REVISTA
DERECHO SOCIAL
Y EMPRESA

IMPACTO DE LA TECNOLOGÍA SOBRE EL EMPLEO. DIÁLOGOS CON HISPANOAMÉRICA

NÚMERO 23, JULIO DE 2025

JULIO A DICIEMBRE DE 2025, FECHA DE CIERRE JUNIO DE 2025

EL IMPACTO LABORAL DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

Jesús R. Mercader Uguina & Belén Velasco Pardo

ENTIDADES EDITORAS

UNIVERSIDAD DE LA RIOJA

Dykinson, S.L.

SAGARDOY BUSINESS&LAW SCHOOL

ISSN 2341-135X

EL IMPACTO EN LO LABORAL DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL

THE IMPACT OF THE ARTIFICIAL INTELLIGENCE REGULATION ON THE EMPLOYMENT SECTOR

JESÚS R. MERCADER UGUINA
Catedrático de Derecho del Trabajo y de la Seguridad Social
Universidad Carlos III de Madrid

BELÉN VELASCO PARDO
Abogada en Uría Menéndez

Fecha de envío: 17/04/2025
Fecha de aceptación: 26/06/2025

7.1. La responsabilidad sancionadora administrativa. 7.2. Responsabilidad civil por daños provocados por sistemas de IA.

RESUMEN: La lógica algorítmica y la inteligencia artificial como instrumentos de toma de decisiones irradian sus efectos en múltiples campos del actuar humano y, como no podía ser de otro modo, también en el laboral. En la Unión Europea, la aprobación del Reglamento de Inteligencia Artificial (RIA) abre las puertas a un marco regulatorio que equilibra la protección tecnológica con la protección de la salud, la seguridad, los derechos fundamentales, la democracia y el Estado de Derecho y que tiene claro impacto en las relaciones laborales y en la cada vez más común implementación de la inteligencia artificial en el lugar de trabajo. En el presente artículo se analiza el RIA, sus principios, el impacto que su regulación tiene en los intereses de los trabajadores por cuenta ajena y el papel del empresario que asume, cuando decide incorporar el uso de sistemas de inteligencia artificial, un conjunto de responsabilidades y obligaciones.

ABSTRACT: The use of algorithms and AI as automated decision-making tools has the potential for wide-reaching effects in multiple fields, which, unsurprisingly, also include the employment sector. In the European Union, the approval of the Artificial Intelligence Act (AIA) paves the way for a regulatory framework that balances technological advancement with the protection of health, safety, fundamental rights, democracy, and the rule of law. This uniform legal framework will have a clear impact on employment relationships and the increasingly common implementation of AI in the workplace. This article provides a detailed analysis of the AIA, outlining its key principles, the impact it has on employees' interests, and the set of responsibilities and obligations employers assume as part of their role when deciding to implement AI systems.

PALABRAS CLAVE: Algoritmo, big data, inteligencia artificial, derechos laborales, riesgo.

KEYWORDS: Algorithm, big data, artificial intelligence, employment rights, risk.

1. EL REGLAMENTO DE LA INTELIGENCIA ARTIFICIAL ATERRIZA EN LO LABORAL

Los recientes progresos en el ámbito de los algoritmos, la proliferación de conjuntos de datos digitales y las mejoras en cuanto al procesamiento, incluidos los aumentos de su potencia, los tiempos logarítmicos de ejecución y, por qué no decirlo, la disminución de los precios, se han unido para dar vida a una nueva clase de tecnología. La lógica algorítmica como instrumento de toma de decisiones irradia sus efectos en múltiples campos del actuar humano y, como no podía ser de otro modo, también en el laboral. Aquí, es la empresa la que, hasta el momento, ha incorporado su uso hasta el punto de que parece que el empresario está dispuesto a delegar o, si se prefiere, a descentralizar parte de sus poderes tradicionales trasladando un importante número de decisiones a la presunta objetividad y plena fiabilidad que proporciona el recurso al Big Data y, por extensión, a la Inteligencia Artificial (**IA**). Y ello en la medida en que su uso actual se proyecta sobre prácticamente la totalidad de las facetas que componen su autonomía organizativa, recorriendo transversalmente la libertad de actuación empresarial, abarcando desde la selección de trabajadores hasta la forma y modo de ejercicio del poder de dirección, incluyendo las decisiones empresariales de despedir[1].

Después de un largo período de tramitación y tras sortear variados obstáculos, el Parlamento Europeo ha dado luz verde al Reglamento de Inteligencia Artificial (Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 por el que

1 El presente estudio ha nutrido de los trabajos de J.R. MERCADER UGUINA, *Algoritmos e inteligencia artificial en el derecho digital del trabajo,* Valencia, Tirant lo Blanch, 2022. También, "En busca del empleador invisible: algoritmos e inteligencia artificial en el derecho digital del trabajo", *El Cronista del Estado Social y Democrático de Derecho*, 2022, nº 100, (Ejemplar dedicado a: Inteligencia artificial y derecho), pp.136-145. "El Reglamento de inteligencia artificial entra en la recta final, una primera lectura en clave laboral, *Revisa General de Derecho del Trabajo y de la Seguridad Social*, 2024, nº 67 (versión electrónica). "El principio de proporcionalidad como límite al control laboral basado en la inteligencia artificial", *Trabajo y Derecho: nueva revista de actualidad y relaciones laborales*, 2024, nº 19 (monográfico Impacto Laboral de la Inteligencia Artificial) (versión electrónica). "El "Big Bang" de la biometría laboral. De la huella dactilar a los neurodatos". *Labos: Revista de Derecho del Trabajo y Protección Social*, 2024, Volumen 5, Número nº 2, pp.4-23. "Inteligencia Artificial y relaciones laborales", Cuadernos de derecho transnacional, 2024, Vol. 16, nº 2, pp. 1114-1128. "*Algoritmos e inteligencia artificial en el derecho digital del trabajo*", en María Emilia Casas Baamonde (Directora) Daniel Pérez del Prado (Coordinador), Derecho y Tecnologías, Madrid, Fundación Areces, 2024, pp. 335-373. También de las entradas en distintos blogs: "El Reglamento de Inteligencia Artificial: frecuentemos el futuro", Brief AEDTSS 42/2024, "Los usos de alto riesgo en el ámbito laboral de la IA y la autocertficación", El Foro de Labos, 9 de mayo de 2024 y "Pintando la caja negra: ¿qué debe entenderse por "derecho a obtener información significativa sobre la lógica aplicada" por los algoritmos?", El Foro de Labos, 7 de mayo de 2025.

se establecen normas armonizadas en materia de inteligencia artificial) (en adelante, **RIA**). Una norma cuyo objetivo declarado en su primer artículo es promover la adopción de una IA centrada en el ser humano y fiable y garantizar un elevado nivel de protección de la salud, la seguridad, los derechos fundamentales, la democracia y el Estado de Derecho frente a los efectos nocivos de los sistemas de IA en la Unión, apoyando al mismo tiempo la innovación. Su escalonada entrada en vigor (art. 113 RIA) permitirá su adecuada comprensión, pero, a su vez, esta vigencia diferida tendrá que luchar contra el acelerado cambio que vive esta materia donde el futuro se convierte en pasado a enorme velocidad.

2. LA DEFINICIÓN JURÍDICA DE LA INTELIGENCIA ARTIFICIAL Y DE LOS "MODELOS FUNDACIONALES"

Un primer avance en el terreno jurídico que incorpora el RIA es la definición de "sistema de inteligencia artificial". Una noción que se alinea definitivamente con la propuesta por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) en mayo de 2019, enmendada en 2024, y que ahora aparece, por primera vez, definida en una norma.

Se entiende, así, por tal "*un sistema basado en una máquina diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras el despliegue y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar información de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos o virtuales*". Una definición que reproduce también la Convención Marco del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho.

El Considerando (12) del RIA puede ser un buen instrumento de interpretación del alcance que ha querido dar la norma de la Unión Europea al citado concepto. Siguiendo lo establecido en el artículo 96.1 f) del RIA, el 4 de febrero de 2025 la Comisión Europea publicó, aunque de manera todavía no oficial, las directrices sobre la definición de IA establecida por el RIA (*Guidelines on the definition of an artificial intelligence system*).

