Revista CESCO de Derecho de Consumo
http://revista.uclm.es/index.php/cesco

Nº 53/2025

ISSN 2254-2582

https://doi.org/10.18239/RCDC_2025.53.3678

DIRECTRICES DE LA UNIÓN EUROPEA SOBRE PRÁCTICAS PROHIBIDAS DE INTELIGENCIA ARTIFICIAL*

Blanca Aparicio Araque

Contratada predoctoral FPU Departamento Derecho civil

Centro de Estudios de Consumo

Universidad de Castilla la Mancha

Resumen: Tras la aprobación por el Parlamento Europeo de la ley de Inteligencia Artificial el 13 de marzo de 2024, se articulan una serie de prohibiciones relacionadas con prácticas en las que se utilizan sistemas de inteligencia artificial, entendiéndose que menoscaban derechos fundamentales de las personas, junto con una serie de excepciones. Estas directrices recientemente aprobadas surgen para clarificar algunos conceptos inexactos, explicando la aplicación práctica del concepto jurídico, de forma que proporciona numerosos ejemplos prácticos en aras a aportar una visión general de las prácticas de inteligencia artificial que se consideran inaceptables debido a sus posibles riesgos para los seres humanos.

Palabras clave: Inteligencia Artificial, prácticas prohibidas, categorización biométrica, sistemas de alto riesgo, aislamiento de navegador remoto.

Title: Guidelines of the European Union on prohibited Artificial Intelligence practices

Abstract: Following the adoption by the European Parliament of the Artificial Intelligence Act on 13 March 2024, a number of prohibitions related to practices involving the use of artificial intelligence systems, which are understood to diminish fundamental rights of individuals, are articulated, along with a number of exceptions. These recently adopted guidelines are intended to clarify some inaccurate concepts, explaining the practical application of the legal concept and providing numerous

* Este trabajo es parte del Proyecto de I+D+i "Protección de consumidores y riesgo de exclusión social: seguimiento y avances" (f. PID2021-128913NB-I00), financiado por MICIU/AEI y "FEDER Una manera de hacer Europa", dirigido por Ángel Carrasco Perera y Encarna Cordero Lobato; del Proyecto de Investigación "El reto de la sostenibilidad en la cadena de suministros y la defensa del consumidor final" (ref. SBPLY/23/180225/000242), cofinanciado por el Fondo Europeo de Desarrollo Regional, en el marco del Programa Operativo de Castilla-La Mancha 2021- 2027, dirigido por Ángel Carrasco Perera y Ana Carretero García; de las Ayudas para la realización de proyectos de investigación aplicada, en el marco del Plan Propio de investigación, cofinanciadas en un 85% por el Fondo Europeo de Desarrollo Regional (FEDER), para el Proyecto de Investigación "Modelos jurídicos eficientes de consumo sostenible" (ref. 2022-GRIN-34487), dirigido por Ángel Carrasco Perera y Ana I. Mendoza Losana:

practical examples in order to provide an overview of artificial intelligence practices that are considered unacceptable due to their potential risks to human beings.

Keywords: Artificial Intelligence, prohibited practices, biometric authentication, high risk systems, remote browser isolation.

I. ANTECEDENTES

La denominada "Ley de Inteligencia Artificial"[1] fue aprobada el 24 de marzo de 2024. Hay algunos autores que consideran que este Reglamento conforma el paquete de acciones sobre Inteligencia Artificial de la Unión Europea más completo, pues consta de varias iniciativas y documentos destinados a apoyar el desarrollo y la implementación de la IA en Europa de manera ética y responsable, entendiendo que dentro de las iniciativas clave se encuentran las siguientes: un marco legal de la UE sobre IA, proponiendo un marco legal que aborda los riesgos asociados con la IA y estableciendo un sistema de calificación de riesgos para diferentes tipos de aplicaciones de IA; un plan de acción para la IA, estableciendo varias iniciativas específicas en áreas como la educación, la investigación y la inversión para apoyar el desarrollo y la implementación de la IA en Europa; un informe sobre la estrategia de datos de la UE, presentando un informe que establece una estrategia de datos ambiciosa para la UE; así como guías éticas para la IA, publicándose nuevas directrices éticas para el desarrollo y la implementación de la IA en la UE[2]. El Reglamento también preveía la creación de un Comité Europeo de Inteligencia artificial, compuesto por representantes de los Estados Miembros y presidido por la Comisión Europea y que tendrá como función facilitar la aplicación uniforme del

[1] 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).

[2] MARTÍN RODRIGUEZ, G., *Nuevos horizontes en las políticas de la UE en materia de inteligencia artificial: hacia el Derecho Europeo de la IA*. GARCÍA SÁNCHEZ, B., JIMÉNEZ GARCÍA, F., (coord.), La atribución de una responsabilidad jurídico penal e internacional de la inteligencia artificial, ed. Iustel, Madrid, 2023, pág. 381.

mismo y emitir dictámenes y recomendaciones sobre cuestiones relacionadas con la inteligencia artificial[3].
Otros autores consideran que esta aproximación realizada por la Unión Europea al mercado interior en este ámbito se basa en normas comunes de seguridad completadas esencialmente por normas sobre la responsabilidad del productor, entendiendo que el régimen de responsabilidad contractual y extracontractual de bienes y servicios será competencia del legislador nacional. Consideran que los Estados miembros disponen de sistemas de responsabilidad civil que, aunque no están armonizados, garantizan el resarcimiento de los daños causados a los afectados y obligan a los responsables al pago de indemnizaciones[4].

Es necesario destacar la especial distinción que realiza de los diversos tipos sistemas de inteligencia artificial, en función del grado de afección de los mismos a los derechos fundamentales de las personas, clasificándolos en: inaceptable, alto, limitado y mínimo. Los sistemas de inteligencia artificial de alto riesgo son aquellos que producen un riesgo significativo derivado del desarrollo, despliegue y el uso de la inteligencia artificial, robótica o tecnologías conexas, pudiendo causar daños o lesiones a las personas o sociedad vulnerando derechos fundamentales o normas de seguridad. Las obligaciones exigidas a estos sistemas tienen que ver con la exigencia de la transparencia en su actuación[5].

Tal y como se establece en el artículo 96.1, apartado b, de la Ley de IA la Comisión debía adoptar directrices sobre la aplicación de las prácticas prohibidas en virtud del artículo 5 de la referida normativa. Las Directrices que sintetizaremos posteriormente, tienen por objeto aumentar la claridad jurídica y proporcionar información sobre la interpretación que la Comisión hace de las prohibiciones del artículo 5 de la Ley de IA, con el fin de garantiza su aplicación coherente, efectiva y uniforme.

El pasado 4 de febrero la Comisión publicó las Directrices sobre prácticas prohibidas de inteligencia artificial[6] (así definidas en la Ley de IA) en aras a concretar cuáles son las prácticas que se consideran inaceptables en cuanto al uso de IA se refiere, debido a la afección de derechos fundamentales que conllevan. De otro lado, también proporciona una serie de excepciones en las que no entra en juego la referida

[3] Últ. ob. cit., pág. 390.
[4] ARAGÄO SEIA C., *Inteligencia artificial: responsabilidad civil 3.0*. El impacto de la era digital en el derecho, LÓPEZ ULLA J.M. (dir.), QUIROGA CORTI, M.P., (coor.), ed. Aranzadi, Pamplona, 2023, pág. 506. En este sentido, se indica que el régimen tradicional de responsabilidad portugués tiene por objeto reparar los daños personales y materiales causados a los particulares. Esta reparación es competencia de los tribunales ordinarios a petición del titular del bien o derecho que ha sido violado y suele ser a título de indemnización.
[5] Esta misma distinción también se recogía en la Resolución del Parlamento Europeo, de 20 de octubre de 2020, que en su artículo 7 hace referencia a "las tecnologías de inteligencia artificial de alto riesgo", incluidos el software y los datos utilizados o producidos por estas tecnologías, indicando que deben desarrollarse, implementarse o utilizarse de manera que se garantice la supervisión humana completa en todo momento.
[6] Approval of the content of the draft Communication from the Commission - Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act).

prohibición, al entender que otro derecho debe prevalecer: el interés público y la seguridad nacional.

El artículo 5 de la Ley de IA prohíbe la comercialización, la puesta en servicio o el uso en la UE de determinados sistemas de IA para prácticas de manipulación, explotación, control social o vigilancia que, por su naturaleza inherente, vulneran los derechos fundamentales y los valores de la Unión. Tal y como se establece en el considerando 28 de la Ley de IA, estas prácticas son especialmente perjudiciales y abusivas y deben prohibirse porque contradicen los valores de la Unión de respeto de la dignidad humana, la libertad, la igualdad, la democracia y el Estado de Derecho, así como los derechos fundamentales consagrados en la Carta de Derechos Fundamentales de la UE, incluido en derecho a la no discriminación (art. 21), derecho a la igualdad (art. 20), derecho a la protección de datos (art. 8), entre otros.

