y famosos: así circuló la información falsa en las elecciones nigerianas", ijnet (red internacional de periodistas), Lucha contra la desinformación, 13 de abril del 2023 ,https://ijnet.org/es/story/funcionarios-y-famosos-as%C3%AD-circul%C3%B3-la-informaci%C3%B3n-falsa-en-las-elecciones-nigerianas) (recuperado el 3 de agosto del 2024).

AFP, "La fiscalía investiga si Meta vulnera la protección de datos de sus usuarios", El Mundo, Empresas, 4 de julio de 2024 (recuperado el 7 de julio del 2024).

ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA ante los deepfakes de desnudos", 2 de octubre del 2023, Cuatrecasas (https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/el-reglamento-de-ia-ante-los-deepfakes-de-desnudos) (recuperado el 1 de mayo del 2024).

ANDRÉS, R., "La última tendencia para bordar las entrevistas de trabajo: entrenar con ChatGPT como reclutador", Xataca, 6 de marzo del 2024 (recuperado el 28 de julio del 2024).

ANNEMANS,R.,"Seguridad de la IA generativa: 8 riesgos que debes conocer". GlobalSing by GMO, 4 de diciembre del 2023 https://www.globalsign.com/es/blog/8-riesgos-de-seguridad-de-la-inteligencia-artificial-generativa (recuperado el 10 de mayo del 2024).

BANDARA, P., "Elon Musk compartió un vídeo engañoso de inteligencia artificial de la vicepresidenta Kamala Harris en X", PetaPixel, 29 de julio del 2024 (https://petapixel-com.translate.goog/2024/07/29/elon-musk-shared-misleading-ai-video-of-kamala-harris-on-x-twitter-deepfake-election-presidential-usa/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc) (recuperado el 3 de agosto del 2024).

BARRIO ANDRÉS, M., "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", *Diario La Ley*, nº 86. Sección Ciberderecho, 30 de julio de 2024.

BENDITO CAÑIZARES, M.T, "Estadio intermedio de reflexión para una futura regulación de la ética en el espacio digital europeo: los principios de transparencia y accountability", *Revista Aranzadi de Derecho y Nuevas Tecnologías*, núm 55/2021, BIB 2021/1465.

BORRACHERO GARRO, A. "Los retos de la tecnología en la publicidad: la campaña de Lola Flores", coordinador: ORTEGA BURGOS, E., *Propiedad intelectual*, 2022, Documento TOL9.141.406.

CALDERÓN C.,"Elecciones México 2024: Deep fakes y fake news ganan en las votaciones", El financiero, 4 de junio del 2024,https://www.elfinanciero.com.mx/elecciones-mexico-2024/2024/06/04/deep-fakes-y-fake-news-marcaron-el-escenario-electoral/ (recuperado el 3 de agosto del 2024).

CASTILLO RAMOS-BOSSINI, S.E., "Regulación europea de la inteligencia artificial", *Nuevas fórmulas de prestación de servicios en la era digital*, dirección Juan Francisco Pérez Gálvez, Dyinson, 2023.

DAVID, DemoCreator, "Los 10 mejores generadores de música IA gratis en 2024", 13 de marzo de 2024 (https://dc.wondershare.es/ai-voice/top-free-ai-music-generators.html) (recuperado el 7 de mayo del 2024).

DEL CASTILLO, C., "Los creadores del canon AEDE quieren una "tasa ChatGPT" para la inteligencia artificial, el Diario.es, 3 de mayo del 2023, (https://www.eldiario.es/tecnologia/creadores-canon-aede-quieren-tasa-chatgpt-inteligencia-artificial_1_10171676.html), (recuperado el 20 de mayo del 2024).

DURAN, I., "Taylor Swift y sus desnudos hechos con IA: CEO de Microsoft dice "hay que actuar ya" ante los deepfakes, In-

fobae, 29 de enero del 2024, https://www.infobae.com/tecno/2024/01/27/taylor-swift-y-sus-desnudos-hechos-ia-ceo-de-microsoft-dice-hay-que-actuar-ya-ante-los-deepfakes/ (rescatado el 3 de agosto del 2024).

ESPUGA TORNÉ, G. "Cómo identificar contenido generado por IA" Linkedin, junio, 2024 (recuperado el 10 de julio del 2024).

FIGUEROA, J.C., "Crear una sola imagen con inteligencia artificial consume tanta energía como cargar tu teléfono", Hipertextual, Tecnología, 12 de diciembre de 2023 (https://hipertextual.com/2023/12/crear-imagen-con-inteligencia-artificial-consume-esta-energia) (recuperado el 20 de abril del 2024).

FERNÁNDEZ HERNÁNDEZ, C. y EGUILUZ CASTAÑEIRA, J., "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", *Diario LA LEY*, Sección Ciberderecho, nº 85, 28 de junio de 2024.

FRANGANILLO, J., "La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos", *methaodos. revista de ciencias sociales* (2023) 11(2) m231102a1010.17502/mrcs.v11i2.710.

GAMERO CASADO, E. "El enfoque europeo de la Inteligencia Artificial", *Revista de Derecho administrativo*, nº 20, 2021, págs.268 y ss.

