Sistema de responsabilidades por el uso de la inteligencia artificial. Un enfoque integral

System of responsibilities for the use of artificial intelligence. A comprehensive approach

Jesús R. Mercader Uguina
Catedrático de Derecho del Trabajo y de la Seguridad Social
Universidad Carlos III de Madrid
ORCID ID: 0000-0001-6301-6788

doi: 10.20318/labos.2024.9038

Resumen: A día de hoy, el sistema de responsabilidad por el uso de sistemas de Inteligencia Artificial sigue reposando, en lo esencial, en los mecanismos que, con matizaciones, adaptaciones y ajustes, han servido a tales fines durante los períodos de cambio y transformación tecnológica que a lo largo de la historia se han sucedido hasta el presente: un modelo basado en sanciones. De este modo, el sistema de responsabilidades por el uso de la Inteligencia Artificial 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, que pretende construirse desde la lógica que proporciona el derecho de daños. Las dificultades a la hora de aplicar y hacer efectivas las técnicas de control y sanción en un escenario incierto sobre los límites y usos de estas técnicas hacen necesario profundizar en su estudio.

Palabras clave: Sistema de responsabilidad, inteligencia artificial, sanción, reparación.

Abstract: The objective of this article is to provide an overview of the regulatory model that has inspired the new regulations on AI and its relations with the specific regulatory framework on occupational risk prevention. The aim is to delve deeper into the general obligations of AI system providers and the specific obligations, in terms of occupational risk prevention, of such providers and of the employer who incorporates such systems into the workplace. An analysis is also made of the current regime regarding liability for damages, with the aim of raising the range of possibilities of claims by the worker who suffers damages derived from AI and its relationship with the general regime of liability of the employer.

Keywords: System of liability, Artificial Intelligence systems, sanctions, restorative subsystem.

1. El último lado del "triángulo de oro" del RIA: La responsabilidad

La antropóloga Mary Douglas sostenía que las sociedades se definen a sí mismas por el modo en que caracterizan y gestionan sus riesgos. El desarrollo de la IA está asociado, de manera inescindible, a los múltiples riesgos que conlleva su incorporación a la dinámica económica y social en su más amplio sentido. Su control constituye la clave esencial en el que se asienta el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 ("RIA"). Y es que, si bien la incorporación de esta nueva realidad conlleva notables incertidumbres, se trata, en última instancia, de construir una incertidumbre medible.

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.

El riesgo está asociado, de manera inescindible, a los cambios que lleva consigo la incorporación de los nuevos sistemas de IA. Así lo pone de manifiesto la RIA que articula su regulación, precisamente, sobre una aproximación desde la idea de "riesgo", definida como "la combinación de la probabilidad de que se produzca un daño y la gravedad de dicho daño". El Reglamento establece una jerarquía de riesgos en función del uso de la IA y sobre las categorías detectadas, establece una serie de obligaciones cuyas proyecciones sobre lo laboral resultan más que evidentes.

En este ámbito quedan directamente proscritos los sistemas de reconocimiento de emociones. 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). 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).

Pero la pieza esencial 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). 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 (…)".

El RIA 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 un 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 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). A ellas se unen el poderoso régimen de obligaciones para todos los sujetos actuantes en el ecosistema de la IA (art. 26-28 RIA).

La última pieza del sistema, el último lado del triángulo del RIA se centra en el régimen de responsabilidades.

2. Un sistema dual de responsabilidad para la IA

A día de hoy, transformaciones tecnológicas al margen, el sistema de responsabilidad por el uso de sistemas de IA sigue reposando, en esencial, en los mecanismos que, con matizaciones, adaptaciones y ajustes, han servido a tales fines durante los períodos de cambio y transformación tecnológica que a lo largo de la historia se han sucedido hasta el presente: un modelo basado en sanciones.