II. BASE JURÍDICA, ÁMBITO MATERIAL Y PERSONAL DE LAS PROHIBICIONES Y ÁMBITO DE EXCLUSIÓN

Se recogen una serie de prohibiciones, cuyo contenido, ejemplos prácticos y excepciones desarrollaremos en las próximas líneas y son la siguientes: (i) manipulación dañina y engaño, (ii) explotación dañina de vulnerabilidades, (iii) puntuación social, (iv) criminalidad individual: evaluación y predicción del riesgo de delitos, (v) raspado no dirigido para desarrollar bases de datos de reconocimiento facial, (vi) reconocimiento de emociones, (vii) categorización biométrica, (viii) identificación biométrica a tiempo real.
Respecto a la base jurídica de las prohibiciones, nos encontramos con el artículo 114 del Tratado de Funcionamiento de la Unión europeo, referente a la base jurídica del mercado interior; así como el art. 16, que versa sobre la base jurídica de protección de datos, que sirve de base jurídica para las normas específicas sobre el tratamiento de datos personales en relación con la prohibición del uso de sistemas de identificación biométrica remota (RBI) con fines policiales, sistemas de categorización biométrica con fines policiales y la evaluación del riesgo individual.

En cuanto al ámbito material, las prácticas prohibidas del art. 5 de la ley de IA se refieren a la comercialización, la puesta en servicio o la utilización de sistemas de IA específicos; mientras que en lo que respecta a los sistemas de identificación biométrica en tiempo real (RBI) la prohibición solo se aplica a su uso. Respecto al concepto *"puesta en mercado"* se refiere a *«la primera puesta a disposición de un sistema de IA […] en el mercado de la Unión» y el concepto "puesta a disposición"* se define como el suministro del sistema «*para su distribución o utilización en el mercado de la Unión en el curso de una actividad comercial, ya sea a cambio de un pago o de forma gratuita*». En cuanto a "puesta en servicio" se refiere a «*el suministro de un sistema de IA para su primer uso al implementador o para uso propio en la Unión para su finalidad prevista*». Por último, el término "usar" abarcar tanto el uso como la implementación del sistema en cualquier momento de su ciclo de vida tras su comercialización y puesta en servicio.

Respecto al ámbito personal, la ley de IA distingue tres categorías diferentes de operadores en relación con los sistemas de IA: proveedores, implementadores, importadores, distribuidores y fabricantes de productos. Define alguna de esta figuras como sigue: los proveedores son: *"personas físicas o jurídicas, autoridades públicas, agencias u otros organismos que desarrollan sistemas de IA o los hacen desarrollar y los comercializan en el mercado de la Unión, o los ponen en servicio con su propio nombre o marca registrada";* los implementadores son "*personas físicas o jurídicas, autoridades públicas, agencias u otros organismos que utilizan sistemas de IA bajo su autoridad, a menos que el uso sea para una actividad personal no profesional".* Destaca que los operadores pueden desempeñar más de una función simultáneamente en relación con un sistema de IA, y, además, el cumplimiento continuo de la Ley de IA es obligatorio durante todas las fases del ciclo de vida de la IA.

En cuanto a la exclusión del ámbito de aplicación de la Ley de IA, el artículo 2 de la misma establece una serie de exclusiones generales. La Ley de IA excluye expresamente de su ámbito de aplicación los sistemas de IA que se introduzcan en el mercado, se pongan en servicio o se utilicen, con o sin modificaciones, exclusivamente con fines militares, de defensa o de *seguridad nacional*, independientemente del tipo de entidad que lleve a cabo dichas actividades. En cuanto al término de "*seguridad nacional*", se refiere al *"interés primordial en la protección de las funciones esenciales del Estado y los intereses fundamentales de la sociedad y abarca la prevención y el castigo de actividades capaces de desestabilizar gravemente las estructuras constitucionales, políticas, económicas o sociales fundamentales de un país y, en particular, de amenazar directamente a la sociedad, a la población o al propio Estado, como las actividades terroristas".* Por ejemplo, no cubra las actividades relacionadas con la seguridad vial o la organización.

III. ACLARACIÓN DE DIVERSOS CONCEPTOS

En cuanto a la cooperación judicial con terceros países, cuando proceda esta exclusión podrá abarcar las actividades de entidades privadas a las que el tercer país en cuestión haya encomendado tareas específicas en apoyo de dicha cooperación policial y judicial. Al mismo tiempo, para que la exclusión sea aplicable, estos marcos de cooperación o acuerdos internacionales deben incluir salvaguardias adecuadas con respecto a la protección de los derechos y libertades fundamentales de las personas, que deberán ser evaluadas por las autoridades de vigilancia del mercado competentes para la supervisión de los sistemas de IA utilizados en el ámbito de la aplicación de la ley y la justicia.
Respecto a la investigación y desarrollo, esta no se aplica a ninguna actividad de investigación, ensayo o desarrollo relativa a sistemas o modelos de IA antes de su comercialización o puesta en servicio. Ahora bien, existirá la obligación de cumplir con la Ley de IA cuando un sistema de IA se comercialice o se ponga en servicio como resultado de dicha investigación y desarrollo

En relación con la actividad profesional y no profesional, el artículo 2, apartado 10, de la Ley de IA establece que esta *«no se aplica a las obligaciones de los*

implementadores que sean personas físicas que utilicen sistemas en el curso de una actividad puramente personal y no profesional». Por ejemplo, una persona que utilice un sistema de reconocimiento facial en su casa quedaría excluida del referido artículo. En cuanto a los sistemas de IA publicado bajo licencias libres y códigos abiertos, tampoco se aplicación la prohibición a los mismos a menos que se comercialicen o pongan en servicio como sistemas de IA de alto riesgo, o el art. 50 que se refiere a las obligaciones de transparencia de determinados sistemas.

En lo referente a la interacción de las prohibiciones con los requisitos para los sistemas de IA de alto riesgo, el uso de sistemas de IA clasificados como de alto riesgo puede, en algunos casos, considerarse una práctica prohibida si se cumplen todas las condiciones de una o más de las prohibiciones del artículo 5 de la Ley de IA. Por el contrario, la mayoría de los sistemas de IA que se acogen a una excepción a una prohibición del artículo 5 de la Ley de IA se considerarán de alto riesgo. Respecto a la interacción entre las prohibiciones y otras disposiciones del Derecho de la Unión, la Ley de IA es un reglamento que se aplica horizontalmente en todos los sectores sin perjuicio de otra legislación de la Unión, en particular sobre la protección de los derechos fundamentales, la protección de los consumidores, el empleo, la protección de los trabajadores y la seguridad de los productos, por lo que la Ley de IA complementa dicha legislación mediante su lógica preventiva y de seguridad.

Respecto a las autoridades de vigilancia del mercado, son aquellas designadas por los Estados miembros, así como el Supervisor Europeo de Protección de Datos (en su calidad de autoridad de vigilancia del mercado para las instituciones, agencias y organismos de la UE), y son responsables de la aplicación de las normas en el Ley de IA para sistemas de IA, incluidas las prohibiciones. En relación con las sanciones, el incumplimiento de las prohibiciones del artículo 5 de la Ley de IA se considera la infracción más grave y, por lo tanto, está sujeto a la multa más elevada: los proveedores e implementadores que incurran en prácticas de IA prohibidas pueden ser multados con hasta 35 millones de euros. Además, es posible que una misma conducta prohibida constituya una infracción de dos o más disposiciones de la Ley de IA.

IV. PRÁCTICAS PROHIBIDAS

4. 1. Manipulación dañina, engaño y explotación

Las dos primeras prohibiciones del artículo 5.1, apartados a y b, de la Ley de IA tienen por objeto proteger a las personas y a las personas vulnerables de los efectos significativamente perjudiciales de la manipulación y la explotación posibilitadas por la IA. Dichas prohibiciones se dirigen a los sistemas de IA que emplean técnicas subliminales, deliberadamente manipuladoras o engañosas que son significativamente perjudiciales e influyen materialmente en el comportamiento de personas físicas o grupos de personas.
Para que se aplique la prohibición recogida en el apartado a (relacionada con el empleo de técnicas subliminales que escapen a la conciencia de una persona o técnicas deliberadamente manipuladoras o engañosas deben darse una serie de

requisitos: (i) la práctica debe constituir la comercialización, la puesta en servicio o la utilización de un sistema de IA; (ii) el sistema de IA debe utilizar técnicas subliminales (más allá de la conciencia de una persona) deliberadamente manipuladoras o engañosas; (iii) las técnicas implementadas por el sistema de IA deben tener como objetivo o efecto distorsionar significativamente el comportamiento de una persona o un grupo de personas; (iv) la conducta distorsionada debe causar o ser razonablemente probable que cause un daño significativo a esa persona, a otra persona o a un grupo de personas. Además, debe existir un vínculo causal entre las técnicas empleadas, la distorsión material del comportamiento de la persona y el daño significativo que ha resultado o es razonablemente probable que resulte de ese comportamiento.