GARAY, J., Wired, 19 de mayo del 2023, "El G 7 promete regular las IA generativas antes de que termine el 2023"(https://es.wired.com/articulos/g7-promete-regular-las-ia-generativas-antes-de-que-termine-el-2023).

GARCÍA MEXÍA, P. "Europa ante el reto de la inteligencia artificial",-The objective, 3 de agosto del 2024, (https://theobjective.com/tecnologia/2024-08-03/europa-ante-el-reto-de-la-inteligencia-artificial/) (recuperado el 5 de agosto del 2024).

GOLDMAN SACHS, Principales riesgos que entraña la inteligencia artificial generativa: enumeración, fundspeople, (https://fundspeople.com/es/principales-riesgos-que-entrana-la-inteligencia-artificial-generativa/), 26 de abril de 2023 (rescatada en 20 de abril del 2024).

HOLCOMBE, J., "Las 9 Mejores Herramientas de Detección de Contenidos con IA que Tienes que Conocer", Kinsta, 5 de abril del 2023 https://kinsta.com/es/blog/deteccion-de-contenidos-ia/ (recuperada el 6 de julio del 2024).

JARQUES, A., "El futuro Reglamento Europeo de Inteligencia Artificial", *Actualidad Jurídica Aranzadi,* nº1003, 2023, Editorial Aranzadi, (BIB 2024/278) (rescatada en 27 de mayo del 2024).

KLEINMAN, Z., El tiempo, novedades tecnológicas, "Las seis nuevas funciones de la última versión de ChatGPT: es capaz de coquetear y detectar emociones" ,15 de mayo del 2024, (https://www.eltiempo.com/tecnosfera/novedades-tecnologia/6-nuevas-funciones-de-la-ultima-version-de-chatgpt-que-es-capaz-de-coquetear-y-detectar-emociones-y-las-fallas-que-cometio-3342771) (recuperado el 30 de mayo del 2024).

LALCHAND, S. et al. (Centro de Servicios Financieros de Deloitte), "Se espera que la IA generativa aumente el riesgo de deepfakes y otros fraudes de la banca", Deloitte, Servicios Financieros, 29 de mayo del 2024, https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html)) (recuperado el 2 de agosto del 2024).

LINDSAY, J. M., "Elecciones 2024: La amenaza falsa a las elecciones del 2024", Council on Foreign Relations, 2 de febrero, 2024, https://www.cfr.org/blog/election-2024-deepfake-threat-2024-election (recuperado el 3 de agosto del 2024).

LUCIO LÓPEZ, L.A., "Deep fake porn, la inteligencia artificial da nueva cara al ciberacoso escolar", Ciem, 2024.

MARTÍNEZ ESPÍN, P., "La propuesta de marco regulador de los sistemas de inteligencia artificial en el mercado de la UE", *Revista CESCO de Derecho de Consumo*, nº46/2023(doi.org/10.18239/RDC_2023.46.3322).

MCMAHON, L., BBC News, 20 de mayo de 2024, "Cuando la escuché me quedé en shock": por qué el programa de IA ChatGPT dejará de usar la voz que se parece a la de Scarlett Johanson, https://www.bbc.com/mundo/articles/cprrzn8g2wqo.) (recuperado el 30 de mayo del 2024).

MEAKER, M., "Deepfakes en elecciones de Eslovaquia rearfiman que IA es un peligro para la democracia", Wired, negocios, 3 de octubre del 2023 https://es.wired.com/articulos/deepfakes-en-elecciones-de-eslovaquia-reafirman-que-ia-es-peligro-para-democracia) (recuperado el día 2 de agosto del 2024).

MEIJOMIL, S., Inboundcycle, "6 generadores de imágenes con IA que no puedes perderte, 8 de noviembre del 2023, (https://www.inboundcycle.com/blog-de-inbound-marketing/generadores-de-imagenes-con-ia).

MICROSOFT SECURITY, IA generativa, La ventaja de los defensores, Linkedln, Microsoft (recuperado en 6 de mayo de 2024).

MORAN, I., Photolari, 2 de mayo, 2024, "20.000 euros por la imagen IA que engañó al jurado de los Sony Wordl Photography Awards el año pasado" (https://www.photolari.com/20-000-euros-por-la-imagen-ia-que-engano-al-jurado-de-los-sony-world-photography-awards-el-ano-pasado/) (recuperado el 20 de mayo del 2024).

MUÑOZ VELA, J. M, "Inteligencia artificial generativa. Desafíos para la propiedad intelectual", *Revista de Derecho UNED*, núm.33, 2024, Premio de artículos jurídicos "García Goyena", 22ª convocatoria (curso 2022-2023), Facultad de Derecho. UNED.

MUÑOZ VELA, J.M, *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*, Thomson Reuters- Aranzadi, Pamplona, 2022.

NIGHTINGALE, S.J, y FARID, H., "Los rostros sintetizados por IA son indistinguibles de los rostros reales y más confiables" Proc Natl Acad Sci US A. 2022 22 de febrero; 119(8): e2120481119. Publicado en línea el 14 de febrero de 2022. doi: 10.1073/pnas.2120481119 (recuperado el 28 de mayo).

PASCUAL, M.G.,Meta no ofrecerá sus nuevos modelos de IA generativa en Europa por su "impredicible entorno regulatorio", El Pais, Tecnología, 18 de julio de 2024 (recuperado el 19 de julio del 2024).