La norma también busca dotar de un régimen jurídico propio a los que califica de "modelos de IA de uso general" (art. 3.63), denominados en versiones anteriores del RIA como "modelos fundacionales", esto es, la inteligencia artificial generativa (**IAG**) que tiene en ChapGPT su mascarón de proa[2]. Como aclara el Considerando (97) del RIA: "*El concepto*

2 Por ejemplo, BERT, uno de los primeros modelos fundacionales bidireccionales, se lanzó en 2018. Se entrenó con 340 millones de parámetros y un conjunto de datos de entrenamiento de 16 GB. En 2023,

de modelos de IA de uso general debe definirse claramente y diferenciarse del concepto de sistemas de IA con el fin de garantizar la seguridad jurídica. La definición debe basarse en las características funcionales esenciales de un modelo de IA de uso general, en particular la generalidad y la capacidad de realizar de manera competente una amplia variedad de tareas diferenciadas. Estos modelos suelen entrenarse usando grandes volúmenes de datos y a través de diversos métodos, como el aprendizaje autosupervisado, no supervisado o por refuerzo".

La IAG utiliza técnicas avanzadas de aprendizaje automático, especialmente las redes neuronales artificiales, para aprender de los datos y producir contenidos que sean realistas, coherentes y variados. El RIA viene también a definir jurídicamente estos sistemas y lo hace entendiendo por tal *"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".*

3. PRINCIPIOS INSPIRADORES DEL REGLAMENTO DE INTELIGENCIA ARTIFICIAL Y LOS SUBPRINCIPIOS DE ALFABETIZACIÓN Y DE NO MENOSCABO DE LOS DERECHOS E INTERESES DE LOS TRABAJADORES POR CUENTA AJENA

Como subraya el Considerando (1) del RIA, "*el objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme, en particular para el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial (en lo sucesivo, «sistemas de IA») en la Unión, de conformidad con los valores de la Unión, a fin de promover la adopción 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*

solo cinco años después, OpenAI entrenó el GPT-4 mediante la utilización de 170 billones de parámetros y un conjunto de datos de entrenamiento de 45 GB. Según OpenAI, la potencia de cómputo requerida para el modelado fundacional se ha duplicado cada 3,4 meses desde 2012. Los modelos fundacionales actuales, como los modelos de lenguaje de gran tamaño (LLM) Claude 2 y Llama 2, y el modelo de conversión de texto a imagen Stable Diffusion de Stability AI, pueden realizar una serie de tareas listas para usar que abarcan múltiples dominios, como escribir publicaciones de blog, generar imágenes, resolver problemas matemáticos, entablar diálogos y responder preguntas basadas en un documento.

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".

El RIA toma como principios inspiradores una serie de ideas definidas por el Grupo independiente de expertos de alto nivel sobre IA. Dicho grupo, como precisa el Considerando (27) del RIA, desarrolló siete principios éticos no vinculantes para la IA que tienen por objeto contribuir a garantizar la fiabilidad y el fundamento ético de la IA. Los siete principios son: (i) acción y supervisión humanas; (ii) solidez técnica y seguridad; (iii) gestión de la privacidad y de los datos; (iv) transparencia; (v) diversidad, no discriminación y equidad; (vi) bienestar social y ambiental, y (vii) rendición de cuentas. En palabras del propio Considerando (27), "*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*". La definición que se establece en el RIA para cada uno de esos principios es la que sigue:

a) «Acción y supervisión humanas»: los sistemas de IA se desarrollarán y utilizarán como una herramienta al servicio de las personas, que respete la dignidad humana y la autonomía personal, y que funcione de manera que pueda ser controlada y vigilada adecuadamente por seres humanos.
b) «Solidez y seguridad técnicas»: los sistemas de IA se desarrollarán y utilizarán de manera que se minimicen los daños imprevistos e inesperados, así como para que sean sólidos en caso de problemas imprevistos y resistentes a los intentos de modificar el uso o el rendimiento del sistema de IA para permitir una utilización ilícita por parte de terceros malintencionados.
c) «Gestión de la privacidad y de los datos»: los sistemas de IA se desarrollarán y utilizarán de conformidad con las normas vigentes en materia de privacidad y protección de datos, y tratarán datos que cumplan normas estrictas en términos de calidad e integridad.
d) «Transparencia»: los sistemas de IA se desarrollarán y utilizarán facilitando una trazabilidad y explicabilidad adecuadas, haciendo que las personas sean conscientes de que se comunican o interactúan con un sistema de IA, informando debidamente a los usuarios sobre las capacidades y limitaciones de dicho sistema de IA e informando a las personas afectadas de sus derechos.
e) «Diversidad, no discriminación y equidad»: los sistemas de IA se desarrollarán y utilizarán incluyendo a diversos agentes y promoviendo la igualdad de acceso, la igualdad de género y la diversidad cultural, evitando al mismo tiempo los efectos discriminatorios y los sesgos injustos prohibidos por el Derecho nacional o de la Unión.

f) «Bienestar social y ambiental»: los sistemas de IA se desarrollarán y utilizarán de manera sostenible y respetuosa con el medio ambiente, así como en beneficio de todos los seres humanos, al tiempo que se supervisan y evalúan los efectos a largo plazo en las personas, la sociedad y la democracia.

Aunque no aparece concretado en la lista de principios recién enunciada, el RIA otorga especial importancia a lo que denomina "alfabetización". El RIA otorga carácter de principio a lo que denomina "alfabetización en materia de inteligencia artificial" que conceptúa en el artículo 3.56 y enuncia en su artículo 4 del siguiente modo: "*Los proveedores y responsables del despliegue de sistemas de IA* [en el caso laboral, los empleadores] *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 grupos de personas en que se utilizarán dichos sistemas*". Como precisa el Considerando (20), "*la alfabetización en materia de inteligencia artificial debe proporcionar a todos los agentes pertinentes de la cadena de valor de la IA los conocimientos necesarios para garantizar el cumplimiento adecuado y la correcta ejecución. Además, la puesta en práctica general de medidas de alfabetización en materia de inteligencia artificial y la introducción de acciones de seguimiento adecuadas podrían contribuir a mejorar las condiciones de trabajo y, en última instancia, sostener la consolidación y la senda de innovación de una IA fiable en la Unión*".

De igual modo, cabe extraer, con expresa proyección en lo laboral, lo que cabría calificar como el principio de "no menoscabo de los derechos e intereses de los trabajadores por cuenta ajena". Definido su ámbito de aplicación, el RIA, como precisa su Considerando (9), conforma un conjunto de normas armonizadas que "*deben aplicarse en todos los sectores y, en consonancia con el nuevo marco legislativo, deben entenderse sin perjuicio del Derecho vigente de la Unión, en particular en materia de protección de datos* (...)*, derechos fundamentales, empleo, protección de los trabajadores* (...)".

Una regla de salvaguardia con expresa proyección en lo laboral que queda materializada en su artículo 2.11: "*El presente Reglamento no impedirá que los Estados miembros o la Unión mantengan o introduzcan disposiciones legales, reglamentarias o administrativas que sean más favorables a los trabajadores en lo que atañe a la protección de sus derechos respecto al uso de sistemas de IA por parte de los empleadores o fomenten o permitan la aplicación de convenios colectivos que sean más favorables a los trabajadores*". Y es que, como viene a aclarar de nuevo el Considerando (9), esta norma "*tampoco debe afectar en modo alguno al ejercicio de los derechos fundamentales reconocidos en los Estados miembros y a escala de*

la Unión, incluidos el derecho o la libertad de huelga o de emprender otras acciones contempladas en los sistemas de relaciones laborales específicos de los Estados miembros y el derecho a negociar, concluir y hacer cumplir convenios colectivos o a llevar a cabo acciones colectivas conforme a la legislación nacional". De igual modo, se añade, "*el presente Reglamento no debe afectar a las disposiciones destinadas a mejorar las condiciones laborales en el trabajo en plataformas digitales*", en los términos definidos por la Directiva (UE) 2024/2831 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, relativa a la mejora de las condiciones laborales en el trabajo en plataformas.