En cuanto a ejemplos de técnicas subliminales, destacan los siguientes: mensajes visuales subliminales, mensajes subliminales auditivos, señalización subvisual y subaudible, imágenes incrustadas, desorientación o manipulación temporal. Como ejemplo práctico, podría ser un juego que aprovecha las tecnologías neuronales basadas en IA y las interfaces máquina-cerebro que permiten a los usuarios controlar partes de un juego con un dispositivo que detecta la actividad cerebral, sin que esa persona sea consciente y que además supongan la captación de información muy sensible, como información bancaria personales. En cuanto a técnicas de manipulación intencionadas, se refiere a aquellas que tienen como objetivo influir, alterar o controlar el comportamiento de una persona de forma que socave su autonomía individual y su libre elección. Como resultado, los individuos se ven obligados a realizar comportamientos que no realizarían si no estuviesen sujetos a dichas técnicas. Como ejemplo de técnica de manipulación intencionada proporciona la manipulación sensorial (en la que, a través de audio, el sistema de IA provoca alternaciones del estado de ánimo) o manipulación personalizada. En tercer lugar, las técnicas engañosas que son aquellas que subvierten o menoscaban la autonomía, la toma de decisiones o la libre elección de una persona de maneras que la persona no es consciente o, cuando lo es, aún puede ser engañada o no puede controlarlas ni resistirse a ellas En este sentido, destaca la obligación del proveedor de garantizar que los sistemas de IA que interactúan con personas estén diseñados de manera que informen a las personas de que interactúan con una IA y no con un humano. Por ejemplo, el etiquetado visible de los *deep fakes* y los chatbots reduce el riesgo de engaño que probablemente surja una vez que el contenido generado por IA se difunda al público y reduce el riesgo de efectos distorsionadores perjudiciales en la formación de opiniones y creencias y el comportamiento del individuo. Por último, pueden ocurrir que nos encontremos con una combinación de técnicas, con el resultado de un impacto compuesto.

Finalmente, para que se aplique esta prohibición del apartado a, se debe tener como resultado que se haya causado o pueda ser razonablemente probable que sea cause un daño significativo. De un lado, describe unos tipos de daños que son: físicos, psicológicos, financiero y económicos. El daño físico abarca cualquier lesión o daño a la vida y a la salud de una persona, así como cualquier daño material a la propiedad, por ejemplo, un chatbot que promueve la autolesión a usuarios. El daño psicológico tiene importancia en los sistemas de IA que explotan vulnerabilidades cognitivas y emocionales, abarcando efectos adversos en la salud mental y el bienestar

psicológico y emocional de una persona, y son significativos al poder acumularse en el tiempo presentando consecuencias de gran impacto y prolongadas en el tiempo (como ejemplo, una aplicación de compañía de IA diseñada para emular patrones de habla, comportamientos y emociones humanas utiliza características antropomórficas y señales emocionales para influir en los sentimientos. El daño financiero y económico abarca unos diversos efectos adversos, como pueden ser: pérdida o exclusión financiera e inestabilidad económica (por ejemplo, chatbot que ofrece productos fraudulentos que termina causando daños financieros).

Además, el daño debe ser significativo, por lo que debe entender que ocasionará impactos adversos significativos sobre la salud física, psicológicas o los intereses financieros de personas. Deberán analizarse varios factores: (i) la gravedad del daño, que se refiere al grado de daño resultante o que es razonable que resulte del uso de un sistema de IA; (ii) contexto y efectos acumulativos, donde se debe incluir el contexto específico, el estado actual así como los efectos acumulativos de diversas acciones, (iii) la escala e intensidad; (iv) la vulnerabilidad de las personas afectas, entiendo que algunos grupos como las personas con discapacidad pueden ser más susceptibles a sufrir daños por estos sistemas; (v) duración y reversibilidad, los daños duraderos o irreversibles suelen considerarse significativos (por ejemplo, un daño significativo que es razonablemente probable que cause un sistema de IA incluye muertes o lesiones, o un impacto grave en la salud de personas. Por último, es precisa la existencia de una relación causal entre la técnica manipuladora o engañosa capaz de distorsionar el comportamiento de la persona y el posible daño significativo. En este sentido, se plantea una gran dificultad: la opacidad o falta de transparencia de los sistemas de IA[7], lo que pueden afectar a la prueba en estos casos.

Como medida preventiva, se proponen una serie de actuaciones diligentes a ser cumplidas por los proveedores e implementadores de sistemas de IA que empleen este tipo de técnicas: (i) transparencia y autonomía de la voluntad, de forma que proporcionan transparencia en el funcionamiento del sistema de IA, así como divulgaciones claras sobre sus capacidades y limitaciones; (ii) cumplimiento de la legislación aplicable pertinente, puesto que en muchos casos el cumplimiento de la misma permite mitigar los riesgos de daño; (iii) prácticas de última generación y estándares de la industria, puesto que puede ayudar a prevenir y mitigar daños no deseados. De otro lado, se indica que los daños y la distorsión del comportamiento de las personas que resultan de factores externos al sistema de IA y que no están bajo el control ni son razonablemente previsibles por el proveedor o el implementador para anticiparse y mitigar los riesgos no serían relevantes para la evaluación de si

[7] Recordemos que este tipo de sistemas son considerados de caja negra o "black box", de forma que es demasiado difícil conocer cual son los "inputs" que se han introducido en el sistema y que le llevan a tomar determinadas decisiones. Es por esto por lo que las diferentes propuestas normativas que han visto la luz en los últimos años están encaminadas a proporcionar diversos mecanismos de facilidad probatoria en tanto en cuanto en muchas ocasiones la víctima se encuentra con una *probatio* diabólica a la hora de solicitar una posible indemnización por los posibles daños que puedan ocasionar los sistemas de IA. En este sentido, véase EBERS, M., La utilización de agentes electrónicos inteligentes en el tráfico jurídico: ¿Necesitamos reglas especiales en el Derecho de la responsabilidad civil?, *Indret: Revista para el análisis del derecho*, julio 2016, pág. 6.

existe un vínculo causal plausible/razonablemente probable entre el comportamiento distorsionado de las personas que interactúan con el sistema y el daño significativo.

4. 2. Explotación dañina de vulnerabilidades

En cuanto a apartado b, para que se pueda prohibir el uso de sistemas de IA que promuevan la explotación perjudicial de vulnerabilidades, deben darse varias condiciones acumulativas: (i) la práctica debe constituir la comercialización, la puesta en servicio o la utilización de un sistema de IA; (ii) el sistema de IA debe explotar las vulnerabilidades debidas a la edad, la discapacidad o la situación socioeconómica; (iii) la explotación posibilitada por el sistema de IA debe tener el objetivo o el efecto de distorsionar materialmente el comportamiento de una persona o de un grupo de personas; (iv) la conducta distorsionada debe causar o ser razonablemente probable que cause un daño significativo a esa persona, a otra persona o a un grupo de personas.
En cuanto a la segunda condición, la ley de IA no especifica el concepto de vulnerabilidad, pero en estas directrices indica que puede entenderse como un amplio espectro de categorías, incluyendo las cognitivas, emocionales, físicas y otras formas de susceptibilidad que pueden afectar la capacidad de una persona o un grupo de personas para tomar decisiones informadas o influir de otro modo en su comportamiento. En cuanto a la explotación, debe entenderse como el uso objetivo de dichas vulnerabilidades de forma perjudicial para las personas explotadas u otras personas, y debe distinguirse claramente de las prácticas lícitas que no se ven afectadas por la prohibición. En cuanto a las categorías de debilidades debemos destacar cuáles son y porqué: (i) la edad, entendiendo que las personas menores de 18 años son especialmente susceptibles a la manipulación al encontrarse en una etapa de desarrollo (por ejemplo, un juego diseñado con IA que fomenta el uso compulsivo), (ii) la discapacidad, con el objetivo de impedir que los sistemas de IA exploten las limitaciones y debilidades cognitivas de los mismos (por ejemplo, los sistemas de IA pueden identificar a mujeres y niñas con discapacidad en línea con contenido sexualmente abusivo y dirigirse a ellas con prácticas de captación más efectivas); (iii) situación socioeconómica específica, pretendiendo proteger a aquellas personas que se encuentran en una situación socioeconómica especifica que haga a las personas afectadas más vulnerables a la explotación, y debiendo interpretarse específico como una condición jurídica o de pertenencia a un grupo social o económico vulnerable específico (por ejemplo un algoritmo predictivo de IA puede utilizarse para dirigir anuncios de productos financieros a personas que viven en códigos postales de bajos ingresos y están en una situación financiera desesperada).

Respecto a la tercera condición, la explotación de las vulnerabilidades examinadas anteriormente debe tener el objetivo o el efecto de distorsionar materialmente el comportamiento de una persona o grupo de ellas. No se requiere necesariamente intención, y la diferencia con el apartado a) es la necesidad que sí está presente en el mismo, pero que no está presente en el apartado b), puesto que las vulnerabilidades específicas de los niños reducen su capacidad para tomar decisiones informadas. En relación con la cuarta condición, el preciso que la distorsión del comportamiento de la persona o grupo de las mismas debe causar o ser

razonablemente probable que cause a esa u otra persona un daño significativo. Por ejemplo, en el caso de personas mayores pueden sufrir un deterior cognitivo y una alfabetización digital reducida, lo que los convierte en blancos perfectos de estafas.

Otra cuestión muy importante es distinguirlo de la persuasión lícita, que es aquella que opera dentro de los límites de la transparencia y el respeto a la autonomía individual, implicando presentar argumento o información de una manera que apela a la razón y a las emociones, pero que explica los objetivos y funcionamientos del sistema de IA, proporciona información relevante y precisa para garantizar una toma de decisiones informada y apoya la capacidad de la persona para evaluar la información y tomar decisiones libres y autónomas (por ejemplo, un sistema de IA que pretende ayudar a los usuarios a aprender un idioma extranjero mejor y más rápido mediante el despliegue de técnicas subliminales no es manipulador si opera de manera transparente y respeta la autonomía individual y la elección libre e informada del usuario de consentir el uso del sistema o no).