PLAZA PENADÉS, J., Dossier, "Las claves de la futura Ley de Inteligencia Artificial Europea", Aranzadi La Ley, Navarra, mayo, 2023.

PUERTO MENDOZA, A., Derecho *digital. Fundamentos básicos*, Ediciones CEF.,2019.

PUFPAFF, M. "¿La tecnología como fuerza para el bien? ¿Cómo se está utilizando la inteligencia artificial para prevenir los suicidios en China?", *Razón y fe*, Tomo 282, nº1447, 2020, págs.205 y ss.

RAA J., "Meta detiene su proyecto para entrenar a la IA con publicaciones de Facebook e Instagram en Europa", Tecnología, El País, 14 de junio de 2024 (recuperado el 1 de julio del 2024).

SÁNCHEZ, L., "Escrivá advierte que la estrategia de IA necesita de entornos seguros y anuncia una nueva ley integral de ciberseguridad", Economist & Jurist, 7 de junio del 2024 (recuperado en 20 de julio del 2024).

SINCLA, A et al." El estado de la IA a principios de 2024: la adopción de la IA generativa aumenta y comienza a generar valor", McKinsey& Company, 30 de mayo de 2024, (https://www.mckinsey.com/locations/south-america/latam/hispanoamerica-en-potencia/el-estado-de-la-ia-a-principios-de-2024-la-adopcion-de-la-ia-generativa-aumenta-y-comienza-a-generar-valor/es-CL) (recuperado el 2 de agosto de 2024).

SOTO ARAMENDARIZ, S "Primer suicidio inducido por inteligencia artificial: algo que temer", 4 de abril de 2023 (https://observatorioblockchain.com/ia/primer-suicidio-inducido-por-inteligencia-artificial-algo-que-temer) (recuperado el 28 de mayo del 2024).

SUÁREZ JAQUET, H. et HINOJAL CUADRADO, E. "El uso del deepfake en producciones audiovisuales: consideraciones jurídicas", coordinador: ORTEGA BURGOS, E., *Propiedad intelectual*, 2022, Documento TOL9.141.396

TENBARGE, K.,"Taylor Swift, deepfakes en X describe falsamente que apoya a Trump", NBC NEWS, 8 de febrero del 2024, https://www.nbcnews.com/tech/internet/taylor-swift-deepfake-x-falsely-depict-supporting-trump-grammys-flag-rcna137620 (recuperado el 3 de agosto del 2024).

VALERO, A., "Deepfakes Porn y violencia contra las mujeres", Fundación Cañada Blanch, 4 de junio de 2024, https://www.fundacioncanadablanch.org/noticias/deepfakes-porn-y-violencia-contra-las-mujeres/ (recuperado el 2 de agosto del 2024).

VIDAL. M, en Linkedln, noticia, (Fuente: https://www.nytimes.com/2024/05/28/technology/ai-chief-executives.html) (recuperado el 7 de junio de 2024).

VIGARIO, D.,"Un año de libertad vigilada para los 15 jóvenes que manipularon y difundieron imágenes con IA de menores desnudas en Almendralejo", El mundo, 9 de julio de 2024 (recuperado en 1 de agosto del 2024).

WIKIPEDIA, Greemlins, https://es.wikipedia.org/wiki/Gremlins (recuperado el 1 de mayo del 2024).

"Las empresas de inteligencia artificial ofrecen ya el servicio de recrear a seres queridos fallecidos e interactuar con ellos", 20 minutos, 20 bits, 3 de diciembre del 2023, (https://www.20minutos.es/tecnologia/empresas-inteligencia-artificial-ofrecen-servicio-recrear-seres-queridos-fallecidos-interactuar-con-ellos-5195903/) (recuperado el 1 de junio del 2024).

"Las obras creadas "exclusivamente" con Inteligencia Artificial no podrán ganar un Premio Nacional de Cultura" El Periódico de España, Cultura, noticias efe, 19 de febrero de 2024 (https://www.epe.es/es/cultura/20240219/obras-creadas-exclusivamente-inteligencia-artificial-codigo-buenas-practicas-ministerio-cultura-98374508) (recuperado el 13 de mayo del 2024).

"Marilyn conoce a James Bond en la película Deepfake del artista", El informe de Marilyn, 5 de marzo de 2024, "https://themarilynreport-com.translate.goog/2024/03/05/marilyn-meets-james-bond-in-artists-deepfake-movie/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc (recuperado el 2 de agosto del 2024).

"El Comisionado de Hamburgo para la protección de Datos y la Libertad de Información pública en documento en el que se analiza los grandes Modelos de Lenguaje desde el punto de vista de la protección de datos, Boletín de julio del 2024", Lks, noticias, (https://www.lksnext.com/es/noticias_boletin/el-comisionado-de-hamburgo-para-la-proteccion-de-datos-y-la-libertad-de-informacion-publica-un-documento-en-el-que-analiza-los-grandes-modelos-de-lenguaje-desde-el-punto-de-vista-de-la-proteccion-de-d/) (recuperado el 2 de agosto del 2024).