El Dictamen del Comité Económico y Social Europeo, recientemente publicado en el DOUE, destaca que el diálogo social y la participación de los trabajadores desempeñan un papel crucial a la hora de preservar los derechos fundamentales de los trabajadores y de promover una «IA fiable» en el mundo laboral, pues constituye uno de los mecanismos para minimizar los riesgos y los efectos nocivos de los sistemas de IA. En sus conclusiones, el CESE hace un firme llamamiento en favor de un diálogo social efectivo en torno a la utilización de los sistemas de IA sobre la base de un instrumento jurídico ad hoc de la UE que incluya disposiciones para lograr lo siguiente de manera más eficaz:

- Permitir la aplicación del artículo 88 del RGPD y ofrecer orientaciones explícitas sobre el consentimiento y el interés legítimo;
- Ampliar el ámbito de aplicación de las disposiciones de la Directiva sobre Trabajo en Plataformas Digitales para abordar los retos que suponen los sistemas de gestión algorítmica para todos los trabajadores;
- Reforzar las normas aplicables en virtud de la Directiva 2002/14/CE cuando se introduzcan sistemas de IA de alto riesgo y ofrecer orientaciones explícitas relativas a las disposiciones de la Directiva 89/391/CEE sobre salud y seguridad en el trabajo;
- Integrar la dimensión del proceso dinámico del diálogo social y las evaluaciones de riesgos de los sistemas de IA, tal como se definen en el Acuerdo marco autónomo de los interlocutores sociales europeos sobre digitalización de 2020;
- Hacer extensiva la comunicación de las evaluaciones de impacto relativas a la protección de datos a los representantes de los trabajadores, tal como se establece en la Directiva sobre Trabajo en Plataformas Digitales;
- Facilitar evaluaciones de impacto ex ante en materia de derechos fundamentales, que deberán llevar a cabo los proveedores antes de que empiecen a utilizarse sistemas de alto riesgo; y
- Establecer orientaciones claras sobre el modo en que pueden emplearse los espacios controlados de prueba y aprovecharse las pruebas en condiciones reales.

4. EL RIESGO COMO PILAR CENTRAL EN LA CONSTRUCCIÓN DEL REGLAMENTO

El riesgo es la dinámica que impulsa el desarrollo de una sociedad empeñada en el cambio y que dirige sus actuaciones hacia la determinación de su propio futuro, en la que se rechaza la posibilidad de que el mismo quede en manos de la religión, la tradición o los caprichos de la naturaleza. La opción ínsita en tal realidad se produce entre el bloqueo y la parálisis o la necesidad de asumir constantes riesgos que se aceptan en virtud de un tácito y difuso consenso en el que, de forma más o menos consciente, se ponderan los riesgos que genera una actividad y las necesidades que con ella se superan. Dentro de una opción deliberada, cada sociedad asume, en relación con su momento histórico, una cierta dosis de riesgo.

El desarrollo tecnológico y el maquinismo han generado multitud de riesgos con los que cotidianamente vivimos. La novedad que, sin embargo, se registra en la actual sociedad de riesgo, y que decisivamente la caracteriza, es que el riesgo excede el marco de la responsabilidad y su genuina funcionalidad reparadora para convertirse en un problema de Estado y, por ende, de responsabilidad política que reclama la decidida intervención de los poderes públicos, y no con una orientación reparadora, sino de prevención, reducción y, en lo posible, eliminación de riesgos. A medida que la sociedad se va complicando tecnológicamente, se convierte de forma progresiva en una sociedad de riesgo[3].

El RIA sitúa, por ello, su centro de gravedad en la valoración del riesgo que conlleva el uso de los sistemas y modelos de IA. La noción de "riesgo", definida en dicha norma como "*la combinación de la probabilidad de que se produzca un daño y la gravedad de dicho daño*" (art. 3.2 del RIA), sirve de base para establecer una "pirámide de riesgos" ascendente (del riesgo medio/bajo hasta el riesgo inaceptable, pasando por el riesgo alto) que se emplea para clasificar una serie de prácticas y casos de uso de la IA en ámbitos específicos, lo que supone reconocer que no todos los tipos de IA suponen un riesgo y que no todos los riesgos son iguales o requieren las mismas medidas de mitigación. Por ello, y en recta correspondencia, el RIA crea un sistema de obligaciones, garantías y responsabilidades para todos los agentes que actúan dentro de este nuevo ecosistema (proveedores, fabricantes, responsables del despliegue y, en el sentido más amplio, afectados por el uso de estos sistemas). Se construye, de este modo, lo que venimos calificando como el "triángulo de oro" del RIA: aproximación desde el riesgo, garantías y responsabilidades.

3 BECK, U., *La sociedad del riesgo. Hacia una nueva modernidad,* Paidos, Barcelona, 1998, *passim*.

4.1. Sistemas de IA prohibidos y sus proyecciones en el ámbito laboral

La lista de prácticas prohibidas abarca todos los sistemas de IA cuyo uso se considera inaceptable por ser contrario a los valores de la Unión. Por ejemplo, porque violan derechos fundamentales. Las prohibiciones engloban aquellas prácticas que tienen un gran potencial para manipular a las personas mediante técnicas subliminales que trasciendan su consciencia o que aprovechan las vulnerabilidades de grupos vulnerables concretos, como los menores o las personas con discapacidad, para alterar de manera sustancial su comportamiento de un modo que es probable que les provoque perjuicios físicos o psicológicos a ellos o a otras personas.

Y es que, como completa el Considerando (28) del RIA, "*Al margen de los múltiples usos beneficiosos de la IA, esta también puede utilizarse indebidamente y proporcionar nuevas y poderosas herramientas para llevar a cabo prácticas de manipulación, explotación y control social. Dichas prácticas son sumamente perjudiciales e incorrectas y deben estar prohibidas, pues van en contra de los valores de la Unión de respeto de la dignidad humana, la libertad, la igualdad, la democracia y el Estado de Derecho y de los derechos fundamentales consagrados en la Carta, como el derecho a la no discriminación, a la protección de datos y a la intimidad y los derechos del niño*".

Con todo, el ámbito donde esta imbricación alcanza su máximo nivel es el de la métrica de las personas, la biometría. Un repaso de las categorías empleadas por el RIA en relación con estos sistemas ["datos biométricos" (art. 3.34), "identificación biométrica" (art. 3.35), "verificación biométrica" (art. 3.36), "sistema de reconocimiento de emociones" (art. 3.39), "sistema de categorización biométrica" (art. 3.40), "sistema de identificación biométrica remota" (art. 3.41), "sistema de identificación biométrica remota en tiempo real" (art. 3.42), o "sistema de identificación biométrica remota en diferido" (art. 3.43)] muestra su relevancia. Una relevancia que se proyecta sobre los distintos modelos que se vinculan con los también distintos niveles de riesgo. La era del capitalismo de la vigilancia, en la conocida expresión de Zuboff[4], supone un enorme desafío al que viene a tratar de poner coto esta norma.

En este ámbito quedan directamente proscritos los sistemas de reconocimiento de emociones[5]. El RIA expresamente prohíbe "*la introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA para inferir las emociones de una persona física en los ámbitos de la aplicación de la ley (...) en lugares de trabajo* (...)" (art. 5.1 f) RIA).

4 ZUBOFF, S., *La era del capitalismo de la vigilancia*, Barcelona, Paidos, 2020.

5 La Comisión publicó la versión aun no oficial, en inglés, de su Comunicación C (2025) 884 final, con directrices sobre las prácticas de IA prohibidas de acuerdo con el Reglamento (EU) 2024/1689 (Commission Guidelines on prohibited artificial intelligence practices).

Las emociones, se ha dicho, no son algo que me ocurre, sino algo que yo hago. Los juicios y cogniciones afectan a las emociones y son la causa de que estás tengan lugar. En suma, son "disposiciones mentales" que generan actitudes y éstas pueden ser objeto de control, valoración y seguimiento. Existe, por ello, un vínculo estrecho entre los sistemas de reconocimiento de emociones y los datos personales de la persona trabajadora en la medida en que los primeros se alimentan de los datos biométricos (expresiones faciales, actividad cerebral o voz, entre otros). El Considerando (18) del RIA viene a precisar que "*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 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*".

Un magnífico ejemplo de esta realidad es la resolución de la Agencia de Protección de Datos húngara (NAIH) en la que se revisaba la práctica llevada a cabo por un banco durante 45 días y que consistía en utilizar un software de procesamiento de señales de voz basado en IA[6]. El mencionado software analizaba y evaluaba los estados emocionales de los clientes y las palabras clave utilizadas en las llamadas. La finalidad de esta tecnología era gestionar las quejas, controlar la calidad de las llamadas y del trabajo y, además, aumentar la eficiencia de los empleados. A continuación, los resultados de este análisis se almacenaban junto con las grabaciones de las llamadas y estos datos se usaban para clasificar las llamadas en orden de prioridad. La justificación del banco para el procesamiento de datos se basó en su interés legítimo de garantizar buenos niveles de retención de clientes y eficiencia. Sin embargo, el NAIH concluyó que el banco no había considerado adecuadamente los intereses en juego y le sancionó con una multa de 670.000 euros, obligándole a suspender el uso del sistema de análisis de emociones descrito.