4. 3. Puntuación social

Los sistemas de IA que permiten prácticas de puntuación social tienen por objeto proteger la dignidad humana y otros derechos fundamentales como son el derecho a la no discriminación y a la igualdad. Para que aplique el artículo 5.1, apartado c, deben cumplirse simultáneamente varias condiciones: (i) la práctica debe constituir la comercialización, la puesta en servicio o la utilización de un sistema de IA; (ii) el sistema de IA debe estar destinado o utilizarse para la evaluación o clasificación de personas físicas o grupos de personas durante un período de tiempo determinado en función de: a) su comportamiento social, o (b) características personales o de personalidad conocidas, inferidas o predichas; (iii) la puntuación social creada con la asistencia del sistema de IA debe conducir o ser capaz de conducir a un trato perjudicial o desfavorable de personas o grupos en uno o más de los siguientes escenarios: (a) en contextos sociales no relacionados con aquellos en los que se generaron o recopilaron originalmente los datos, y/o b) un trato injustificado o desproporcionado en relación con su comportamiento social o su gravedad.
En cuanto a la segunda condición, mientras el término "evaluación" siguiere la participación de alguna forma de evaluación o juicio sobre una persona o grupo de persona, el término "clasificación" parece indicar que no precisa una evaluación. Por tanto, el término clasificación es más amplio, al abarcar otro tipo de personas o grupos de personas en función de características como su edad, sexo, o altura. Además, exige que la evaluación o clasificación se base en datos que abarquen más de "un cierto período de tiempo", lo que sugiere que la evaluación no debe limitarse a una calificación única o inmediata con datos o comportamientos de un contexto individual muy específico (por ejemplo, una autoridad de migración y asilo implementa un sistema de vigilancia parcialmente automatizado en campamentos de refugiados, basado en diversas infraestructuras de vigilancia, como cámaras y sensores de movimiento). En cuanto al tercer requisito, las prácticas de evaluación y clasificación deben basarse en el procesamiento de datos habilitados por IA en relación con el comportamiento social de individuos (que puede incluir acciones, comportamientos, hábitos, interacciones dentro de la sociedad, etc., y generalmente

cubre puntos de datos relacionados con el comportamiento de múltiples fuentes) y sus características personales y de personalidad conocidas (pueden incluir información diversa sobre una persona, por ejemplo, sexo, orientación sexual o características sexuales, género, identidad de género, raza, etnia, situación familiar, domicilio, ingresos, miembros del hogar, profesión, empleo u otra situación legal, rendimiento laboral, situación económica, entre otras). "Las características conocidas" se basan en información proporcionada al sistema de IA como entrada, y que en la mayoría de los casos es información verificable; "las características inferidas" se basan en información que se ha inferido de otra información, y la inferencia generalmente la realiza un sistema de IA y "las características previstas" son aquellos que se estiman basándose en patrones con una precisión inferior al 100 %.

En cuanto al cuarto requisito, debe mediar una relación causal entre la puntuación social y el tratamiento. Por tanto, el tratamiento debe ser consecuencia de la puntuación. Ahora bien, no se exige que la evaluación realizada por el sistema de IA sea la única causa del trata perjudicial o desfavorable. Es importante destacar que una puntuación puede dar lugar a un tratamiento perjudicial o desfavorable incluso si es producida por una o más organizaciones diferentes de la que utiliza la puntuación (por ejemplo, una autoridad pública puede obtener una puntuación para la evaluación de solvencia de una persona física elaborada por otra empresa especializada en evaluaciones de solvencia y riesgo, que se basa en información sobre las personas y su comportamiento procedente de diversas fuentes). En cuanto a la condición final, es que el uso de la puntuación debe resultar en un trato perjudicial (lo que significa que la persona o grupo de ellas sufran ciertos daños y perjuicios a causa del mismo) o desfavorable (lo que significa que, como resultado de la puntuación, la persona o el grupo de persona reciban un trato menos favorable respecto a otras).

En cuanto al primer escenario, hablamos de un trato perjudicial o desfavorable en contextos sociales no relacionados. Para encontrarnos ante esta situación, el trato perjudicial o desfavorable resultante de la puntuación debe ocurrir en un contexto social no relacionado a los contextos en los que se generaron o recopilaron originalmente los datos. En la mayoría de los casos, ocurre en contra de las expectativas razonables de personas y violando la legislación de la Unión en materia de protección de datos. Como ejemplo, podría ser una situación en la que una agencia de bienestar social utiliza un sistema de IA para estimar la probabilidad de fraude por parte de los beneficiarios de asignaciones familiares que se basa en características recopiladas o inferidas de contextos sociales sin conexión o relevancia aparente para la evaluación del fraude, como tener un cónyuge de cierta nacionalidad u origen étnico, tener una conexión a Internet, comportamiento en plataformas sociales o desempeño en el lugar de trabajo, etc.

Respecto al segundo escenario, sería un trato desfavorable o perjudicial pero desproporcionado al comportamiento social. Esto supone que la gravedad del impacto y la interferencia con los derechos fundamentales de la persona afectada resultante de la puntuación social, en comparación con la gravedad de su comportamiento social, deben determinar si dicho trato es desproporcionado para el fin legítimo

perseguido, teniendo en cuenta el principio general de proporcionalidad. Para determinarlo se requiere una evaluación caso por caso, considerando varios factores como circunstancias pertinentes del caso, consideraciones éticas generales, principios de equidad, justicia social, entre otros. Como ejemplo, una agencia pública que utiliza un sistema de IA para perfilar a las familias y detectar tempranamente a niños en riesgo, basándose en criterios como la salud mental y el desempleo parental, así como en información sobre el comportamiento social de los padres derivada de múltiples contextos. Es importante destacar que independiente de si son proporcionados o utilizados por personas públicas o privadas, si bien la puntuación en el sector público puede tener consecuencias muy significativas para las personas debido al desequilibrio de poder y la dependencia de los servicios públicos, también pueden producirse consecuencias igualmente perjudiciales en el sector privado, donde las empresas y otras entidades también realizan cada vez más prácticas de puntuación. De otro lado, quedan fuera de alcance de la prohibición los casos en los que la evaluación no se basa en características personales o de personalidad o comportamiento social de individuos, incluso si en algunos casos los individuos pueden verse afectados indirectamente por la puntuación

4.4. Criminalidad individual: evaluación y predicción del riesgo de delitos

En el artículo 5.1., apartado d, la Ley de IA prohíbe que los sistemas de IA evalúen o predigan el riesgo de que una persona física cometa un delito penal basándose únicamente en la elaboración de perfiles o la evaluación de rasgos y características de personalidad. Su justificación radica en que las personas físicas deben ser juzgadas por su comportamiento real y no por el comportamiento previsto por la IA basándose únicamente en su perfil, rasgos de personalidad o característica. Para que se aplique la misma deben darse varias condiciones acumulativas: (i)la práctica debe constituir la comercialización, la puesta en servicio para este fin específico o la utilización de un sistema de IA; (ii) el sistema de IA debe realizar evaluaciones de riesgos que evalúen o predigan el riesgo de que una persona física cometa un delito penal; (iii) la evaluación del riesgo o la predicción debe basarse únicamente en uno o ambos de los siguientes factores: a) la elaboración de perfiles de una persona física, (b) evaluar los rasgos y características de personalidad de una persona física. En cuanto al segundo requisito, la predicción de delitos se refiere en general a una variedad de tecnologías avanzadas en IA y métodos analíticos aplicados a gran cantidad de datos, a menudos históricos, que, en combinación con teorías criminológicas, se utilizan para pronosticar el crimen como base para informar a la policía y las estrategias de aplicación de la ley y acciones para combatir, controlar y prevenir el crimen. Los sistemas de IA para la predicción de delitos identifican patrones en datos históricos, asociando indicadores con la probabilidad de ocurrencia de un delito y generando puntuaciones de riesgo como resultados predictivos. Sin embargo, este uso de datos históricos sobre delitos cometidos para predecir el comportamiento futuro de otras personas puede perpetuar o incluso reforzar sesgos[8],

[8] Especial atención merecen la aparición de los mismos en el ámbito sanitario. En este sentido, véase GIL MEMBRADO, C., *Riesgos del uso de algoritmos en el diagnóstico y en la investigación biomédica,* ed. Reus, Madrid, 2023, págs. 33-34.

y puede dar lugar a que se pasen por alto circunstancias individuales cruciales cuando estas no forman parte del conjunto de datos ni se consideran en los algoritmos con los que opera el sistema de IA en particular[9]. En cuanto a la tercera condición, es que la evaluación de riesgos debe basarse únicamente en la elaboración del perfil de la persona o b) en la evaluación de sus rasgos y características de personalidad. La elaboración de perfiles de personas físicas se refiere a la construcción y aplicación de un perfil descriptivo para un grupo determinado, por ejemplo, categorías de autores de delitos (por ejemplo, terroristas) construidas a partir de datos históricos sobre delitos cometidos previamente por otros. Respecto a la evaluación de rasgos y características de personalidad, constituyen una categoría amplia de características relacionadas con una persona física en particular, para la cual no existe una taxonomía generalmente aceptada. El considerando 42 de la Ley de Inteligencia Artificial ofrece ejemplos de rasgos y características de personalidad que pueden evaluarse para predecir el riesgo de que una persona cometa un delito, como la nacionalidad, el lugar de nacimiento, entre otros. Por ejemplo, una autoridad policial utiliza un sistema de IA para predecir la conducta delictiva en delitos como el terrorismo, basándose únicamente en la edad, la nacionalidad, la dirección, el tipo de vehículo y el estado civil de las personas. Con este sistema, se considera que las personas son más propensas a cometer delitos futuros que aún no han cometido, basándose únicamente en sus características personales. Cabe suponer que este sistema está prohibido por el artículo 5(1)(d) de la Ley de IA.