LA INNOVACIÓN EN LA LEY EUROPEA DE INTELIGENCIA ARTIFICIAL

Por PABLO GARCÍA MEXÍA

Consultor-director de Derecho digital en Herbert Smith Freehills
Co-director del Posgrado en Privacidad, sociedad digital e IA en la UAM
Letrado de las Cortes Generales

REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

Algunos tachan la innovación de ideología, "la ideología del capitalismo tecnológico" (Aibar 2023). Una de las metas de este trabajo es demostrar que no es así. En cualquier caso, considerar la innovación como un aspecto determinante del crecimiento, y por tanto de la prosperidad económica, es hoy prácticamente general. Es lo que explica que la Unión Europea la acoja como elemento esencial de su estrategia en materia de inteligencia artificial (IA), y más en concreto del Reglamento del Parlamento Europeo y del Consejo por el que se establecen reglas armonizadas sobre inteligencia artificial (Ley de inteligencia artificial, en adelante LIA).

I.- ANTECEDENTES Y CONTEXTO

La LIA no llega en efecto en el vacío, sino que tuvo un antecedente clave en el Libro Blanco de la Comisión sobre la IA, publicado en 2020, que esbozaba una visión pretendidamente equilibrada entre excelencia y confianza en este campo. Al tiempo, la LIA forma parte de un conjunto de políticas más amplio, que incluye el Plan Coordinado sobre IA y el Paquete de Innovación en IA.

El Plan Coordinado sobre Inteligencia Artificial, publicado por primera vez en 2018, representa un compromiso conjunto de la Comisión, los Estados miembros de la UE, Noruega y Suiza para mejorar la competitividad mundial de Europa en IA. En el plan inicial se detallaban las acciones y los instrumentos de financiación para apoyar el desarrollo de la IA en diversos sectores, animando a los Estados miembros a crear sus propias estrategias nacionales. La última actualización del plan se publicó en 2021, haciendo hincapié en acelerar las inversiones en tecnologías de IA y armonizar las políticas de IA para reducir la fragmentación a escala de la Unión.

El Paquete de Innovación en IA se publicaba en 24 de enero de 2024, de la mano de la Comunicación de la Comisión sobre el impulso de las empresas emergentes y la innovación en inteligencia artificial fiable, con el fin de reforzar las acciones anteriores. Sobre esta base, la Comisión proponía reforzar la capacidad informática europea, ofreciendo la red de supercomputación de la UE a las empresas emergentes (o startups) de manera asequible; desarrollar el talento, mediante la creación de programas educativos más especializados para satisfacer la creciente demanda de profesionales de IA y apoyar a las empresas emergentes, mediante la inversión en iniciativas que las ayuden, sin perjuicio de mantener la fiabilidad de sus diseños conforme a los valores europeos.

Este Paquete de Innovación obtenía el respaldo del Consejo de la UE, mediante el Reglamento (UE) 2024/1732 del Consejo, de 17 de junio de 2024. El Reglamento precisa que el acceso a la red europea de supercomputación deberá ser lo más amplio y justo posible, teniendo especialmente en cuenta a las empresas emergentes y a las pequeñas y medianas empresas (Pymes); garantiza una contribución financiera de la Unión de hasta el 50 % de los costes de adquisición y funcionamiento de los superordenadores dedicados a IA; y establece que los superordenadores dedicados a IA deben utilizarse principalmente para desarrollar soluciones de IA en la Unión, especialmente generativa y a gran escala.

II.- LA INNOVACIÓN EN LA LIA

Ninguna mejor declaración de intenciones de la LIA a propósito de la innovación que su artículo 1 (y los considerandos [1] y [2]), en cuanto establece como fines de dicha norma: la mejora del funcionamiento del mercado interior, la promoción de una IA centrada en el ser humano y fiable, la protección frente a los

daños que la IA pueda generar a la democracia y los derechos fundamentales (entre ellos la salud, la seguridad, y el medio ambiente), y, en último término, el apoyo a la innovación (previsto también en el apartado 2[g] de dicho artículo).

Aunque la menciona en último lugar, la UE reconoce pues que la innovación resulta determinante en general, y muy particularmente en una tecnología tan emergente como la IA, a la hora de detallar los fines de la LIA. Estos son ante todo humanismo y fiabilidad, como desde hace ya bastantes años se viene proclamando en textos internacionales dedicados a la materia (Declaración de Asilomar, textos de Naciones Unidas, OCDE, Consejo de Europa, etc.). Es bien razonable que así sea. La anteposición del funcionamiento del mercado interior se explica por la idea de hacer de la IA "un producto", garantizando un alto nivel de protección y seguridad jurídica para los usuarios (considerandos [3] y [8]), cuyos derechos se protegen justamente en la medida en que puedan ser dañados por estas tecnologías. Otra cosa es que la enumeración citada conduzca a una lectura contrapuesta de derechos y democracia, por un lado, frente a innovación por otro; aunque sobre esta cuestión volveremos al final del trabajo.