Igualmente se encuentra prohibida "*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* (...)" (art. 5.1 g) RIA).

6 Analizado por MUÑOZ RUIZ, A.B., *Biometría y sistemas automatizados de reconocimiento de emociones: Implicaciones Jurídicos-Laborales*, Valencia, Tirant lo Blanch, 2023, pp. 149-153.

4.2. La señalización laboral de los sistemas de alto riesgo

La pieza nuclear del sistema que construye el RIA se asienta en el establecimiento de límites al uso de los sistemas que califica de "alto riesgo" (art. 6.1 y 2 por relación con lo establecido en el Anexo III del RIA), a los que dedica la norma la mayor parte de su extenso contenido. Una noción que, como precisa el Considerando (46), se diseña tomando en cuenta sus efectos, por cuanto incluye entre tales sistemas aquellos que "*tengan consecuencias perjudiciales importantes para la salud, la seguridad y los derechos fundamentales de las personas de la Unión, y dicha limitación reduce al mínimo cualquier posible restricción del comercio internacional, si la hubiera*".

Lo laboral ocupa, también aquí, un papel protagonista. Y es que, como viene a subrayar el considerando (48), "*la magnitud de las consecuencias adversas de un sistema de IA para los derechos fundamentales protegidos por la Carta es especialmente importante a la hora de clasificar un sistema de IA como de alto riesgo. Entre dichos derechos se incluyen 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, la no discriminación, (y) los derechos de los trabajadores* (...)".

La lectura del artículo 6.2 en relación con el Anexo III de RIA ratifica de modo concluyente la referida proyección sobre lo social. En particular, el Anexo III incorpora entre los sistemas de IA de alto riesgo los que afecten al "*empleo, gestión de los trabajadores y acceso al autoempleo*" (apartado 3). Y en concreto, considera sistema de IA de alto riesgo "*a) Sistemas de IA destinados a ser utilizados para la contratación o la selección de personas físicas, en particular para publicar anuncios de empleo específicos, analizar y filtrar las solicitudes de empleo y evaluar a los candidatos*" y "*b) Sistemas de IA destinados a utilizarse para tomar decisiones o influir sustancialmente en ellas que afecten a la iniciación, promoción y resolución de relaciones contractuales de índole laboral, a la asignación de tareas basada en la conducta individual o en rasgos o características personales, o al seguimiento y evaluación del rendimiento y la conducta de las personas en el marco de dichas relaciones*".

El Considerando (57) justifica dicha inclusión sobre la base de que los mismos: "*pueden perpetuar patrones históricos de discriminación, por ejemplo, contra las mujeres, ciertos grupos de edad, las personas con discapacidad o las personas de orígenes raciales o étnicos concretos o con una orientación sexual determinada, durante todo el proceso de contratación y en la evaluación, promoción o retención de personas en las relaciones contractuales de índole laboral. Los sistemas de IA empleados para controlar el rendimiento y el comportamiento de estas personas también pueden socavar sus derechos fundamentales a la protección de los datos personales y a la intimidad*". El uso de sistemas de IA para el control, el seguimiento y la evaluación de trabajadores plantea, por tanto, serias preocupaciones con respecto de sus derechos funda-

mentales de las personas trabajadoras y al establecimiento de unas condiciones de trabajo equitativas y justas, como ya hemos tenido oportunidad de señalar en trabajos anteriores.

Estas preocupaciones se ven recogidas en el propio RIA, que incorpora un importante sistema de garantías que se anudan a los requisitos generales que deberán cumplir los sistemas de IA de alto riesgo *"teniendo en cuenta sus finalidades previstas, así como el estado actual de la técnica generalmente reconocido en materia de IA"* (art. 8). Ello lleva consigo el establecimiento, implantación, documentación y mantenimiento de un sistema de gestión de riesgos entendido como "*un proceso iterativo continúo planificado y ejecutado durante todo el ciclo de vida de un sistema de IA de alto riesgo, que requerirá revisiones y actualizaciones sistemáticas periódicas*" (art. 9). A ello se une la necesaria "*gobernanza de los datos*" (art. 10), las exigencias de una precisa documentación técnica (art. 11) y la necesidad de garantizar un nivel de trazabilidad del funcionamiento del sistema (art. 12). La transparencia suficiente para que los responsables del despliegue (en el caso laboral, los empleadores) interpreten y usen correctamente la información de salida que incorporen los sistemas de IA, constituye un principio maestro (art. 13) al que se une la necesidad de que su diseño y desarrollo permita cumplir con el principio de humano al mando (art. 14). En fin, unos sistemas que se diseñarán y desarrollarán de modo que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme durante todo su ciclo de vida (art. 15).

La arquitectura de la privacidad se está transformando radicalmente en nuestros días. Los algoritmos y la IA se están convirtiendo en instrumentos en los que la empresa post-material delega funciones centrales de su poder. Por el momento son las grandes empresas las que implementan estos modelos, pero es cuestión de tiempo que los mismos se expandan también a las pequeñas y medianas. Su uso se está generalizando y va desde los procesos de selección hasta las múltiples formas de ejercicio de las facultades de dirección y control, llegando, incluso, a las propias decisiones extintivas. En algún trabajo he calificado este proceso como la "gran delegación empresarial"[7].

5. GARANTÍAS FRENTE LOS SISTEMAS DE ALTO RIESGO

Como explica Ferrajoli, "*las garantías no son otra cosa que las técnicas previstas por el ordenamiento para reducir la distancia estructural entre normatividad y efectividad, y, por tanto, para posibilitar la máxima eficacia de los derechos fundamentales*". Por ello,

7 Ampliamente, MERCADER UGUINA, J.R., "En busca del empleador invisible: algoritmos e inteligencia artificial en el derecho digital del trabajo", *El Cronista del Estado Social y Democrático de Derecho*, 2022, nº 100, (Ejemplar dedicado a: Inteligencia artificial y derecho), pp.136-145.

continúa, "*reflejan la diversa estructura de los derechos fundamentales para cuya tutela y satisfacción han sido previstas*". En todo caso, "*el garantismo de un sistema jurídico es una cuestión de grado que depende de la precisión de los vínculos positivos o negativos impuestos a los poderes por las normas constitucionales y por el sistema de garantías que aseguran una tasa más o menos elevada de eficacia de tales vínculos*"[8]. Los sistemas de alto riesgo deben cumplir con una serie de requisitos:

5.1. La transparencia y explicabilidad algorítmica como principio maestro

La IA puede suponer un cambio de paradigma que se centra en el abandono de la causalidad como criterio central y su sustitución por la correlación, una transformación tan relevante que puede generar dificultades en la explicabilidad de las decisiones que se tomen basadas en algoritmos de Big Data. La explicabilidad es una característica activa del modelo que se refiere a la capacidad de generar una explicación sobre el comportamiento del modelo a partir de los datos utilizados, de los resultados obtenidos y del proceso completo de la toma de decisión en función de la audiencia o perfil de los destinatarios a los que se dirige la explicación.

En esta línea se viene insistiendo desde que, en 2019, la OCDE publicó sus "Principios sobre Inteligencia artificial" con el objetivo de promover una IA innovadora y que genere confianza, y que respete los derechos humanos y los valores democráticos. El libro blanco de la Comisión sobre Inteligencia Artificial, preludio de la futura propuesta legislativa europea sobre el particular, identifica criterios para determinar qué usos de la IA pueden representar un riesgo importante para el ser humano, para evaluarlo y mitigarlo. La propuesta de Reglamento para garantizar la equidad en la asignación de valor en toda la economía de datos (Data Act) de la Comisión, vendrá a establecer que siempre que un titular de datos esté legalmente obligado a proporcionar datos a un tercero dicha información se realizará en términos justos, razonables y no discriminatorios y de manera transparente.

Dichos requisitos en materia de transparencia y de explicabilidad de la toma de decisiones por la IA también deben ayudar a contrarrestar los efectos disuasorios de la asimetría digital y los llamados «patrones oscuros», dirigidos contra las personas y su consentimiento informado.

El artículo 13 de la RIA, apunta en esta dirección al señalar que los sistemas deberán estar diseñados de forma que los usuarios interpreten y usen correctamente la información suministrada por el sistema; además, deberán ir acompañados de unas instrucciones de

8 FERRAJOLI, L., *Derechos y garantías. La Ley del más débil*, Madrid, Trotta, 1999, p. 25.

uso concisas, correctas, completas y claras. En concreto, precisa, los sistemas de IA de alto riesgo se diseñarán y desarrollarán de un modo que se garantice que funcionan con "*nivel de transparencia suficiente para que los proveedores y usuarios entiendan razonablemente el funcionamiento del sistema*". En este sentido, deberán garantizar un tipo y un nivel de transparencia adecuados para que el usuario y el proveedor cumplan las obligaciones oportunas previstas normativamente.