La sanción en sentido amplio es, según Guasp[1], la consecuencia que el ordenamiento jurídico adopta para aquellos supuestos en que se ha producido un resultado que desaprueba o rechaza. Los dos grandes tipos de sanciones son las civiles y las punitivas. Las primeras tratan de reparar el efecto producido por la infracción; las segundas, entre las que se encuentran las penas y las sanciones administrativas, son "sanciones artificiales o males intrínsecos y heterogéneos con la infracción", en los que ya no se trata de resarcir o reparar, sino de castigar para retribuir o para prevenir. Para abreviar hablaremos aquí de sanciones punitivas (penales y administrativas) y resarcitorias. Por otra parte, hay que tener en cuenta que cuando se trata de infracciones que son objeto de una sanción punitiva es posible que, junto a la lesión del interés general que justifica la imposición de una sanción de este tipo –penal o administrativa–, se produzca una lesión específica para una determinada persona, que resulta concretamente perjudicada por la acción ilícita, y en estos casos se aplicarán las dos sanciones (la punitiva y la resarcitoria).

De este modo, 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 proporcional el derecho de daños.

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" [2]. 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.

3. La responsabilidad sancionadora administrativa: Pluralidad de agentes

Por lo que hace al primero de los sistemas, el RIA parte de la funcionalidad de las sanciones como instrumentos para la efectividad de las garantías y obligaciones que el mismo impone. Señala en el Considerando (168) que: "*Se debe poder exigir el cumplimiento del presente Reglamento mediante la imposición de sanciones y otras medidas de ejecución. Los Estados miembros deben tomar todas las medidas necesarias para garantizar que se apliquen las disposiciones del presente Reglamento, incluso estableciendo sanciones efectivas, proporcionadas y disuasorias para las infracciones, lo que incluye respetar el principio de non bis in ídem*". Por otro lado, desde una lógica que favorece la coordinación interadministrativa

[1] J. GUASP, *Derecho*, Madrid, Gráficas Hergón, 1971, pp. 522-524.

[2] 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.

y el respeto al reparto de competencias entre los distintos órganos que, también, desde distintas perspectivas están llamados a actuar en este ámbito, el Considerando (157) establece que: "*El presente Reglamento se entiende sin perjuicio de las competencias, funciones, poderes e independencia de las autoridades u organismos públicos nacionales pertinentes que supervisan la aplicación del Derecho de la Unión que protege los derechos fundamentales, incluidos los organismos de igualdad y las autoridades de protección de datos*".

El RIA, a pesar de su ánimo de ser una ley unificadora de un régimen general aplicable a los sistemas de IA, es una norma que recibirá una aplicación desde diversas perspectivas, principalmente la innovación y la protección de los derechos fundamentales, y por organismos con competencias y funciones diversas tanto a nivel europeo (Comité Europeo de Inteligencia Artificial, Oficina de la IA, Comité Europeo de Protección de Datos) como a nivel estatal (autoridad de vigilancia del mercado, autoridad notificantes, autoridades de protección de datos personales, Inspección de Trabajo y Seguridad Social).

La implantación de la IA se encuentra aún en un estadio incipiente y, tanto el grado de litigiosidad en relación con los daños causados por su uso, como las actuaciones que en esta materia se están desarrollando por las autoridades públicas, es escaso. Sin embargo, a medida que su uso vaya generalizándose es previsible que la conflictividad y los problemas a la hora de resolver los concretos espacios en los que deban moverse los órganos responsables de los distintos ámbitos vayan creciendo en complejidad y dificultad. Procede, por ello, realizar un breve examen de los distintos agentes con competencias en relación con cuestiones que directa o indirectamente pueden afectar a los sistemas de IA y, al hilo del mismo, analizar el arsenal sancionador asociado a cada ámbito de competencia.