Más allá de estos preceptos, la innovación se halla indiscutiblemente presente en otros lugares de la LIA, como atestiguan los siguientes aspectos:

a. Foco en pymes y startups:

- Se incluyen medidas especiales para apoyar a las pymes y a las empresas emergentes, facilitando su capacidad para innovar dentro de un marco normativo seguro y coherente (considerandos [143] y [145]); así como excepciones regulatorias relativas a las microempresas (considerando [146]).

b. Entornos controlados de pruebas (sandboxes):

- La LIA fomenta el establecimiento de sandboxes regulatorios a escala nacional. Estos sandboxes proporcionan

un entorno controlado para el desarrollo y las pruebas de sistemas de IA innovadores, bajo supervisión regulatoria, antes de su lanzamiento al mercado (considerandos [138] y [139]).

c. Investigación y desarrollo:

- La LIA apoya explícitamente la innovación al excluir de su ámbito de aplicación los sistemas y modelos de IA desarrollados exclusivamente para la investigación científica (considerando [25]).
- En esta misma línea, el software, los datos y los modelos de código abierto, en cuanto pueden compartirse y modificarse libremente, son reconocidos por sus contribuciones a la investigación y la innovación (considerando [102]), lo que lleva a la LIA a dispensarles un tratamiento de alguna manera privilegiado en determinados supuestos.

d. Estandarización:

- Se destaca la estandarización como un elemento crucial para brindar soluciones técnicas que garanticen el cumplimiento de la normativa, lo que puede redundar en una mayor innovación, competitividad y crecimiento (considerando [121] y art. 40.3, que impulsa a incrementar la cooperación global en este ámbito).

e. Modelos de IA generativa:

- Se reconoce que, aun cuando estos modelos presentan grandes problemas, ofrecen un importante potencial de innovación (considerando [105]).

f. Formación y condiciones de trabajo en materia de IA:

- Una formación adecuada puede mejorar las condiciones de trabajo y por ende consolidar la innovación (considerando [20]).

Más allá de todo ello, la LIA dedica un Capítulo completo, concretamente el VI, a las que considera medidas más relevantes de entre las citadas. Tales "Medidas de apoyo a la innovación" son los entornos controlados de pruebas (o sandboxes), a los que se debe añadir las pruebas de sistemas de alto-riesgo en condiciones del mundo real; y las medidas en relación con las pymes (y las microempresas).

II.1.- LOS ENTORNOS CONTROLADOS DE PRUEBAS (SANDBOXES)

Los Estados miembros deben establecer al menos un entorno oficial de pruebas en un plazo de 24 meses desde la entrada en vigor de la LIA. Puede ser nacional, regional o conjunto con otros Estados miembros, y la Comisión puede prestar apoyo a cualquiera de ellos (arts. 57.1 y 57.2).

El fin de estos sandboxes es proporcionar un entorno controlado que fomente la innovación y facilite el desarrollo, la formación, el ensayo y la validación de sistemas de IA innovadores durante un tiempo limitado, antes de su introducción en el mercado o puesta en servicio, con arreglo a un plan específico acordado entre el correspondiente proveedor y la autoridad nacional. Dichos sandboxes pueden incluir pruebas en condiciones del mundo real para su supervisión en los mismos (art. 57.5). Es además claro que estos objetivos benefician especialmente a las pymes (art. 57.9).

Gracias a las potestades de supervisión que se les concede, las autoridades nacionales pueden suspender las actividades de un determinado espacio de pruebas a la vista de la identificación de riesgos significativos en el proyecto de que se trate (art. 57.11). Por su parte, los proveedores no quedan exentos de responsabilidad sobre los posibles daños, pero sí de cualesquiera sanciones administrativas, siempre y cuando se ajusten a los planes y orientación del correspondiente sandbox (art. 57.12).

La LIA establece otras disposiciones en materia de cooperación entre las autoridades nacionales y éstas y la Oficina de Inteligencia Artificial y el Comité de Inteligencia Artificial de la Comisión; y de información pública que deberán proporcionar aquellas y la mencionada Oficina de Inteligencia Artificial (listado de sandboxes e interfaz única de acceso) (arts. 57.13 a 57.17). Queda por otro lado en manos de la Comisión la adopción de medidas que especifiquen el funcionamiento de los entornos de pruebas, incluidos los requisitos de admisibilidad, los procedimientos y las condiciones, que deberán garantizar la transparencia, flexibilidad para las autoridades nacionales, un acceso amplio (que será prioritario para las pymes), apoyo para la realización de las evaluaciones de conformidad y la participación de los agentes pertinentes (arts. 58.2 y 58.3). Los sandboxes implementarán asimismo servicios adicionales, como los de orientación en el cumplimiento de esta normativa o el apoyo en la elaboración de documentos de estandarización o certificación, entre otros (art. 58.3). Respecto de las pruebas en el mundo real, los acuerdos que se establezcan entre autoridades nacionales y proveedores deberán proteger los derechos fundamentales, la salud y la seguridad; y, cuando fuera procedente, se cooperará con otras autoridades nacionales (art. 58.4).

Atención especial de la LIA merece el tratamiento de datos personales en entornos regulatorios de pruebas. No es de extrañar, dado que, como es sabido, el de privacidad es uno de los principales riesgos derivados de la IA, a su vez debido a la enorme voracidad de datos de estos sistemas y modelos, imprescindible para su entrenamiento. La LIA es pese a todo generosa respecto de los sandboxes, ya que se permite el tratamiento de datos personales para sistemas de IA en áreas de interés público (salud, medio ambiente, energía, etc.), aunque sea bajo condiciones estrictas. Y sea como fuere, el tratamiento habrá de cumplir con la legislación de protección de datos de la Unión, muy singularmente en lo que hace a su seguridad (arts. 57.10 y 59).