Por otra parte, los sistemas de IA de alto riesgo irán acompañados de las instrucciones de uso correspondientes en un formato digital o de otro tipo adecuado, las cuales incluirán información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los usuarios. Esa información especificará, entre otras cosas: (i) la identidad y los datos de contacto del proveedor y, en su caso, de sus representantes autorizados; (ii) las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo, y en particular, el grado en que el sistema de IA pueda ofrecer una explicación de las decisiones que adopte; así como su funcionamiento en relación con las personas o los grupos de personas en relación con los que se pretenda utilizar el sistema; (iii) los cambios en el sistema de IA de alto riesgo y su funcionamiento predeterminado por el proveedor en el momento de efectuar la evaluación de la conformidad inicial, en su caso; (iv) las medidas de vigilancia humana a que se hace referencia en el artículo 14 de la PRIA, incluidas las medidas técnicas establecidas para facilitar la interpretación de la información de salida de los sistemas de IA por parte de los usuarios; o, en fin (v) la vida útil prevista del sistema de IA de alto riesgo, así como las medidas de mantenimiento y cuidado necesarias para garantizar el correcto funcionamiento de dicho sistema, también en lo que respecta a la actualización del software, durante toda su vida útil prevista. Particularmente relevante es el establecimiento de una exigencia complementaria a todas las anteriores relativa al hecho de que "*los proveedores y los usuarios garantizarán un nivel suficiente de alfabetización en materia de IA*".

5.2. Principio de gobernanza de los datos y lucha contra los sesgos

El gobierno de los datos constituye un principio medular en la gestión y control de los sistemas de IA. Como señala el Considerando (67) de la RIA:

> "*El acceso a datos de alta calidad es 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 funciona del modo previsto y en*

condiciones de seguridad y no se convierte en una fuente de alguno de los tipos de discriminación prohibidos por el Derecho de la Unión. Es preciso instaurar prácticas adecuadas de gestión y gobernanza de datos para lograr que los conjuntos de datos de entrenamiento, validación y prueba sean de buena calidad".

Aunque en el desarrollo y entrenamiento de los procesos desarrollados por sistemas de IA no será necesario el tratamiento de datos personales al utilizarse, normalmente, datos anonimizados[9], lo cierto es que la inseguridad que proporcionan estos sistemas y, de forma más relevante, el hecho de que se proyecten en un terreno que, como el laboral, resulta especialmente sensible a los peligros derivados del tratamiento de datos (dada su perdurabilidad, su carácter personal, la diversidad de escenarios para los que pueden ser relevantes; y, en fin, el elevado número de personas trabajadoras que quedan afectadas), imponen controles de calidad muy severos sobre su tratamiento.

Especialmente importantes son, por ello, las exigencias derivadas del principio de minimización, que obligan a escoger aquella tecnología que resulte menos intrusiva desde el punto de vista de la protección de datos. El artículo 5.1 c) del Reglamento general de protección de datos (RGPD) establece con carácter general que los datos deben ser adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados. Una exigencia que cobra especial importancia en el terreno de los tratamientos automatizados porque, como pone de manifiesto el Dictamen 2/2017 (WP251) del Grupo de Trabajo del Artículo 29, relativo al tratamiento de datos en el trabajo, elaborado bajo el marco de la Directiva 95/46/CE, "*los algoritmos de aprendizaje automático están diseñados para procesar grandes volúmenes de información y generar correlaciones que permitan a las organizaciones crear perfiles de personas muy exhaustivos y sólidos. Aunque, en el caso de la elaboración de perfiles, conservar datos puede presentar ventajas, dado que el algoritmo podrá aprender de un mayor número de datos, los responsables del tratamiento deben cumplir el principio de minimización de datos al recoger datos personales y garantizar que conservan dichos datos durante no más tiempo del necesario y de forma proporcional a los fines del tratamiento de los datos personales*".

La calidad de los datos a través de los cuales se entrenan los algoritmos resulta, por todo ello, esencial en la medida en que dicha depuración evitará la existencia de sesgos selectivos. Sobre esta base, el artículo 10.3 del RIA viene a precisar que "*los conjuntos de*

9 Tal y como señala el Considerando (26) RGPD: "(…) *los principios de protección de datos no deben aplicarse a la información anónima, es decir información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo. En consecuencia, el presente Reglamento no afecta al tratamiento de dicha información anónima, inclusive con fines estadísticos o de investigación*".

datos de entrenamiento, validación y prueba deben ser lo suficientemente pertinentes y representativos, carecer de errores y ser completos en vista de la finalidad prevista del sistema". Asimismo, deberán tener las propiedades estadísticas adecuadas, también en lo que respecta a las personas o los grupos de personas en las que en un principio se usará el sistema de IA de alto riesgo. En concreto, concluye el apartado 4 del citado artículo 10 que: "*los conjuntos de datos de entrenamiento, validación y prueba tendrán en cuenta, en la medida necesaria en función de su finalidad prevista, las características o elementos particulares del contexto geográfico, conductual o funcional específico en el que se pretende utilizar el sistema de IA de alto riesgo*". Se admite la posibilidad de tratar categorías especiales de datos personales "*ofreciendo siempre las salvaguardias adecuadas para los derechos y las libertades fundamentales de las personas físicas, lo que incluye establecer limitaciones técnicas a la reutilización y la utilización de las medidas de seguridad y protección de la privacidad más recientes, tales como la seudonimización o el cifrado, cuando la anonimización pueda afectar significativamente al objetivo perseguido*" (art. 9.1 RGPD).

Con todo, la normativa es especialmente sensible a los posibles sesgos que incorporen los datos objeto de tratamiento. Así lo subraya el Considerando (67) del RIA:

> "*Los conjuntos de datos deben tener las propiedades estadísticas adecuadas, también en lo que respecta a las personas o los colectivos de personas en relación con los que esté previsto utilizar el sistema de IA de alto riesgo, prestando una atención especial a la mitigación de los posibles sesgos en los conjuntos de datos que puedan afectar a la salud y la seguridad de las personas físicas, tener repercusiones negativas en los derechos fundamentales o dar lugar a algún tipo de discriminación prohibida por el Derecho de la Unión, especialmente cuando los datos de salida influyan en la información de entrada de futuras operaciones (bucles de retroalimentación). 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 pertenecientes a determinados colectivos vulnerables, incluidos colectivos raciales o étnicos* (…)".

Sobre esta base, el texto de RIA incorpora diversos mandatos en esta dirección. Así el artículo 10.2 señala que: "*Los conjuntos de datos de entrenamiento, validación y prueba se someterán a una gobernanza adecuada al contexto del uso, así como a la finalidad prevista del sistema de IA*", estableciendo que, entre otros, dichas medidas se centrarán, en particular,

en: "*f) el examen atendiendo a posibles sesgos que puedan afectar a la salud y la seguridad de las personas, afectar negativamente a los derechos fundamentales o dar lugar a una discriminación prohibida por el Derecho de la Unión, especialmente cuando los datos de salida influyan en los datos de entrada en futuras operaciones («bucle de retroalimentación») y medidas adecuadas para detectar, prevenir y mitigar posibles sesgos*".

Añade el apartado 5 del referido artículo 10 que, "*En la medida en que sea estrictamente necesario para garantizar la detección y la corrección de los sesgos negativos asociados a los sistemas de IA de alto riesgo*", los proveedores de dichos sistemas podrán tratar excepcionalmente las categorías especiales de datos personales que se mencionan en el artículo 9, apartado 1, del Reglamento (UE) 2016/679; el artículo 10 de la Directiva (UE) 2016/680, y el artículo 10, apartado 1, del Reglamento (UE) 2018/1725, ofreciendo siempre las salvaguardias adecuadas para los derechos y las libertades fundamentales de las personas físicas, lo que incluye establecer limitaciones técnicas a la reutilización y la utilización de las medidas de seguridad y protección de la privacidad más recientes. En particular, se aplicarán todas las condiciones siguientes para que se produzca este tratamiento, precisando, entre otras medidas: "*a) que el tratamiento de datos sintéticos o anonimizados no permita alcanzar eficazmente la detección y corrección de sesgos*"; o que "*e) que las categorías especiales de datos personales se eliminen una vez que se haya corregido el sesgo o los datos personales hayan llegado al final de su período de conservación*".