Es necesario destacar que la prohibición no si aplicará a los sistemas de IA utilizados para respaldar la evaluación humana de la participación de una persona en una actividad delictiva, puesto que se basa en hechos objetivos y verificables directamente vinculados a ella. Además, el concepto de "intervención humana" ha sido objeto de jurisprudencia del TJUE, en particular en el contexto de la toma de decisiones exclusivamente automatizada que predice el riesgo de que los pasajeros aéreos se vean involucrados en delitos graves. El TJUE interpretó la norma de la Directiva (UE) 2016/681 ("Directiva PNR") que prohíbe las decisiones judiciales adversas basadas únicamente en el procesamiento automatizado y exigió evaluación

[9] En este sentido, cabe mencionar el sistema Compas, que fue una herramienta de IA través de la cual se evalúa el riesgo de reiteración delictiva de los delincuentes en algunas jurisdicciones de Estados Unidos. El resultado de esta investigación se admitió como prueba por los tribunales estadounidenses, pero solo se admitió cuando se utilizaba de manera complementarias a otras pruebas. Esto es así porque se descubrió que COMPAS poseía un evidente sesgo racista fruto de un uso totalmente inadecuado de los datos estadísticos de criminalidad y sus resultados no eran diferentes a aquellos que podrían obtener un mero observador sin experiencia. También destaca El Protocolo S.A.R.A. (Spoke Abuse Risk Assessment) es diseñado en 1995 con la finalidad de valorar el riesgo de sufrir violencia física y/o sexual intrafamiliar. Se utiliza en la actualidad, desde 2017, por los Mossos d'Esquadra para valorar el riesgo de la víctima de la comisión de nuevos hechos delictivos. Consta de 20 parámetro que se pueden dividir en 5 tipos de factores: i) historial delictivo distinto a la relación de pareja; ii) historial psicosocial; iii) historial de violencia de pareja; iv) delito actual; v) otros. Y por último, el protocolo RisCanvi (del catalán Risc-Canvi, Riesgo-Cambio) se implementó en la Administración Penitenciaria catalana en 2018 con la finalidad de obtener una herramienta objetiva que ayudara a prever y evitar los riesgos derivados de la salida del centro a los reclusos. Se evalúan un total de 5 riesgos: violencia autodirigida, violencia institucional, reincidencia general, reincidencia violenta y quebrantamientos de condena. En este sentido, véase ÁLVAREZ SÁEZ K., La actual utilización de la inteligencia artificial para evaluar el riesgo de reincidencia, MARTÍN OSTOS, J (dir.), *Inteligencia artificial y derecho*, ed. Astigi, Sevilla, 2024, págs. 19-29.

humana individual y revisar si hay coincidencias positivas por medios no automatizados para identificar falsos positivos y garantizar resultados no discriminatorios. Por último, no solo las actividades de las autoridades policiales están comprendidas en esta prohibición, sino que las actividades de las entidades privadas también: cuando la ley confía a agentes privados el ejercicio de la autoridad pública y de los poderes públicos para la prevención, investigación, detección o enjuiciamiento de delitos o la ejecución de sanciones penales. Además, la prohibición puede aplicarse a entidades privadas que evalúen o predigan el riesgo de que una persona cometa un delito cuando esto sea objetivamente necesario para el cumplimiento de una obligación legal a la que esté sujeto ese operador privado de evaluar o predecir el riesgo de que las personas cometan delitos penales específicos (por ejemplo, en caso de blanqueo de dinero o financiación del terrorismo).

Como en los casos anteriores, hay supuestos a los que la prohibición no alcanza. Las predicciones de delitos basadas en la ubicación, geoespaciales, el lugar del delito, o la probabilidad de que en esas zonas se cometa un delito no implicarían un tipo de vigilancia que evalúan a un individuo específico. Como ejemplo, una autoridad aduanera utiliza herramientas de análisis de riesgos de IA para predecir la probabilidad de ubicación de narcóticos o mercancías ilícitas, por ejemplo, basándose en rutas de tráfico conocidas.

4. 5. Raspado no dirigido para desarrollar bases de datos de reconocimiento facial

El artículo 5.1., apartado e, prohíbe la puesta en servicio para este fin específico o el uso de sistema de IA que crean o amplían bases de datos de reconocimiento facial mediante la extracción no dirigida de imágenes fáciles de Internet o grabaciones de CCTV (que son un sistema de cámaras de seguridad que permiten la captura y grabación de imágenes captadas de forma remota). Para que se aplique la misma deben cumplirse varias condiciones acumulativas: (i) la práctica debe constituir la comercialización, la puesta en servicio para este fin específico o la utilización de un sistema de IA; (ii) con el fin de crear o ampliar bases de datos de reconocimiento facial; (iii) los medios para poblar la base de datos son a través de herramientas de IA para el raspado no dirigido; y (iv) las fuentes de las imágenes proceden de Internet o de grabaciones de CCTV.
Respecto al segundo requisito, "bases de datos" se refiere a cualquier recopilación de datos o información especialmente organizada para su rápida búsqueda y recuperación por ordenador. Una base de datos de reconocimiento facial es aquella capaz de cotejar un rostro humano de una imagen digital o un fotograma de vídeo con una base de datos de rostros, comparándolo con las imágenes de la base de datos y determinando si existe una coincidencia probable entre ambos. En cuanto al tercer requisito, el "scraping" se refiere típicamente al uso de rastreadores web, bots u otros medios para extraer automáticamente datos o contenido de diferentes fuentes, como cámaras de seguridad, sitios web o redes sociales. El término "sin objetivo" se refiere a una técnica que funciona como una «aspiradora», absorbiendo la mayor cantidad posible de datos e información, sin dirigirse específicamente a los

sujetos de la extracción (ahora bien, la recopilación selectiva de imágenes centrada en un grupo de víctimas, mediante el uso de rastreadores para seleccionar imágenes de víctimas que los traficantes de personas publican o anuncian en redes sociales, no está amparada por la prohibición). En cuanto al cuarto requisito, la fuente de las imágenes faciales puede ser internet o grabaciones de CCTV. En este sentido, se aclara que el hecho de que una persona haya publicado imágenes fáciles de sí misma en una plataforma de redes sociales no significa que consienta que las mismas se incluyan en una base de datos de reconocimiento facial. No se encuentran dentro del alcance de esta prohibición la extracción de datos biométricos distintos de las imágenes faciales.

4. 6. Reconocimiento de emociones

El artículo 5.1., apartado f, prohíbe que los sistemas de IA infieran las emociones de una persona física en el ámbito laboral y educativo, exceptuando el uso del sistema cuando esté previsto por razones médicas o de seguridad. En cuanto a su justificación, el reconocimiento de emociones puede conducir a resultados discriminatorios y puede ser intrusivo para los derechos y libertades de las personas afectadas, entre otros muchos motivos. Para que se aplique la misma, es preciso que se den una serie de condiciones acumulativas: (i) la práctica debe constituir la comercialización, la puesta en servicio para este fin específico o la utilización de un sistema de IA; (ii) sistema de IA para inferir emociones; (iii) en el ámbito del lugar de trabajo o de las instituciones de educación y formación; y (iv) quedan excluidos de la prohibición los sistemas de IA destinados a fines médicos o de seguridad.
En cuanto al segundo criterio, la identificación se produce cuando el procesamiento de datos biométricos (por ejemplo, de la voz o una expresión facial) de una persona física permite comparar e identificar directamente una emoción con una previamente programada en el sistema de reconocimiento de emociones. En cuanto al concepto "emociones" o intenciones, debe entenderse en sentido amplio y no hacerlo de forma restrictiva. No se refiere a actitudes, y además se aclara que las emociones o intenciones no incluyen estados físicos, como el dolor o la fatiga (por ejemplo, la observación de que una persona está sonriendo no es un reconocimiento de emociones). En cuanto al tercer criterio, esta prohibición se lomita a los sistemas de reconocimiento de emociones en los ámbitos del lugar de trabajo e instituciones educativas. El concepto de "lugar de trabajo" se debe entender de forma amplia, y se refiere a cualquier espacio físico o virtual específico donde las personas físicas desempeñan tareas y responsabilidades asignadas por su empleador o por la organización a la que están afiliadas, por ejemplo, en el caso del trabajo por cuenta propia (por ejemplo, está prohibido el uso de cámaras web y sistemas de reconocimiento de voz por parte de un centro de llamadas para rastrear las emociones de sus empleados, como la ira). En cuanto a instituciones educativas, se incluyen tanto las públicas como las privadas, y algunos requisitos para identificarlas son: suelen estar acreditadas o autorizadas por las autoridades educativas nacionales competentes o autoridades equivalentes y además las instituciones educativas pueden otorgar un certificado. Ahora bien, hay que estudiar el caso. Por ejemplo, la

Ley de IA no prohíbe una aplicación basada en IA que utilice reconocimiento de emociones para aprender un idioma en línea fuera de una institución educativa.