Como se sabe España se adelantó a la aprobación de la LIA con el lanzamiento de la regulación nacional que la desarrollaba en materia de entornos controlados de pruebas. Se trata del Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la (entonces Propuesta de) LIA. Está abierto a proveedores y usuarios de sistemas de alto-riesgo conforme a la normativa indicada, quienes serán seleccionados, tras las oportunas convocatorias de la Administración, conforme a criterios entre los que la innovación ocupa lugar destacado.

Por último, la LIA abre la posibilidad de efectuar pruebas de sistemas de IA de alto-riesgo en el mundo real, incluso fuera de un entorno regulatorio controlado. Las condiciones para ello son de todos modos muy estrictas, pues: ha de tratarse de sistemas de alto-riesgo del Anexo III (sistemas biométricos, infraestructuras críticas, educación, salud, justicia, etc.); debe existir un plan de pruebas aprobado por la correspondiente autoridad nacional de supervisión de mercado, conforme a su vez a muy exigentes condiciones; y las pruebas deben ajustarse a pautas que incluyen el consentimiento informado de los destinatarios, la protección de datos y la responsabilidad por los daños y perjuicios que el sistema pudiera causar (art. 60). Obsérvese que, a diferencia de la que lo hace respecto de los entornos controlados, la autoridad nacional que debe velar por el cumplimiento de esta normativa en lo que se refiere a las pruebas fuera de sandbox, es la mencionada y correspondiente "autoridad nacional de supervisión de mercado", que el art. 3(26) LIA concreta en la que "lleva a cabo las actividades y adopta las medidas derivadas del Reglamento (UE) 2019/1020", relativo a la vigilancia del mercado y la conformidad de los productos, y que en España es la llamada Oficina de Enlace Única, incardinada en la Dirección General de Consumo. Tiene sentido, fundamentalmente a la vista de la densa red de cooperación que se prevé entre Estados miembros, y estos y la Comisión Europea, existente respecto de los entornos

controlados, mejor servidos por autoridades específicamente encargadas de la supervisión de la IA (y con independencia de que desarrollen solo esa u otras actividades).

Un precepto específico (art. 61), de perfil muy similar al que en el Reglamento general de protección de datos se ocupa del consentimiento, regula el consentimiento informado para la realización de pruebas en el mundo real. Este deberá ser: otorgado libremente, previo a la participación en las pruebas; posterior a haber recibido una información concisa, clara, relevante y comprensible; y revocable sin sufrir prejuicio alguno. El consentimiento deberá documentarse, a la par que deberá entregarse una copia a los interesados.

II.2.- EL TRATAMIENTO DE LAS PYMES

La segunda gran línea de acción en pro de la innovación establecida en la LIA se refiere a las pymes, incluyendo en ellas a las empresas emergentes (o startups) y a las microempresas.

El objetivo es apoyar especialmente a estas empresas en el cumplimiento de esta normativa, pues es bien sabido que las empresas grandes, no digamos las que lo son mucho, están en mejores condiciones de hacer frente a exigencias regulatorias de calado, como sin duda son las derivadas de la LIA. El artículo 62 prevé tres grandes medidas en este sentido: a) acceso prioritario de las pymes a los entornos regulatorios de pruebas, actividades de formación y canales de comunicación específicos, y facilidades de participación en procesos de estandarización; b) tasas reducidas a efectos de la obtención de una evaluación de conformidad; y c) especial atención de la Oficina de Inteligencia Artificial de la Comisión, quien deberá proporcionar plantillas a efectos del cumplimiento de la LIA, mantener plataformas de información y campañas de comunicación ad hoc, y promover buenas prácticas en materia de contratación pública relativa a sistemas de IA.

El artículo 63.1 establece por su parte que las microempresas podrán cumplir de manera simplificada con determinados elementos de los requisitos de gestión de calidad previstos para los proveedores de sistemas de alto-riesgo (art. 17 LIA). Ello se llevará a cabo conforme a directrices elaboradas por la Comisión.

El propio artículo 63 LIA remite a la Recomendación 2003/361/CE de la Comisión Europea, a efectos de establecer la definición de microempresa, siendo esta "una empresa que ocupa a menos de 10 personas y cuyo volumen de negocios anual o cuyo balance general anual no supera los 2 millones de euros" (art. 2.3 del Anexo de la Recomendación 20337361/CE).

Llama no obstante la atención el tenor del apartado segundo del artículo 63 LIA. Hace notar que la simplificación de requisitos prevista en el apartado primero "no se deberá interpretar como una exención respecto de las microempresas del cumplimiento de cualesquiera otros requisitos u obligaciones" en la propia LIA establecidos, "incluyendo los previstos en los artículos 9, 10, 11, 12, 13, 14, 15, 72 y 73." Es una disposición superflua. El artículo 63.1 dispone lo que dispone, una simplificación, nada más; y perfectamente concretada en torno al artículo 17 de la propia norma. No existiría base alguna para construir exención de ningún tipo a partir de esa simple disposición. Si esa interpretación que se pretende evitar sería del todo improcedente, sobra pues esa aclaración, que por ello mismo termina siendo antipática hacia justamente las empresas más débiles, las que se pretende proteger. Y tanto más antipática cuanto se enfatiza con la expresión "cualesquiera otras", y se remacha, para cerrar, con una expresa mención de artículos que en todo caso se les impondrán. Inmerecido por la microempresa europea (e incluso por quienes siendo tan pequeños pretendan acceder a nuestro mercado).