Conviene recordar que los datos sintéticos son datos generados artificialmente. Los datos sintéticos se crean de forma algorítmica y se utilizan como soporte para los conjuntos y pruebas y producción de datos, para validar modelos matemáticos y cada vez más, para entrenar modelos de aprendizaje automático. Estos datos algorítmicos nacen "*a partir de un conjunto de datos que queremos proteger pero que necesitamos compartir con terceros, generar un nuevo conjunto de datos que conserva las características informacionales del conjunto origen pero que no permite recomponer los datos originales a partir de los creados artificialmente*"[10].

5.3. Protección desde el diseño y por defecto. La solidez técnica como requisito clave

La aproximación basada en el riesgo constituye un principio medular del sistema de garantías aplicable a la protección de datos en el RGPD y se articula a través de diversas instituciones vinculadas a la misma. De manera singular, el artículo 25 RGPD bajo el epígrafe "Protección de datos desde el diseño y por defecto", incorpora a la normativa de protección

10 https://manueldelgado.com/que-son-los-datos-sinteticos/. También, https://www.ciospain.es/big-data/que-son-los-datos-sinteticos-datos-generados-para-ayudar-a-tu-estrategia-de-ia.

de datos la práctica de considerar los requisitos de privacidad desde las primeras etapas del diseño de productos y servicios[11].

El Considerando (69) del RIA viene a subrayar esta idea:

> "*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*".

La RIA incorpora nuevas ideas y dimensiones de las inicialmente proyectadas por el RGPD. Así, su artículo 11.1 establece una garantía que posee un claro aire de familia con los sistemas de protección desde el diseño y por defecto, al establecer que "*la documentación técnica de un sistema de IA de alto riesgo se preparará antes de su introducción en el mercado o puesta en servicio, y se mantendrá actualizada*". Esta documentación, se añade, se redactará de modo que demuestre que el sistema de IA de alto riesgo cumple los requisitos establecidos normativamente y proporcionará a las autoridades nacionales competentes y los organismos notificados toda la información que necesiten para evaluar si el sistema de IA de que se trate cumple con dichas exigencias.

Un factor ligado al anterior es la solidez técnica de estos sistemas. Como concreta el Considerando (75) de la RIA:

> "*La solidez técnica es un requisito clave para los sistemas de IA de alto riesgo, que deben ser resilientes a los riesgos asociados a las limitaciones del sistema (p. ej., errores, fallos, incoherencias o situaciones inesperadas), así como a acciones maliciosas que pueden poner en peligro su seguridad y dar lugar a conductas perjudiciales o indeseables por otros motivos. La incapacidad de protegerlos frente a estos riesgos podría tener consecuencias para la seguridad o afectar de manera*

11 Nos remitimos para la definición de su alcance a AEPD, *Guía de Privacidad desde el Diseño*, Madrid, AEPD, 2019. https://www.aepd.es/sites/default/files/2019-11/guia-privacidad-desde-diseno.pdf y por defecto https://www.aepd.es/sites/default/files/2020-10/guia-proteccion-datos-por-defecto.pdf.

negativa a los derechos fundamentales, por ejemplo, debido a la adopción de decisiones equivocadas o a que el sistema de IA en cuestión genere una información de salida errónea o sesgada. Los usuarios del sistema de IA deben tomar medidas para garantizar que la posible compensación entre solidez y precisión no conduzca a resultados discriminatorios o negativos para subgrupos minoritarios".

5.4. Auditorías algorítmicas

Igualmente, se establece como garantía la necesidad de contar con un sistema de gestión de riesgos. Según detalla el artículo 9 de la RIA: "*Se establecerá, implantará, documentará y mantendrá un sistema de gestión de riesgos asociado a los sistemas de IA de alto riesgo*". Este sistema consistirá en un proceso dinámico que se llevará a cabo durante todo el ciclo de vida de un sistema de IA de alto riesgo y que requerirá actualizaciones sistemáticas periódicas. Los sistemas de IA de alto riesgo serán sometidos a pruebas destinadas a determinar cuáles son las medidas de gestión de riesgos más adecuadas y comprobarán que los sistemas de IA de alto riesgo funcionan de un modo adecuado para su finalidad prevista y cumplen los requisitos previstos normativamente. Y es que, como explica el Considerando (125) del RIA: "*Dada la complejidad de los sistemas de IA de alto riesgo y los riesgos asociados a ellos, es importante desarrollar un procedimiento adecuado de evaluación de la conformidad de los sistemas de IA de alto riesgo en el que participen organismos notificados, denominado «evaluación de la conformidad de terceros». No obstante, habida cuenta de la experiencia actual de los profesionales que realizan la certificación previa a la comercialización en el campo de la seguridad de los productos y de la distinta naturaleza de los riesgos implicados, procede limitar, al menos en la fase inicial de aplicación del presente Reglamento, el alcance de las evaluaciones externas de la conformidad a los sistemas de IA de alto riesgo que no están asociados a productos*".

Las auditorías algorítmicas cobran, por tanto, una indudable importancia. El fin de una auditoría es identificar o anticipar errores, riesgos o amenazas (actuales o potenciales) y ayudar a corregirlos[12]. Esto puede darse en cualquiera de las fases del desarrollo del sistema, tanto en su diseño y puesta en marcha, como en la fase de funcionamiento y posteriormente a él. Los supuestos más problemáticos se plantean cuando resulta difícil "replicar" la aplicación de la norma sin el algoritmo, para compararla con el resultado que éste arroja. En estos casos el control puede materializarse a través de sistemas de auditoría inversa y usando datos administrativos, entrevistas y escrutando tanto informes como

12 GARCÍA HERRERO, J., "¿Cómo auditar un algoritmo?", entrada de 28 de febrero de 2017 en el blog https://jorgegarciaherrero.com/auditar-algoritmo/.

guiones del diseño del modelo para después recopilar resultados a escala contrastando, de este modo, la regularidad del funcionamiento y aplicación del algoritmo o sistema. Esto es que ese sistema aplicado a una concreta persona y proyectado al resto de personas funcionará exactamente igual que ha funcionado conmigo (o sea que las mismas premisas, sobre el set de datos de otra persona igual o análoga a mí, arrojaría el mismo resultado).

5.5. Principio de humano al mando

La "reserva de humanidad", como se la ha calificado[13], la idea del humano al mando (*human-in-the-loop*, *human-on-the-loop*, o *human-in-comand*)[14] constituye un soporte fundamental en el diseño de los nuevos sistemas de IA y en dicha dirección ya caminaba el artículo 22.3 RGPD al establecer que: "*En los casos a que se refiere el apartado 2, letras a) y c), el responsable del tratamiento adoptará las medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado, como mínimo el derecho a obtener intervención humana por parte del responsable, a expresar su punto de vista y a impugnar la decisión*".

El Acuerdo Marco sobre Digitalización que se presenta como un acuerdo "*para apoyar el éxito de la transformación digital de la economía europea y gestionar sus grandes repercusiones en los mercados laborales, el mundo del trabajo y la sociedad en general*" ha subrayado la garantía que supone el control humano.

El PIA da un paso más en esta dirección y precisa que los sistemas de IA de alto diesgo deberán estar diseñados de forma que puedan ser vigilados de manera efectiva por personas físicas para prevenir o reducir al mínimo los riesgos derivados de su uso. El objetivo de la vigilancia humana será, conforme establece el artículo 14 de la RIA, prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando un sistema de IA de alto riesgo se utiliza conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persisten a pesar de aplicar otros requisitos establecidos en el presente capítulo.

Las medidas a adoptar permitirán que las personas a quienes se encomiende la vigilancia humana puedan, en función de las circunstancias:

13 PONCE, J., *Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico*, Revista General de Derecho Administrativo, 2019, nº 50.