Por último, en cuanto al cuarto requisito, es el que establece la excepción en caso de razón médica y de seguridad. Tal y como se indica, debe entender por "usos terapéuticos" el uso de productos sanitarios con marcado CE (sin comprender la excepción el uso de sistemas de reconocimiento de emociones para detectar aspectos generales del bienestar). Por ejemplo, el reconocimiento de emociones se puede implementar por razones médicas para ayudar a empleados o estudiantes con autismo y mejorar la accesibilidad para personas ciegas o sordas.

4. 7. Categorización biométrica para ciertas características "sensibles"

El artículo 5.1., apartado g, prohíbe los sistemas de categorización biométrica que categorizan individualmente a personas físicas basándose en sus datos biométricos para deducir o inferir su raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, vida sexual u orientación sexual. No se incluyen dentro de la prohibición los datos biométricos obtenidos de conformidad con la legislación de la Unión o nacional, que puedan utilizarse, por ejemplo, con fines policiales. En cuanto a su justificación, este tipo de mecanismos puede dar lugar a situaciones de trato injusto y discriminatorio, como cuando se deniega un servicio por considerarse que alguien pertenece a una determinada raza. Para que se aplique la prohibición, deben darse las siguientes características: (i) la práctica debe constituir la comercialización, la puesta en servicio para este fin específico o la utilización de un sistema de IA; (ii) el sistema deberá ser un sistema de categorización biométrica; (iii) las personas individuales deben ser categorizadas; (iv) sobre la base de sus datos biométricos; (v) deducir o inferir su raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, vida sexual u orientación sexual.
En cuanto a la segunda condición, la categorización de un individuo mediante un sistema biométrico consiste típicamente en determinar si sus datos biométricos pertenecen a un grupo con alguna característica predefinida. La categorización biométrica puede basarse en categorías de características físicas (por ejemplo, rasgos faciales y forma, color de piel) que asignan a las personas a categorías específicas. Algunas de estas categorías pueden ser de naturaleza «sensible» o estar protegidas por la legislación de la Unión contra la discriminación, como la raza. Sin embargo, la categorización biométrica también puede basarse en el ADN o en aspectos conductuales, como el análisis de pulsaciones de teclas o la forma de andar de una persona. Hay algunos usos que sí están permitidos, como por ejemplo los filtros que categorizan las características faciales o corporales utilizados en los mercados en línea para permitir que un consumidor obtenga una vista previa de un producto. En cuanto a la tercera condición, la utilización de datos biométricos para la categorización de personas físicas es un elemento esencial para que se aplique la prohibición. Además, para que la prohibición sea aplicable, las personas físicas deben estar categorizadas individualmente. En cuanto a la cuarta condición, la Ley prohíbe únicamente los sistemas de categorización biométrica que tengan como objetivo deducir o inferir un número limitado de características sensibles: raza, opiniones

políticas, afiliación sindical, creencias religiosas o filosóficas, vida sexual u orientación sexual. Por ejemplo, entraría dentro de la prohibición un sistema de categorización biométrica que pretende ser capaz de deducir la orientación religiosa de un individuo a partir de sus tatuajes o rostros.

Por último, hay muchos usos que se encuentran fuera del alcance de esta prohibición. La prohibición no se aplica a los sistemas de IA que participan en el etiquetado o filtrado de conjuntos de datos biométricos adquiridos legalmente, como imágenes, basados en datos biométricos, incluso en el ámbito de las fuerzas de seguridad. Por ejemplo, La categorización de pacientes mediante imágenes según el color de su piel o de sus ojos puede ser importante para el diagnóstico médico, por ejemplo, el diagnóstico de cáncer.

4. 8. Sistemas de identificación biométrica a tiempo real (RBI) para fines de aplicación de la ley

El artículo 5.1., apartado h, prohíbe el uso de sistemas de información basada en la información (RBI) en tiempo real es espacios de acceso público con fines de aplicación de la ley (sujeto a excepciones limitadas que establece la ley). En cuanto a la justificación, este tipo de sistemas pueden generar resultados sesgados y conllevar efectos discriminatorios, así como la inmediatez del impacto y las limitades para realizar comprobaciones o correcciones adicionales en relación con el uso de dichos sistemas que operan en tiempo real, conllevan mayores riesgos para los derechos y libertades de las persones afectadas. Ahora bien, en algunos casos el uso de estos sistemas sí que está justificado en favor de la protección del interés público. Por tanto, el uso de estos sistemas están prohibido, a no ser que se quiera lograr uno de estos objetivos: i) la búsqueda selectiva de víctimas específicas de secuestro y trata de seres humanos o la explotación sexual de seres humanos, así como la búsqueda de personas desaparecidas; ii) la prevención de una amenaza específica, sustancial e inminente a la vida o a la integridad física la seguridad de las personas físicas o una amenaza real y presente o real y previsible de un ataque terrorista; iii) la localización o identificación de una persona sospechosa de haber cometido un delito. Para que se aplique la condición se deben cumplir varias condiciones: (i) el sistema de IA debe ser un sistema RBI; (ii) la actividad consiste en el «uso» de ese sistema; (iii)en 'tiempo real', (iv) en espacios de acceso público, y (v) para fines de cumplimiento de la ley. En cuanto a la primera condición, un sistema RBI es un sistema de IA cuyo fin es identificar personas físicas, sin su participación activa, normalmente a distancia, mediante la comparación de los datos biométricos de una persona con los datos biométricos contenidos en una base de datos de referencia. En cuanto a la lejanía, el uso de sistemas biométricos para confirmar la identidad de una persona física con el único fin de acceder a un servicio, desbloquear un dispositivo o acceder de forma segura a instalaciones queda excluido del concepto de «remoto». En cuanto a base de datos de referencia, la identificación no es posible sin una base de datos de referencia que contenga datos biométricos con fines de comparación. Respecto a tiempo real, significa que el sistema captura y procesa datos biométricos 'de manera instantánea, casi instantánea o, en cualquier caso, sin demora

significativa'. En cuanto a "en espacios de acceso público" se refiere a cualquier espacio físico de propiedad pública o privada accesible a un número indeterminado de personas físicas, independientemente de que se apliquen determinadas condiciones de acceso y de las posibles restricciones de aforo. Por último, la prohibición se aplica al uso de los sistemas del RBI para fines de aplicación de la ley, independientemente de la entidad, autoridad u organismo que lleve a cabo las actividades de aplicación de la ley. Los fines de aplicación de la ley comprenden la investigación, la detección y el enjuiciamiento de delitos. También abarcan actividades relacionadas con la prevención de delitos, incluyendo la protección y la prevención de amenazas a la seguridad pública, antes de que se cometa el delito. Por ejemplo, la policía puede adoptar medidas coercitivas en manifestaciones, grandes acontecimientos deportivos o disturbios en el contexto de la prevención del delito.

De otro lado, se articulan tres excepciones. Se permite su uso para para la búsqueda específica de víctimas de secuestro, trata de seres humanos o explotación sexual de seres humanos, así como para la búsqueda de personas desaparecidas; prevención de amenazas inminentes a la vida o ataques terroristas; y localización e identificación de sospechosos de ciertos delitos. En primer lugar, en el primer supuesto la Ley de IA introduce esta cuestión en aras a ayudar a las autoridades policiales a buscar víctimas de esos delitos graves (por ejemplo, hay un niño secuestrado y hay indicios concretos de que el secuestrador pretende llevar al niño a otro lugar en coche, la policía puede utilizar este tipo de sistemas en tiempo real para buscar al mismo). El segundo escenario que se contempla en este primer supuesto es el de la búsqueda de personas desaparecidas. Hace una distinción sobre menores y adultos desaparecidos, puesto que en ocasiones la desaparición voluntaria de un adulto no siempre da lugar a búsqueda (tienen derecho a desaparecer). Para apreciar la necesidad de su utilización, la búsqueda podría estar vinculada a la situación jurídica de la persona ("bajo curatela"), su estado de salud (enfermedad mental), la existencia de una cota suicida o la partida sin pertenencias personales. En algunos Estados miembros la búsqueda puede realizarse a través de un procedimiento administrativo y no con fines policiales (por ejemplo, cuando una persona vulnerable está desaparecida, pero no existe sospecha de delito ni ningún otro fin policial, el uso de sistemas de información basada en la investigación (RBI) en tiempo real para buscar a esa persona no se considerará que sea para fines de aplicación de la ley y, por lo tanto, estaría sujeto a las reglas para dicho uso según el RGPD.

En segundo lugar, en aquellas situaciones en la que se quiera prever amenazas inminentes a la vida o ataques terroristas. Los criterios relativos a la amenaza de vida para permitir el uso de sistemas de información basada en la responsabilidad (RBI) en tiempo real en espacios de acceso públicos son los siguientes: (i) existencia de una amenaza específica; (ii) que sea sustancial; (iii) que sea inminente para la vida o la integridad física de personas físicas. La amenaza no tiene que referirse a personas o grupos identificados, sino que puede referirse a personas físicas en general. Además, aclara que también se incluye una amenaza inminente a la infraestructura crítica *"cuando la perturbación o destrucción de dicha infraestructura crítica dé lugar a una amenaza inminente para la vida, la integridad física o la seguridad de una persona, incluido un daño grave al suministro de suministros*

básicos a la población o al ejercicio de las funciones esenciales del Estado." (por ejemplo, una grave perturbación y destrucción de una central eléctrica). De cualquier manera, lo que constituye una amenaza inminente para la vida o la seguridad física de las personas físicas se define y evalúa en última instancia a nivel del Estado miembro con base en sus leyes nacionales, de conformidad con el Derecho de la Unión Europea. Un ejemplo sería un policía que recibe información de que un exalumno planea un atentado mortal en su antigua universidad para vengarse de varios excompañeros. El segundo supuesto es una amenaza real y presente o real y previsible de un ataque terrorista. En este caso, el nivel de amenaza terrorista se define a nivel nacional por lo que puede variar de un Estado miembro a otro. Según el TJUE, en estos contextos, «una amenaza a la seguridad nacional debe ser real y presente, o al menos previsible, lo que presupone la concurrencia de circunstancias suficientemente concretas»[10]. También es importante destacar que no se especifica el uso de la información basada en la evidencia en tiempo real para localizar o identificar a una persona concreta, sino que su finalidad es procurar la prevención de una amenaza específica.