3.1. La Inspección de Trabajo y Seguridad Social

La arquitectura del actual Derecho digital del Trabajo se está adquiriendo reconstruyendo sobre nuevos pilares de sustentación. Los algoritmos y la Inteligencia Artificial (IA) se están convirtiendo en son los instrumentos maestros piezas maestras en desde los las que la empresa post-material delega funciones centrales de su poderse lanza hacia una transformación radical de sus tradicionales estructuras. Por el momento son las grandes empresas las que implementan de modo generalizado estos modelos, pero es cuestión de tiempo que los mismos se expandan al resto. Esta "gran delegación empresarial", se produce, además, a un ritmo incontenible Pero el ritmo de expansión del en el que el futuro se convierte en pasado a enorme velocidad[3].

La Inteligencia Artificial aporta sistemas que permiten la dirección y control de la prestación de servicios, ya sea en la asignación de tareas, en el control de los tiempos

[3] De estas y otras muchas cuestiones vinculadas con este conjunto de transformaciones, me he ocupado en mi monografía, J.R. MERCADER UGUINA, *Algoritmos e inteligencia artificial en el derecho digital del trabajo,* Valencia, Tirant lo Blanch, 2022. Posteriormente, en *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). También en "*Los usos de alto riesgo en el ámbito laboral de la IA y la autocertificación*", El Foro de Labos, 9 de mayo de 2024.

de trabajo, en el ejercicio de funciones de control y supervisión, o en el establecimiento de métodos de medición del rendimiento. Esto es así, por ejemplo, en la prestación de servicios a través de plataformas digitales, en las que un algoritmo gestiona la recepción de solicitudes de servicio y toma las decisiones necesarias para gestionar su prestación. Pero, en general, cualquier profesional, empresario, compañía, que incorpore un sistema de IA para el desarrollo de sus actividades, prestación de servicios, relaciones comerciales se encuentra sometido al régimen de control de las autoridades administrativas y, en particular, de la Inspección de Trabajo.

En los últimos años, la Inspección de Trabajo y Seguridad Social ha tenido que realizar actuaciones inspectoras en estos nuevos entornos tecnológicos. Así, en el año 2017 se llevaron a cabo las primeras inspecciones en empresas que prestaban servicios de reparto a través de plataformas digitales. El eje 3 del Plan Estratégico de la Inspección de Trabajo y Seguridad Social 2021-2023, tiene por objeto fortalecer y modernizar el sistema de la Inspección de Trabajo y Seguridad Social para mejorar la capacidad del servicio prestado a los ciudadanos. El objetivo 30 prevé mejorar la planificación de las actuaciones inspectoras utilizando las herramientas más avanzadas de IA. En concreto, en la actuación 30.3 referida a una estrategia más activa para reducir los comportamientos de incumplimiento y fraude, pero, en ningún caso, se prevén expresas actuaciones dirigidas a controlar, por el momento, el uso de los sistemas de IA por las empresas.

No obstante, los espacios en los que proyectará su actividad la Inspección de Trabajo en su conexión con el uso por las empresas de sistemas de IA son sin duda múltiples. La Ley de infracciones en laboral y de prevención de riesgos laborales deberá, a no tardar, incorporar nuevos tipos sancionadores que contemplen las también nuevas situaciones. Particular interés tendrán estos tipos de infracciones graves o muy graves, que, además, de dar lugar a responsabilidad administrativa podrán generar acciones de responsabilidad civil.

De momento, lo que si podemos observar es un sistema de sanciones que, comparado con el que se aprecia en materia de protección de datos e IA resulta, a priori, modesto. Así, la falta muy grave por vulneración del derecho a la intimidad y la consideración debida a la dignidad de los trabajadores pueden alcanzar multas de hasta 187.515€. Así, en caso, en los que la empresa implementara de manera agresiva e intrusiva mecanismos de vigilancia al trabajador la referida multa resultaría la máxima posible. Por su parte, el montante de las multas por infracciones en materia de prevención de riesgos es más elevada: las graves, en su grado máximo, de 24.586 a 49.180 euros; las muy graves, en su grado mínimo, de 49.181 a 196.745 euros; en su grado medio, de 196.746 a 491.865 euros; y en su grado máximo, de 491.866 a 983.736 euros.