III.- CÓDIGOS DE BUENAS PRÁCTICAS Y CÓDIGOS DE CONDUCTA

El tratamiento de la innovación en la LIA no estaría completo sin una alusión al llamado soft-law en este campo, materializado en los códigos de buenas prácticas y los códigos de conducta. Como se sabe, esta técnica regulatoria dista de ser nueva para la UE, estando sin ir más lejos presente, entre otros muchos posibles ejemplos, en el citado Reglamento General de Protección de Datos. Si abordamos aquí esta cuestión es porque el soft-law constituye una atractiva vía intermedia para incentivar el cumplimiento de determinadas pautas que el poder público considera deseables, sin acudir a la imposición de la legislación, por naturaleza imperativa, y por ello mismo arriesgada, cuando lo que a la par se pretende es salvaguardar la innovación.

Esta es justamente la razón por la que otros modelos regulatorios mundiales de la IA a escala global evitan, al menos de momento, las normaciones imperativas del estilo de la europea, para decantarse en cambio por el soft-law, en forma de códigos de seguimiento voluntario por parte de las empresas. Es fundamentalmente el caso del modelo estadounidense, plasmado en el Decreto presidencial Biden de 30 de octubre de 2023 sobre un desarrollo y uso seguros y fiables de la IA, sin perjuicio de sus disposiciones vinculantes en materia de defensa y seguridad nacional y de la voluntad de hacer cumplir normativas ya imperativas existentes (como es la de consumo).

En cualquier caso, la LIA introduce también el soft-law. Aun cuando no lo hace de un modo uniforme. Como acabamos de decir, introduce dos tipos de códigos que nada tienen que ver entre sí. Uno de ellos, el de los códigos de buenas prácticas, no pretende en modo alguno fomentar la innovación, sino, por el contrario, reforzar el cumplimiento de la propia LIA, que en cualquier caso resulta vinculante para los propios destinatarios de esos códigos de buenas prácticas. Es cierto en cambio, que los que la LIA denomina códigos de conducta, y que también prevé,

Revista de Privacidad y Derecho Digital
ISSN digital 2444-5762, NÚM. 34, Madrid, Mayo-Agosto (2024), págs. 316-330

sí que se orientan a promover la innovación, lo que, salvadas todas las demás distancias, asimilaría su función a las "guías" que tan profusamente prevé el Decreto Biden de los EE.UU.

III.1.- LOS CÓDIGOS DE BUENAS PRÁCTICAS

La LIA prevé a su vez un doble régimen para los códigos de buenas prácticas. El artículo 50.7 regula los establecidos respecto de sistemas de IA. La idea es que sea la Oficina de Inteligencia Artificial quien fomente y facilite su elaboración a escala de la Unión, para facilitar la aplicación efectiva de determinadas obligaciones, en concreto las relativas a la detección y el etiquetado de contenidos generados o manipulados artificialmente. Tales códigos serán aprobados por la Comisión, en el supuesto de resultar a su juicio adecuados. En caso contrario, impondrá reglas comunes para dar cumplimiento a tales obligaciones.

Para los modelos, de nuevo la Oficina de Inteligencia Artificial, y esta vez según el artículo 56 (en concordancia con los arts. 53, 55 y 89.1), tiene la tarea de promover la creación de códigos de buenas prácticas para apoyar la correcta aplicación de la LIA. Estos códigos deben abordar, entre otras, las siguientes obligaciones: a) Mantener la información actualizada en función de los desarrollos tecnológicos y de mercado; b) proporcionar resúmenes relativos a la formación que se lleve a cabo; c) identificar y gestionar los riesgos sistémicos a escala de la Unión; y d) documentar las medidas de control de riesgos, teniendo en cuenta su gravedad y probabilidad.

La Oficina de Inteligencia Artificial podrá invitar a proveedores de modelos de IA, autoridades nacionales, organizaciones de la sociedad civil, empresas, instituciones académicas y otras partes interesadas a participar en la elaboración de estos códigos. Los códigos deberán describir claramente sus objetivos e incluir compromisos o medidas con indicadores clave de rendimiento (KPIs) para garantizar su cumplimiento. La citada Oficina y el

Comité de Inteligencia Artificial supervisarán la eficacia de los códigos en el logro de sus objetivos y publicarán evaluaciones acerca de su idoneidad. Por su parte, la propia Comisión podrá aprobar códigos de buenas prácticas, de validez extensiva a toda la Unión. Los códigos de buenas prácticas deberán completarse en un plazo de nueve meses a partir de la entrada en vigor de la LIA; si un determinado código no se completara en ese plazo, o, si pasado un plazo de 12 meses, la Comisión no lo considerase adecuado, esta podrá establecer normas comunes de obligado cumplimiento.