En tercer lugar, el referido artículo permite el uso en tiempo real de RBI en espacios de acceso público para la localización e identificación de una persona sospechosa de haber cometido un delito penal, con el fin de llevar a cabo una investigación o un enjuiciamiento penal contemplados en el Anexo II de la Ley y punibles en Estados miembros (terrorismo, la trata de seres humanos, la explotación sexual de niños y pornografía infantil, tráfico ilícito de estupefacientes, de armas, entre otros). Se abre entonces un abanico de concepto que analizar: (i) localización e identificación: un estado miembro puede autorizar este sistema para localizar e identificar a un sospechoso de un delito penal a fin de llevar a cabo una investigación penal, procesar a esa persona por el delito cometido o ejecutar una sentencia vigente); (ii) sospechosos y perpetradores: un sospechoso es una persona sobre la que existen motivos fundados para creer que ha cometido un delito y se han recabado pruebas suficientes de su participación en el mismo; mientras que un autor es una persona acusado o condenada por un delito; (iii) lista de delitos graves: aunque todos los delitos enumerados en el anexo II pueden dar lugar a la emisión de una orden de detención europea («ODE») contra un sospechoso o autor, el uso de información en tiempo real para que el RBI localice e identifique a un sospechoso de uno de estos delitos penales graves no es necesario que se haya emitido una orden de detención europea. Además, para utilizar la información de base sobre la investigación en tiempo real con este fin, el delito penal en cuestión debe ser punible en el Estado miembro de que se trate con una pena o una orden de detención por un período máximo de al menos cuatro años. Como ejemplo, la RBI puede utilizarse en casos de tráfico ilegal de drogas. Sin embargo, los delitos sexuales no figuran en la lista de delitos, a menos que estén relacionados con la explotación sexual infantil, material de abuso sexual infantil o violación.

[10] Sentencia del Tribunal de Justicia de 20 de septiembre de 2022, SpaceNet, C-793/19 (Asuntos acumulados C-793/19, C-794/19), ECLI:EU:C:2022:702, apartado 93.

En todos los casos citados anteriormente, es de especial mención indicar que los proveedores y los implementadores deben evaluar cuidadosamente si existe otra legislación nacional y de la Unión que tenga aplicación en estos casos, existiendo una legislación más especifica que regule estrictamente los mismos.

V. SALVAGUARDIAS Y CONDICIONES PARA LAS EXCEPCIONES

5. 1. Evaluación de Impacto sobre la Protección de Datos y Registro de los Sistemas Autorizados del RBI

El uso de sistemas RBI en tiempo real para uno de los objetivos enumerados en el artículo 5.1, apartado h, está sujeta a ciertas garantías y condiciones, que se detallan en los artículos 5.2. a 5.7. de la Ley de IA. La primera condición indica que este uso solo está permitido para confirmar la identidad de la persona específicamente. Esta expresión debe entenderse en el sentido de que el uso de la información basada en la investigación (RBI) en tiempo real solo puede iniciarse para buscar a personas específicas respecto de las cuales las autoridades policiales tengan motivos para creer o estén informadas de que son víctimas de los delitos enumerados en el artículo 5, apartado 1. La segunda condición advierte que, antes de utilizar el sistema, debe evaluarse la naturaleza de la situación que da lugar a su posible uso, en particular la gravedad, la probabilidad y la magnitud del daño que se le causaría a las personas físicas, a la sociedad y a los fines policiales si no se utilizara el sistema en relación con las consecuencias de su uso para los derechos y libertades de las personas afectadas, en particular la gravedad, probabilidad y magnitud de dichas consecuencias. La tercera condición alude a que el uso del RBI en tiempo real debe estar claramente limitado en cuanto a su alcance geográfico, duración y destinatario, lo que garantiza que el sistema RBI solo se utilice cuando sea estrictamente necesario. En cuarto lugar, antes de la implementación, la autoridad policial que implemente el sistema RBI en tiempo real debe haber realizado una Evaluación de Impacto sobre los Derechos Fundamentales, y registrado el sistema en la base de datos de la UE (excepto en un caso debidamente justificado).

Una Evaluación de Impacto sobre la Protección de Datos (EIPD) es un nuevo tipo de evaluación de impacto que busca identificar el impacto que ciertos sistemas de IA de alto riesgo, incluidos los sistemas de IDR, pueden tener sobre los derechos fundamentales, siendo una herramienta de rendición de cuentas. En la misma se debe incluir una en primer lugar, descripción del uso de RBI y los procesos del implementador para el uso, junto con una serie de datos: (i) el nombre del implementador, (ii) los fines de aplicación de la ley para los cuales se utilizará el sistema RBI en tiempo real; (iii) la descripción de la base de datos de referencia con la que se comparará la identificación biométrica, incluidas las fuentes de los datos biométricos (imágenes faciales, muestras de voz, etc.) que se utilizarán; (iv) la descripción de la tecnología subyacente al sistema para explicar su funcionamiento (haciendo referencia a la documentación disponible proporcionada por el proveedor y su nombre); (v) la base jurídica sobre la que se implementará el RBI en tiempo real. En segundo lugar, el período y la frecuencia de uso, de forma que ada uso

individual de un sistema RBI en tiempo real para una de las excepciones permitidas debe ser autorizado previamente a su implementación por una autoridad judicial u otra autoridad independiente. En tercer lugar, las categorías de personas y grupos afectados por el sistema, de forma que se debería distinguir entre: (i) el individuo objetivo, que puede ser víctima de un delito, autor o sospechoso; (ii) las personas físicas cuyos datos biométricos consten en la base de datos de referencia, y(iii) las categorías de personas que están presentes en las zonas circundantes donde se desplegará el sistema RBI. En cuarto lugar, los riesgos específicos de daño a las personas afectadas, así como los derechos afectados por el uso de RBI en tiempo real en espacios de acceso público (por ejemplo, el derecho a la vida privada y familiar, incluida la expectativa razonable de las personas de mantener el anonimato en los espacios públicos; el derecho a la protección de datos, ya que los sistemas del RBI se basan en el procesamiento de datos biométricos y otros datos personales, entre otros). En quinto lugar, las medidas de supervisión humana que sean precisas, pues tal y como hemos indicado anteriormente, ninguna decisión que afecte negativamente a una persona podrá tomarse basándose únicamente en los resultados del sistema RBI en tiempo real. En sexto lugar, las medidas de mitigación de riesgos, puesto que el implementador debe indicar las medidas de reparación en caso de que se materialice un riesgo, incluidos los procedimientos de gobernanza y los mecanismos de queja.

En segundo lugar, la Ley de IA obliga al implementador de un sistema de información basada en información RBI en tiempo real utilizado en espacio con fines policiales a registrar este sistema en la base de datos de la UE. Pero, en casos de emergencia justificada, el despliegue puede comenzar antes de ese registro, siempre que la autoridad policial registre el sistema sin perpetuarse en el tiempo. Tal y como se indica en la Ley, el reiterado no puede deberse a una acción deliberada. Por ejemplo, solicitar a las autoridades policiales que registren el sistema RBI dentro de las 24 horas siguientes a su uso podría considerarse una demora razonable cuando el sistema se implementó en una situación de amenaza inminente a la vida, como en el escenario de un tirador real.

5. 2. Necesidad de una autorización previa

Tal y como se establece en el artículo 5.3. cada uso individual de un sistema de RBI ten tiempo precisa una autorización previa, prohibiendo la toma decisiones automatizada basada únicamente en el resultado de dicho sistema, puesto que produce un resultado jurídico adverso. La autoridad judicial competente o una autoridad administrativa independiente cuya decisión sea vinculante concederá la autorización únicamente si, basándose en pruebas objetivas o indicaciones claras que se le presenten, está convencida de que el uso del sistema de identificación biométrica remota «en tiempo real» en cuestión es necesario y proporcionado para alcanzar uno de los objetivos especificados en el apartado 1. En cuanto al objetivo de exigir una autorización precia para cualquier uso está se encuentra en estudiar y analizar si cualquier uso previsto de dicho sistema es necesario y proporcionado para cumplir los objetivos indicados (búsqueda específica de víctimas específicas, la prevención de amenazas específicas o la localización o identificación de los

delincuentes). En cuanto a la excepción en caso de urgencia, debe estar debidamente justificado. Debe entenderse urgencia como "*situaciones en las que la necesidad de utilizar los sistemas en cuestión es tal que resulta efectiva y objetivamente imposible obtener una autorización antes de comenzar".* En estos casos el sistema de IA debe restringirse a lo mínimo absoluto necesario y estar sujeto a los condiciones y garantías adecuadas.
En cuanto a la solicitud previa, debe realizarse por la autoridad competente (el implementador), entendiéndose por autoridad policial *"cualquier otro organismo o entidad a quien la legislación del Estado miembro haya encomendado el ejercicio de la autoridad pública y de los poderes públicos a efectos de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de sanciones penales, incluida la protección y la prevención frente a amenazas a la seguridad pública» se considera una autoridad policial",* y también podría ser la autoridad competente responsable de presentar la solicitud de autorización previa. En cuanto a los usos, se precisa la indicada autorización para el uso de estos sistemas en espacios de acceso público, y para fines de aplicación de la ley, incluso si los sistemas son operados por otras partes en nombre de autoridades policiales. Respecto al momento, debe hacerse con cada uso, de forma que el momento decisivo para obtener dicha autorización no es el momento previo a la instalación de los sistemas de información basada en la información en tiempo real (RBI), sino cada uso concreto de los mismos. Además, la solicitud debe estar motivada, de forma que la misma debe estar razonada y fundamentada.