14 VIDA FERNÁNDEZ, J., "*Los retos de la regulación de la inteligencia artificial: algunas aportaciones desde la perspectiva europea*", en T. DE LA QUADRA-SALCEDO, J.L. PIÑAR MAÑAS, (Dir.) *Sociedad Digital y Derecho*. Madrid. BOE, 2018, p. 218.

a) ser conscientes de las capacidades pertinentes y limitaciones del sistema de IA de alto riesgo, entenderlas suficientemente y controlar debidamente su funcionamiento, de modo que puedan detectar indicios de anomalías, problemas de funcionamiento y comportamientos inesperados y ponerles solución lo antes posible;
b) ser conscientes de la posible tendencia a confiar automáticamente o en exceso en la información de salida generada por un sistema de IA de alto riesgo («sesgo de automatización»), en particular con aquellos sistemas que se utilizan para aportar información o recomendaciones con el fin de que personas físicas adopten una decisión;
c) interpretar correctamente la información de salida del sistema de IA de alto riesgo, teniendo en cuenta en particular las características del sistema y las herramientas y los métodos de interpretación disponibles;
d) decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o desestimar, invalidar o revertir la información de salida que este genere; e
e) intervenir en el funcionamiento del sistema de IA de alto riesgo o interrumpir el sistema accionando un botón específicamente destinado a tal fin o mediante un procedimiento similar que permita detener el sistema en un estado seguro, excepto si la injerencia humana eleva los riesgos o repercute negativamente en el rendimiento habida cuenta del estado actual de la técnica generalmente reconocido.

6. LAS EMPRESAS COMO RESPONSABLES DEL DESPLIEGUE DE LOS SISTEMAS DE ALTO RIESGO Y EL PAPEL DE CONTROL DE LOS REPRESENTANTES DE LOS TRABAJADORES

Define el artículo 3.4 del RIA como "responsable del despliegue" a "*una persona física o jurídica, o autoridad pública, órgano u organismo que utilice un sistema de IA bajo su propia autoridad, salvo cuando su uso se enmarque en una actividad personal de carácter no profesional*". La empresa es, por tanto, el implementador natural en el ámbito de las relaciones laborales.

La empresa asume, por tanto, cuando decide incorporar el uso de sistemas de IA, un conjunto de responsabilidades y obligaciones. Los responsables del despliegue de un sistema de IA de alto riesgo desempeñan un papel crítico a la hora de garantizar que se protegen los derechos fundamentales, complementando las obligaciones del proveedor al desarrollar el sistema de IA. Son los más indicados para comprender cómo se utilizará concretamente el sistema de IA de alto riesgo y pueden, por lo tanto, identificar potencia-

les riesgos significativos que no se previeron en la fase de desarrollo, debido a un conocimiento más preciso del contexto de uso y de las personas o los grupos de personas que podrían verse afectados.

Además de las obligaciones aplicables con carácter general a los sistemas de IA de alto riesgo, la Sección 3 de este Título III, titulado "Obligaciones de los proveedores y responsables del despliegue de los sistemas de IA de alto riesgo y de otras partes", establece las obligaciones que, en relación con los sistemas IA, deben cumplir determinados actores que intervienen en su diseño, fabricación, comercialización y uso.

En concreto, los responsables del despliegue de los sistemas de IA que incorporen estas técnicas se encuentran obligados a cumplir con un ineludible presupuesto [art. 26.7 RIA y Considerando (92)]: "*Antes de poner en servicio o utilizar un sistema de IA de alto riesgo en el lugar de trabajo, los responsables del despliegue que sean empleadores informarán a los representantes de los trabajadores y a los trabajadores afectados de que estarán expuestos a la utilización del sistema de IA de alto riesgo. Esta información se facilitará, cuando proceda, con arreglo a las normas y procedimientos establecidos en el Derecho nacional y de la Unión y conforme a las prácticas en materia de información a los trabajadores y sus representantes*". Una exigencia esencial si atendemos a los términos del Considerando (92) que considera que "*este derecho de información es accesorio y necesario para el objetivo de protección de los derechos fundamentales que subyace al presente Reglamento* (...)" y, además, se impone "*incluso aunque no se cumplan las condiciones de las citadas obligaciones de información o de información y consulta previstas en otros instrumentos jurídicos*" (la referencia se hace a la Directiva 2002/14/CE).

La negociación colectiva está comenzando a concretar el alcance de las referidas exigencias. Buena muestra de ello es el XXI Convenio colectivo general de la industria química y las previsiones de su artículo 10. En el mismo se establece, por ejemplo, el derecho de la representación legal de los trabajadores a ser informada sobre los sistemas concretos de IA que se pretenden implantar, especificando sus objetivos y razones. Se exige además que se evalúen los posibles impactos en el empleo, incluyendo la identificación de los puestos de trabajo que puedan verse afectados y los cambios potenciales en las condiciones laborales. Los representantes de las personas trabajadoras tendrán un plazo de 10 días para elaborar un informe con sus observaciones, el cual deberá ser valorado por la dirección de la empresa antes de implementar la IA.

De igual modo, el XXV Convenio colectivo del sector de la banca establece específicos derechos ante la IA. En concreto, en su artículo 80.5 se establece que las personas trabajadoras tienen derecho a no ser objeto de decisiones basadas única y exclusivamente en variables automatizadas, salvo en aquellos supuestos previstos por la ley, así como derecho a la no discriminación en relación con las decisiones y procesos, cuando ambos estén basados únicamente en algoritmos, pudiendo solicitar, en estos supuestos, el concurso e intervención

de las personas designadas a tal efecto por la Empresa, en caso de discrepancia. Además, las Empresas deben informar a los representantes de las personas trabajadoras sobre el uso de la analítica de datos o los sistemas de IA cuando los procesos de toma de decisiones en materia de recursos humanos y relaciones laborales se basen, exclusivamente en modelos digitales sin intervención humana. Dicha información, como mínimo, abarcará los datos que nutren los algoritmos, la lógica de funcionamiento y la evaluación de los resultados.

Una exigencia que se ve reforzada por el artículo 86 del RIA que reconoce el derecho a recibir una explicación individualizada de aquellos usos de la IA que afecten a una persona trabajadora. Esto supone que "*toda persona que se vea afectada por una decisión que el responsable del despliegue adopte basándose en los resultados de un sistema de IA de alto riesgo que figure en el anexo III, con excepción de los sistemas enumerados en su punto 2, y que produzca efectos jurídicos o le afecte considerablemente del mismo modo, de manera que considere que tiene un efecto perjudicial para su salud, su seguridad o sus derechos fundamentales, tendrá derecho a obtener del responsable del despliegue explicaciones claras y significativas acerca del papel que el sistema de IA ha tenido en el proceso de toma de decisiones y los principales elementos de la decisión adoptada*".

La Sección 5 de este Título III del RIA ("Normas, evaluación de conformidad, certificados, registro") incorpora la exigencia de autocertificación de los sistemas de alto riesgo, proceso mediante el cual los proveedores de dichos sistemas evalúan y declaran que cumplen con los requisitos y estándares establecidos normativamente. El fundamento de dicha exigencia se extrae del Considerando (125) del RIA que viene a precisar que: "*Dada la complejidad de los sistemas de IA de alto riesgo y los riesgos asociados a ellos, es importante desarrollar un sistema adecuado para el procedimiento de evaluación de la conformidad de los sistemas de IA de alto riesgo en el que participen organismos notificados, denominado «evaluación externa de la conformidad». No obstante, habida cuenta de la experiencia actual de los profesionales que realizan la certificación previa a la comercialización en el campo de la seguridad de los productos y de la distinta naturaleza de los riesgos implicados, procede limitar, al menos en la fase inicial de aplicación del presente Reglamento, el alcance de las evaluaciones externas de la conformidad a los sistemas de IA de alto riesgo que no están asociados a productos. En consecuencia, el proveedor es quien, por norma general, debe llevar a cabo la evaluación de la conformidad de dichos sistemas bajo su propia responsabilidad, con la única excepción de los sistemas de IA que están destinados a utilizarse para la biometría*".

Establece el artículo 43.2 RIA que: "*En el caso de los sistemas de IA de alto riesgo mencionados en los puntos 2 a 8 del anexo III, los proveedores se atendrán al procedimiento de evaluación de la conformidad fundamentado en un control interno a que se refiere el anexo VI, que no contempla la participación de un organismo notificado*".

De este modo, los sistemas de IA laborales quedan incluidos dentro del radio de la autocertificación. Una exigencia que lleva consigo para los proveedores de estos sistemas el

cumplimiento de ciertos requisitos, a saber: (i) comprobar que el sistema de gestión de la calidad establecido reúne los requisitos establecidos en el artículo 17 del RIA; (ii) examinar la información de la documentación técnica para evaluar la conformidad del sistema de IA con los requisitos esenciales pertinentes establecidos en el capítulo III, sección 2; y (iii) comprobar que el proceso de diseño y desarrollo del sistema de IA y la vigilancia poscomercialización del mismo a que se refiere el artículo 72 son coherentes con la documentación técnica. Con todo, el modelo de autocertificación plantea numerosas dudas y hace más necesaria que nunca la consolidación de empresas tecnológicamente responsables.