III.2.- LOS CÓDIGOS DE CONDUCTA

No puede descartarse que uno de los objetivos de los que la LIA denomina códigos de conducta sea reforzar el cumplimiento de esta legislación. Si quienes los suscriben se ajustan a sus términos, sin duda reforzarían esa vigencia. Un objetivo de mayor relieve de estos es no obstante promover tal cumplimiento de modo voluntario, entre destinatarios no obligados por la LIA, o solo obligados parcialmente a la misma. Y todo ello con el fin primordial de fomentar la innovación, gracias a evitar una imperatividad que, en estos casos, se considera innecesaria.

El Art. 95.1 abre la posibilidad de suscribir estos códigos de conducta a proveedores de sistemas IA que no sean de alto-riesgo para sujetarse a todas o algunas de las obligaciones de los de alto-riesgo (previstas en el Capítulo III, sección segunda). Adicionalmente, el artículo 95.2 brinda esta posibilidad de acogerse a códigos de conducta a proveedores y usuarios de todo tipo de sistemas, para sujetarse a requisitos específicos conforme a las pautas allí previstas.

Los códigos de conducta se hallan directamente emparentados con una interesante iniciativa de la Comisión Europea, adoptada tras el acuerdo formal entre Comisión, Consejo y Parlamento sobre la LIA de 8 de diciembre de 2023. Se trata del denominado

AI Pact, encaminado a fomentar el intercambio de experiencias e información entre empresas (de la documentación disponible no se desprende que puedan participar entes públicos) y la Comisión acerca de sus expectativas y proyectos de cumplimiento de la LIA, así como compromisos específicos a este respecto. El beneficio, especialmente reputacional, aunque también formativo, para los firmantes, parece claro. Cientos de empresas se han adherido ya, mediante un procedimiento que, por cierto, es muy sencillo y se completa enteramente en línea.

III.3.- NOTA COMÚN SOBRE GOBERNANZA

El Comité europeo de inteligencia artificial, con arreglo al artículo 66(e)(i) LIA, está encargado de emitir recomendaciones y dictámenes acerca del desarrollo y aplicación, tanto de los códigos de buenas prácticas, como de los códigos de conducta.

IV.- VALORACIÓN FINAL

En el mencionado Paquete de Innovación en IA, la propia Comisión Europea nos recuerda que tres superordenadores europeos están entre los diez mayores del mundo (Leonardo, Lumi y Mare Nostrum 5, este último en Barcelona). Mientras que un prestigioso ranking de universidades especializadas en IA arrojaba el quizá sorprendente resultado de que la Universidad de Granada es la número 30 del mundo (por delante de Oxford o Harvard), y número 1 de toda la UE, en investigación sobre IA.

Sin embargo, según datos de 2024 de la OCDE, más del 90% de la inversión mundial de capital-riesgo en IA, que se disparó de 2.700 millones de euros en 2022 a 24.000 millones de euros en 2023, se hace en Estados Unidos. La UE solo cuenta con 9 empresas entre las 100 mayores digitales del mundo (España

tiene una), frente a 58 que son estadounidenses. Y de entre los 16 mayores modelos de IA generativa del mundo, solo hay uno europeo (alemán en concreto), siendo los otros 15 de los EE.UU.

Pese a que esas notas de talento en IA y esos datos de infraestructura digital citados invitan a la esperanza, la Unión parte pues de muy atrás. Cosa poco deseable, cuando es de nuevo la Comisión Europea quien deja constancia de que solo la IA generativa creará un valor empresarial de entre 2,4 y 4 billones de euros anuales.

Nada de esto se le escapa a la UE, razón por la cual aborda incuestionablemente la innovación en la LIA. El tiempo dirá si las disposiciones sobre entornos regulatorios de pruebas, sobre trato especial en favor de pymes o sobre asuntos como investigación y estandarización terminan siendo suficientes para revertir esos datos en nuestro favor. Que terminen siendo suficientes en una norma en la que la innovación, sí, se cuenta entre los principales objetivos, pero donde priman otros dos fines, la seguridad de una IA como producto, y la salvaguarda de los derechos y la democracia frente a sus posibles daños.

Innovar es prosperar. Y prosperar, no solo en términos de crecimiento económico, siendo esto ya de por sí deseable. Innovar es también prosperar en derechos, sean del tipo que sean.

Como cualquier otra, la tecnología digital, la IA, puede poner en riesgo derechos. Aunque también puede garantizarlos y reforzarlos: por solo poner un ejemplo de uso real, empleada en la predicción de los servicios de urgencias de un hospital, puede salvar vidas, al optimizar los tiempos de atención del personal médico. Iniciativas como el Paquete de Innovación en IA tienen esto muy en cuenta. La LIA se ha erigido en la mejor salvaguarda mundial de los derechos y libertades frente al uso de la IA. Y sus esfuerzos en materia de innovación son desde luego loables, aunque quizá pudieran haber sido mayores a fin de lograr que Europa utilice en la mayor medida posible estas tecnologías, cosa que, por lo dicho, debe tratar de hacer.

Síganos en Linked

Visite nuestra web e infórmese de las novedades y actividades formativas que realizamos

www.rdu.es

2444-5762