De otro lado, la autorización sólo podrá ser concedida por una autoridad judicial o administrativa independiente cuya decisión sea vinculante y la misma deberá dirigirse a la autoridad competente según el Derecho nacional. Se concederá la autorización cuando se considere que está lo suficientemente motivada, pues cualquier interferencia en los derechos y libertades fundamentales se debe respetar siempre la esencia de los derechos y libertades. En cuanto a la excepción a su solicitud, en casos de urgencia, un usuario puede presentar una solicitud de autorización en un plazo de 24 horas a partir del momento en que se utilice el sistema RBI en tiempo real. En la práctica, este suele ser el momento en que se activan y despliegan las cámaras biométricas compatibles y se realiza la primera comparación biométrica con el sistema. Se establece que, si se rechaza una solicitud de autorización en caso de urgencia, el uso del sistema RBI en tiempo real debe suspenderse de inmediato. En caso de que la autoridad policial impugne la denegación, un administrador podrá conservar los datos hasta que se tome una decisión definitiva sobre la solicitud. Durante ese período, normalmente no se pondrán a disposición de la autoridad policial.

Algo fundamental en que no se puede tomar ninguna decisión que produzca un efecto jurídico adverso sobre una persona basándose *únicamente* en el resultado del sistema RBI en tiempo real. Por ejemplo, una persona es arrestada y encarcelada por un delito grave únicamente con base en su identificación mediante un sistema de reconocimiento facial, sin ninguna otra verificación. Tal y como se establece en el artículo 14, apartado 5, de la Ley de IA, el implementador no podrá adoptar ninguna medida ni decisión basándose en la identificación resultante del sistema, *"a menos que dicha identificación haya sido verificada y confirmada por separado por al menos*

dos personas físicas con la competencia, la formación y la autoridad necesarias" o a menos *que "el Derecho de la Unión o nacional considere desproporcionada la aplicación de este requisito".*

Por último, se requieren leyes nacionales para hacer operativo el uso de sistemas de información basada en la responsabilidad (RBI) en tiempo real en espacios de acceso público con fines policiales. Si se adopta una ley nacional que autorice el uso de RBI en tiempo real, la Ley de IA especifica los elementos sustantivos que deben contener las leyes nacionales para cumplir con los requisitos establecidos en la Ley de IA. Indica que la legislación nacional debe respetar los límites y condiciones establecidos en el artículo 5.1., apartado h. Esto implica que los Estados miembros no podrán ampliar los objetivos para los que se puede utilizar la información basada en la información (RBI) en tiempo real en espacios de acceso público con fines policiales mencionados anteriormente. Además, es competencia de cada Estado miembro que desee permitir el uso de los sistemas en cuestión, especificar en su Derecho nacional dichas normas, en aras proporcionar información pertinente y completa sobre el uso de sistemas de información basada en la responsabilidad en tiempo real a la autoridad competente para que pueda determinar la estricta necesidad y proporcionalidad de dicho uso. Además, el artículo 70 de la Ley de IA obliga a los Estados miembros a *"establecer al menos una autoridad de notificación y una autoridad de vigilancia del mercado",* así como a presentarán a la Comisión informes anuales sobre dicho uso (artículo 5.6. Ley IA). También tendrá la Comisión que presentar informes anuales sobre dicho uso, que no incluirán datos sensibles de las actividades policiales conexas.

Para finalizar, se incluyen una serie de supuestos que se encuentran fuera de alcance, como son los sistemas de verificación o autenticación biométrica y el uso retrospectivo. Por ejemplo, la legislación nacional podría autorizar a las autoridades policiales a realizar reconocimiento facial retrospectivo para comparar imágenes de sospechosos de delitos con imágenes faciales registradas en una base de datos de delitos. Otro uso que queda fuera del alcance es el uso de sistemas RBI en tiempo real con fines policiales, ya sea en un espacio privado (por ejemplo, en casa de alguien) o en línea (como el uso de una sala de chat o un juego en línea para identificar a un sospechoso de difundir material de abuso sexual infantil).

VI. CONCLUSIONES

Una vez plasmado el principal contenido de las referidas directrices, podemos afirmar que se convierte en un instrumento clarificador de diversos conceptos jurídicos, en tanto en cuanto la Ley de Inteligencia Artificial incluía ciertos conceptos indeterminados que no respondían a definiciones concretas, y que podrían dar lugar a múltiples dudas en la práctica. Si bien es cierto que, aunque todavía queden propuestas legislativas en el tintero dignas de ser aprobadas, como es la Propuesta de Directiva del Parlamento Europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial, a través de este instrumento se escenifican diversas prácticas prohibidas con la intención de proporcionar la información necesaria para que podamos ser conocedores de todas

las prácticas que no se pueden llevar a cabo a través del uso de este tipo de tecnología. Ahora bien, tal y como se relata en el documento, que algunas prácticas estén prohibidas no es óbice para que, gracias al uso de sistemas de inteligencia artificial se articulen nuevas formas de vigilancia que permitan un control efectivo de reiteradas conductas delictivas o que sirvan para encontrar e identificar en un menor tiempo a personas desaparecidas. Este instrumento es un buen ejemplo de lo que realmente ocurre con este tipo de sistemas: se pueden convertir en nuestro mejor aliado, o en el peor de los enemigos. Deben ponerse encima de la balanza y valorarse en todo momento los derechos fundamentales recogidos tanto en la Constitución Española como en la Carta de Derechos Fundamentales de la Unión Europea, teniendo en cuenta que el interés público resulta vencedor en esta lucha en algunas ocasiones, en las que se podrán utilizar este tipo de sistemas. No podemos olvidar que cuando determinados individuos se encuentran videovigilados, una gran cantidad de derechos fundamentales pueden verse vulnerados, como puede ser el derecho al honor, a la intimidad personar y familiar y a la propia imagen, u otro tipo de derechos que han surgido con el auge de la tecnología como son los neuroderechos[11], entre los que destacan en estos casos la protección contra los sesgos. Sin embargo, tal y como se recoge en la comentada normativa, se entenderá lícito el uso de estos sistemas cuando lo que esté en juego sean cuestiones como la seguridad nacional. Se prohíbe el uso de estas prácticas para impedir que nuestros derechos puedan verse vulnerados en todo momento, pero se permite el mismo cuando un interés superior deba verse protegido y este tipo de sistemas se conviertan en el método más rápido y efectivo para lograr ese objetivo.

VII. BIBLIOGRAFÍA

ÁLVAREZ SÁEZ K., La actual utilización de la inteligencia artificial para evaluar el riesgo de reincidencia, MARTÍN OSTOS, J (dir.), *Inteligencia artificial y derecho*, ed. Astigi, Sevilla, 2024.

ARAGÄO SEIA C., *Inteligencia artificial: responsabilidad civil 3.0*. El impacto de la era digital en el derecho, LÓPEZ ULLA J.M. (dir.), QUIROGA CORTI, M.P., (coor.), ed. Aranzadi, Pamplona, 2023.

EBERS, M., La utilización de agentes electrónicos inteligentes en el tráfico jurídico: ¿Necesitamos reglas especiales en el Derecho de la responsabilidad civil?, *Indret: Revista para el análisis del derecho*, julio 2016.

GIL MEMBRADO, C*., Riesgos del uso de algoritmos en el diagnóstico y en la investigación biomédica,* ed. Reus, Madrid, 2023.

MARTÍN RODRIGUEZ, G., *Nuevos horizontes en las políticas de la UE en materia de inteligencia artificial: hacia el Derecho Europeo de la IA*. GARCÍA SÁNCHEZ, B., JIMÉNEZ GARCÍA, F., (coord.), La atribución de una responsabilidad jurídico penal e internacional de la inteligencia artificial, ed. Iustel, Madrid, 2023.

[11] En este sentido véase CARRASCO GONZÁLEZ, M.C., APARICIO ARAQUE, B., Una aproximación a la categoría de los neuroderechos, *Publicaciones Jurídicas Centro de Estudios de Consumo*, diciembre 2024.

CARRASCO GONZÁLEZ, M.C., APARICIO ARAQUE, B., Una aproximación a la categoría de los neuroderechos, *Publicaciones Jurídicas Centro de Estudios de Consumo*, diciembre 2024.