3.2. La Agencia Española de Protección de datos

La proyección laboral de la competencia de las autoridades de protección de datos en el control de los sistemas de IA, la podemos encontrar en la Resolución legislativa del Parlamento Europeo, de 24 de abril de 2024, sobre la propuesta de Directiva del Parlamento

Europeo y del Consejo relativa a la mejora de las condiciones laborales en el trabajo en plataformas digitales ("PD-CLPD").

La misma viene a poner de manifiesto la necesaria conexión entre autoridades laborales y de protección de datos como queda subrayado en su Considerando (65): "*Los sistemas automatizados de supervisión o de toma de decisiones utilizados en el contexto del trabajo en plataformas implican el tratamiento de datos personales de las personas que realizan trabajo en plataformas y afectan a las condiciones laborales y a los derechos de los trabajadores de plataformas que se encuentran en este grupo de personas, lo que plantea problemas relativos a la legislación en materia de protección de datos y a otros ámbitos, como la legislación laboral. Así, las autoridades de control de la protección de datos y otras autoridades competentes deben cooperar, en particular en el ámbito transfronterizo, en el cumplimiento efectivo de la presente Directiva mediante el intercambio de información pertinente, entre otros medios, sin menoscabo de la independencia de las autoridades de control de la protección de datos*".

Sobre esta base el art. 24.1 del PD-CLPD establece que: "*las autoridades de control responsables de supervisar la aplicación del RGPD también serán responsables de supervisar y garantizar la aplicación de los artículos 7 a 11 de la presente Directiva en lo que respecta a las cuestiones de protección de datos, de conformidad con las disposiciones pertinentes de los capítulos VI, VII y VIII del Reglamento (UE) 2016/679. El límite máximo para las multas administrativas a que se refiere el artículo 83, apartado 5, de dicho Reglamento será aplicable a las infracciones de los artículos 7 a 11 de la presente Directiva*".

Según el art.83.5 RGPD, se sancionan con multas administrativas de 20.000. 000 € como máximo o, si es una empresa, de una cuantía equivalente al 4 % como máximo del volumen de negocio total anual global del ejercicio financiero anterior, eligiendo la de mayor cuantía, los atentados contra los principios básicos para el tratamiento, incluidos los requisitos para el consentimiento (principios relativos al tratamiento, licitud del mismo, condiciones para el consentimiento, y el tratamiento en las categorías especiales de datos personales) o los derechos de los afectados (transparencia y modalidades, información y acceso a datos personales, derecho de rectificación y supresión, derecho de oposición y decisiones individuales automatizadas).

Por su parte, el art. 24.2 del PD-CLPD precisa que "*las autoridades a que se refiere el apartado 1 y otras autoridades nacionales competentes cooperarán, cuando proceda, en la garantía de cumplimiento de la presente Directiva, dentro del ámbito de sus competencias respectivas, en particular cuando surjan cuestiones sobre las consecuencias de los sistemas automatizados de supervisión o de toma de decisiones para las personas que realizan trabajo en plataformas. A tal fin, las mencionadas autoridades intercambiarán entre ellas la información pertinente, incluida la obtenida en el contexto de inspecciones o investigaciones, bien previa solicitud, bien por propia iniciativa*".

En atención a lo previsto en el Considerando (11) del Reglamento (UE) 2016/679, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos ("RGPD"), la protección efectiva de los datos personales en la Unión exige, entre otras consideraciones, que las infracciones a las obligaciones previstas se castiguen con sanciones equivalentes (económicas

o no). En definitiva, que la vulneración de este derecho fundamental no quede impune en ninguno de los Estados miembros. En el mismo sentido se pronuncian los considerandos (148) y (150) RGPD, al indicar que cualquier infracción debe ser castigada con sanciones, incluidas multas administrativas, sin perjuicio de que se atienda a las circunstancias concretas de cada caso. Una de las características distintivas del RGPD es su enfoque integral, que abarca tanto la prevención como la sanción y la indemnización (art. 82 RGPD).