7. UN SISTEMA DUAL DE RESPONSABILIDAD PARA LA IA

El sistema de responsabilidades por el uso de la IA se encuentra integrado por dos subsistemas de diferente alcance y contenido: de un lado, el subsistema punitivo, basado en sanciones administrativas y asentado sobre una pluralidad de agentes que tutelan tal actividad de control; y, de otro, el subsistema reparador, asentado en la lógica que proporciona el derecho de daños[15].

Las dificultades a la hora de aplicar y hacer efectivas las técnicas de control y sanción son múltiples. La autonomía y el carácter opaco de los sistemas de IA, es decir, la dificultad de comprender y explicar cómo han tomado sus decisiones, por las propias características de la tecnología que utilizan, como sucede en el caso de algunos métodos de aprendizaje profundo (*deep learning*), dificulta de modo especial, tanto la proyección de régimen sancionador, como "*la prueba no solo de la culpa sino también de la relación de causalidad*"[16]. La interconectividad es otra de las características distintivas de los productos que incorporan sistemas de IA que pueden plantear numerosos problemas a la hora de detectar y controlar su alcance.

7.1. La responsabilidad sancionadora administrativa

El RIA, en su capítulo XII, presenta un marco sancionador e impone el mandato a los Estados miembros de desarrollar el régimen de sanciones y otras medidas de ejecución, como advertencias o medidas no pecuniarias, aplicables a las posibles infracciones. Este

15 Ampliamente, MERCADER UGUINA, J.R., *Sistema de responsabilidades por el uso de la inteligencia artificial. Un enfoque integral.* Labos, 2024, Vol.1, Extra 1, pp. 211-227.

16 M. MARTÍN CASALS, *Las propuestas de la Unión Europea para regular la responsabilidad civil por los daños causados por sistemas de inteligencia artificial*, InDret: Revista para el Análisis del Derecho, 2023, nº 2, p. 75.

régimen sancionador resulta imprescindible para asegurar el cumplimiento de las obligaciones que establece el RIA a los proveedores, responsables del despliegue, importadores y distribuidores, representantes autorizados, potenciales proveedores en espacios de pruebas y personas afectadas por sistemas de IA establecidos o ubicados en la Unión Europea. Como expresa el Considerando (168) del RIA, "*se debe poder exigir el cumplimiento del presente Reglamento mediante la imposición de sanciones y otras medidas de ejecución*". Sus efectos parecen, a priori, demoledores.

A título de ejemplo, el artículo 99.3 RIA establece que: "*El no respeto de la prohibición de las prácticas de IA a que se refiere el artículo 5 estará sujeto a multas administrativas de hasta 35.000.000 EUR o, si el infractor es una empresa, de hasta el 7 % de su volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior*". Por su parte, el artículo 99.4 establece que: "*El incumplimiento por parte de un sistema de IA de cualquiera de las disposiciones que figuran a continuación en relación con los operadores o los organismos notificados, distintas de los mencionados en el artículo 5, estará sujeto a multas administrativas de hasta 15.000.000 EUR o, si el infractor es una empresa, de hasta el 3 % de su volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior*", resultando a nuestros efectos relevante el apartado e) que proyecta los anteriores efectos sobre "*las obligaciones de los responsables del despliegue con arreglo al artículo 26*".

El Anteproyecto de Ley para el Buen Uso y la Gobernanza de la Inteligencia Artificial, aprobado por el Consejo de Ministros del pasado 11 de marzo de 2025, consta de treinta y siete artículos agrupados en cuatro capítulos, además de dos disposiciones adicionales, cuatro disposiciones finales y un anexo. Su Capítulo IV está dedicado a establecer, a nivel nacional, el régimen sancionador administrativo derivado del RIA, cumpliendo así la obligación que impone a los Estados miembros el artículo 99 RIA.

El referido texto clasifica las infracciones en muy graves, graves o leves, determinando las correspondientes sanciones. Como novedad, introduce que, en el caso de las infracciones muy graves por prácticas de IA prohibidas y en las infracciones en que un sistema de IA haya causado un incidente grave (en los términos del artículo 3.49 del RIA), la sanción incluirá adicionalmente la retirada del producto, la desconexión o la prohibición del sistema de IA en el ámbito territorial de la autoridad de vigilancia del mercado sancionadora.

Dado que las sanciones máximas pueden quedar limitadas por el volumen anual de negocio de la entidad infractora, se considera que, cuando la empresa directamente responsable de la infracción es parte de un grupo de empresas, será el volumen de negocios del grupo el que fije el límite superior de la sanción.

En cuanto a la tipificación, contempla un catálogo de infracciones muy graves que engloba tanto las relativas a las prácticas de IA prohibidas recogidas en el artículo 5 del RIA (art. 14.1) como las relativas a los operadores de sistemas de alto riesgo (art. 14.2).

También se tipifican las infracciones graves, que divide según sean aplicables a cualquier operador de sistemas de IA (art. 15), a los proveedores y proveedores potenciales de sistemas de IA (art. 16.1), a los proveedores de sistemas de IA de alto riesgo (art. 16.2), al proveedor o proveedor potencial de sistemas de IA de alto riesgo (art. 16.3), a los representantes autorizados de sistemas de IA (art. 17), a los importadores de sistemas de IA (art. 18), a los distribuidores de sistemas de IA (art. 19), a los responsables del despliegue de sistemas de IA y de los responsables del despliegue de sistemas de IA de alto riesgo (art. 20) y a los organismos notificados (art. 21).

Se tipifican, por último, las infracciones leves por parte de cualquier operador (art. 22), por parte de los proveedores de sistemas de IA (art. 23) por parte de los importadores de sistemas de IA (art. 24), por parte de los responsables del despliegue de sistemas de IA (art. 25) y por parte de los organismos notificados (art. 26).

Sobre dicha base, se establece el mecanismo de graduación de las sanciones, que deberá tener en cuenta la adecuación entre la sanción y el hecho constitutivo de la infracción, considerando especialmente su repercusión y su trascendencia por lo que respecta a la salud y seguridad de las personas y a sus derechos fundamentales.

Para el caso de que el sujeto infractor sea una pyme y la infracción no fuera muy grave, se establece un régimen de requerimiento por el que la resolución podrá terminar el procedimiento sancionador con la adopción por el responsable, en plazo y forma, de las medidas correctoras pertinentes y la indemnización total de los daños y perjuicios causados, en su caso.

Finalmente, se establece un plazo de prescripción de un año para las infracciones leves, tres años para las graves y de cinco años para las muy graves, a contarse desde el día en que la infracción se hubiera cometido.

7.2. Responsabilidad civil por daños provocados por sistemas de IA

Sin perjuicio de la sanción que se pudiera imponer, el infractor quedará obligado a la reposición de la situación alterada por el mismo a su estado originario, así como a la indemnización de los daños y perjuicios causados, que podrán ser determinados por la autoridad competente.

El texto de la RIA aprobado por el Consejo no introduce disposiciones concretas sobre responsabilidad civil. Impone, ciertamente, un gran número de obligaciones a los así llamados «operadores» de sistemas de IA, expresión que incluye una gran variedad de sujetos: proveedores, distribuidores e importadores de sistemas de IA, responsables del despliegue de sistemas de IA, fabricantes de productos que introduzcan en el mercado o pongan en servicio un sistema de IA junto con su producto y con su propio nombre o mar-

ca; y representantes autorizados de los proveedores de sistemas o modelos de IA. La RIA, sin embargo, se limita a regular la supervisión gubernamental de la actividad de dichos operadores y a imponerles régimen sancionatorio administrativo por la transgresión de aquellas obligaciones en los términos que hemos analizado.

En 2022 se adoptaron dos propuestas de Directivas, con diferente grado de especialización, llamadas a regular en un futuro próximo la responsabilidad por daños derivados de la inteligencia artificial, aunque su trayectoria o tramitación ha sido muy desigual. Por un lado, la Propuesta de Directiva del Parlamento europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos, 28 de septiembre de 2022 y que ha visto la luz bajo la forma de Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, del 23 de octubre de 2024 (en adelante, DPD), y, por otro, una Propuesta de 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) (COM/2022/496 final) (en adelante, "PD-RCIA"). El programa de trabajo de la Comisión Europea para 2025 *(Commission work programme 2025 - Moving forward together: A Bolder, Simpler, Faster Union)* ha procedido a retirar este proyecto dado que no se considera previsible un acuerdo sobre el texto, por lo que la Comisión evaluará si debe presentarse otra propuesta u optar por otro tipo de enfoque.