El art. 83 RGPD parte del establecimiento de las condiciones generales para la imposición de multas de carácter administrativo. El legislador no ha dudado en indicar que estas deben corresponderse, básicamente, con una serie de características en su imposición. Por su parte, el art. 84 RGPD prevé que los Estados miembros puedan imponer «otras sanciones aplicables a las infracciones del presente Reglamento, en particular las infracciones que no se sancionen con multas administrativas de conformidad con el art. 83, y adoptarán todas las medidas necesarias para garantizar su observancia. Dichas sanciones serán efectivas, proporcionadas y disuasorias».

La responsabilidad de dicha actuación sancionadora, y que la misma se ajuste a estas características señaladas, es evidente que corresponde a cada autoridad de control. La RGPD establece una serie de circunstancias concurrentes en el ejercicio de esta función sancionadora, que deben ser tenidas en cuenta por cada regulador, a la hora de establecer la sanción correspondiente a cada caso concreto y determinado, así como con relación al hecho material de fijar de manera pormenorizada la cuantía de multa que se ha imponer.

3.3. La Agencia Española de Supervisión de la Inteligencia Artificial

La Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), creada por la DA 130, de 28 de diciembre, de Presupuestos Generales del Estado para el año 2022 y desarrollada por el Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial ("RD. 729/2023"), es una Agencia Estatal dotada de personalidad jurídica pública, patrimonio propio y autonomía en su gestión, con potestad administrativa, y se crea con el objetivo de ser el organismo público de referencia en materia de esta tecnología. Entre sus funciones está el desarrollo de las funciones que le asigna el Reglamento europeo de IA, supervisando los sistemas de IA de alto riesgo, coordinando la supervisión con las autoridades de vigilancia del mercado, promoviendo estándares y buenas prácticas y evaluando los modelos de IA (art. 4 y 10).

En este punto el art. 4.1 de RD 729/2023 establece que: "*la Agencia tendrá la función de inspección, comprobación, sanción y demás que le atribuya la normativa europea que le resulte de aplicación y, en especial, en materia de inteligencia artificial. Todo ello sin menoscabo de las competencias y funciones que en este ámbito vienen ejerciendo (…) el Ministerio de Trabajo y Economía Social y la Inspección de Trabajo y Seguridad Social, en su función de vigilancia del cumplimiento de las normas del orden social y exigencia de responsabilidades, en el ámbito de las relaciones laborales*".

Precisando el apartado 2 de ese mismo artículo que: "*La Agencia dentro del ámbito de competencias correspondientes al Estado y, de acuerdo con lo dispuesto en los artículos 108 bis a 108 sexies de la Ley 40/2015, de 1 de octubre, tiene por objeto la minimización de los riesgos que puede suponer el uso de esta nueva tecnología, el adecuado desarrollo y potenciación de los sistemas de inteligencia artificial. En el ámbito de la competencia estatal, ejercerá las funciones de autoridad responsable de la supervisión, y en su caso sanción, de los sistemas de inteligencia artificial con el objeto de eliminar o reducir los riesgos para la integridad, la intimidad, la igualdad de trato y la no discriminación, en particular entre mujeres y hombres, y demás derechos fundamentales que pueden verse afectados por el mal uso de los sistemas*".

Las referidas competencias se encuentran asociadas al potente régimen de responsabilidades y sanciones asociadas a los sistemas de IA que queda patente en el RIA. 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.

El art. 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 art. 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".

4. Responsabilidad por daños provocados por sistemas de IA

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 marca; 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.

Pero el daño ha de diferenciarse de la infracción administrativa en cuanto que la misma puede servir para castigar la creación de riesgos que llegarán a generar daños o

a quedarse, simplemente, en esferas de peligro que no se hayan llegado a concretar en daños efectivos. De este modo, las personas afectadas por una infracción del RIA deben tener también derecho a recibir una compensación por los daños efectivamente sufridos.

Las sanciones administrativas están diseñadas para castigar a las entidades que incumplen el reglamento y para disuadir futuras infracciones. Se basan en criterios que buscan asegurar el cumplimiento de las normas y prevenir comportamientos indebidos. En contraste, la indemnización tiene un objetivo reparador, orientado a compensar a las personas afectadas por el daño sufrido como resultado de la infracción.

Si bien con referencia al RGPD, pero con una proyección que alcanza también al esquema de responsabilidades que resulta aplicable a los sistemas de IA, la Sentencia del Tribunal de Justicia de la Unión Europea de 20 de junio de 2024, C-590/22, ha distinguido las distintas funciones y fines que poseen la indemnización por daños y perjuicios y las sanciones administrativas. El Tribunal de Justicia de la Unión Europea aclara que los criterios utilizados para imponer multas administrativas no deben aplicarse a la determinación del importe de la indemnización. La indemnización debe reflejar el daño real sufrido por la persona, mientras que las multas administrativas cumplen una función diferente, que es principalmente disuasoria y correctiva en relación con la conducta de las entidades responsables. Ciertamente, "al mantener esta distinción, el Tribunal de Justicia de la Unión Europea protege el derecho a una compensación adecuada para las víctimas sin que se vean afectadas por las medidas punitivas destinadas a los infractores. Ello también previene la posibilidad de que las entidades sean penalizadas en exceso por las mismas infracciones, ya que las sanciones administrativas y las indemnizaciones tienen propósitos distintos y deben aplicarse de manera separada" [4].

4.1. Antecedentes: Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial

Los principios y objetivos de la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial, que incorpora una Propuesta de Reglamento en esta materia (en adelante, "PR-RCIA")[5] buscó tratar de alcanzar la seguridad

[4] D. FIERRO RODRÍGUEZ, *Análisis de los daños derivados de infracciones administrativas del Reglamento General de Protección de Datos a la luz del asunto C-590/22*, Práctica de Derecho de Daños, 2024, nº 160.

[5] 2020/2014(INL). Sobre los aspectos principales de dicha Propuesta, puede verse P. ÁLVAREZ OLALLA, *Propuesta de Reglamento en materia de responsabilidad civil por el uso de inteligencia artificial, del Parlamento europeo, de 20 de octubre de 2020*, Revista CESCO de Derecho de consumo, 2021, nº 38, versión digital. También me ocupé de esta Propuesta de Reglamento en J. R. MERCADER UGUINA, *Riesgos, garantías y responsabilidades frente al uso de sistemas de inteligencia artificial*", en A. Abadías Selma y G. García González (Coord.), *Protección de los trabajadores e inteligencia artificial. La tutela de los derechos sociales en la cuarta revolución industrial*, Barcelona, Atelier, 2022, pp. 135-157.

jurídica a lo largo de la cadena de responsabilidad, en particular para el productor, el operador, la persona afectada y cualquier otro tercero.

El PR-RCIA definía un régimen de responsabilidad civil en materia de IA basado en el grado de control sobre un riesgo asociado al funcionamiento y la operación de un sistema de IA. Por ello, el cálculo de la responsabilidad entre los distintos agentes dependería de la interacción entre la finalidad de uso para la que se comercializa el sistema de IA, la forma en que se usa el sistema de IA, la gravedad del daño o perjuicio potencial, el grado de autonomía de la toma de decisiones que puede resultar en daños y de la probabilidad de que el riesgo se materialice. Además, la propuesta tiene sinergia y concordancia con las definiciones y principios de la Propuesta de Reglamento sobre Ley de Inteligencia Artificial, muestra de ello, son las definiciones que se plasman tanto en lo que debe entenderse por sistema de IA y su distinción como de "alto riesgo".

El régimen de responsabilidad se vertebraba desde la figura del operador de un sistema de IA. El «operador» es el «humano» que tiene el control sobre el sistema experto y que puede corresponder al usuario, al poseedor o al propietario. Es aquel que tiene el control del riesgo conectado con la operación de que se trate y que se beneficia de ella («risk management»)[6]. El Considerando (10) PR-RCIA afirmaba que "la responsabilidad civil del operador se justifica por el hecho de que controla un riesgo asociado al sistema de IA, comparable al del propietario de un automóvil" y, sobre esta base, entiende que, "debido a la complejidad y conectividad de un sistema de IA, el operador será, en muchos casos, el primer punto de contacto visible para la persona afectada". Sobre esta base el art. 3 de la Propuesta de Reglamento sobre un régimen de responsabilidad civil en materia de IA, distingue en operador inicial y final.

La PR-RCIA sentaba como premisa fundamental que "el operador de un sistema de IA de alto riesgo será objetivamente responsable de cualquier daño o perjuicio causado por una actividad física o virtual, un dispositivo o un proceso gobernado por dicho sistema de IA". Como se ha señalado, la responsabilidad objetiva de los sistemas de IA de alto riesgo es una responsabilidad objetiva estricta o pura (por riesgo)[7], "los operadores de un sistema de IA de alto riesgo no podrán eludir su responsabilidad civil alegando que actuaron con la diligencia debida o que el daño o perjuicio fue causado por una actividad, un dispositivo o un proceso autónomos gobernados por su sistema de IA. Los operadores no serán responsables si el daño o perjuicio ha sido provocado por un caso de fuerza mayor". El art. 4.4 PR-RCIA establecía que "el operador final de un sistema de IA de alto riesgo garantizará que las operaciones de dicho sistema de IA estén cubiertas por un seguro de responsabilidad civil adecuado en relación con los importes y el alcance de la indemnización previstos en los artículos 5 y 6 del presente Reglamento".

Especial relieve poseía el régimen de indemnizaciones. El art. 5 PR-RCIA establecía las cantidades máximas por las que respondería un operador de un sistema de IA de

[6] Como señala S. NAVAS NAVARRO, *Sistemas expertos basados en inteligencia artificial y responsabilidad civil. Algunas cuestiones controvertidas*, Diario La Ley, 2019.

[7] E. GOÑI HUARTE, *La causalidad incierta en la responsabilidad civil derivada de la Inteligencia Artificial*, Revista General de Derecho Europeo, 2022, nº 58, p. 349.

alto riesgo que hubiera sido considerado responsable de un daño o perjuicio en caso de fallecimiento o de daños causados a la salud o a la integridad física de una persona afectada como resultado del funcionamiento de un sistema de IA de alto riesgo; de los "daños morales significativos" que resultasen en una pérdida económica comprobable o en daños a bienes, también cuando distintos bienes propiedad de una persona afectada resulten dañados como resultado de un único funcionamiento de un único sistema de IA de alto riesgo; o, en fin, cuando la persona afectada dispusiera de un derecho a reclamar por responsabilidad contractual contra el operador en función de la cuantía de los perjuicios materiales o el daño moral. Y añadía que "cuando la indemnización combinada que deba abonarse a varias personas que sufran daños o perjuicios causados por el mismo funcionamiento de un mismo sistema de IA de alto riesgo supere los importes totales máximos previstos, los importes que deban abonarse a cada persona se reducirán proporcionalmente de forma que la indemnización combinada no supere los importes máximos establecidos". Dentro de los límites para el importe establecidos, se establecía un régimen particular para la indemnización en caso de daños físicos seguidos de la muerte de la persona afectada.

En materia de prueba, la PR-RCIA señalaba que tanto el operador como el perjudicado, podrán utilizar y disponer para demostrar la negligencia de los datos generados por el sistema de IA siempre bajo las salvaguardas del RGPD para hacer valer sus derechos en un proceso judicial ("Un operador considerado responsable podrá utilizar los datos generados por el sistema de IA para demostrar la negligencia concurrente de la persona afectada, de conformidad con el RGPD y otras leyes en materia de protección de datos relevantes. La persona afectada también podrá usar esos datos con fines probatorios o aclaratorios en la demanda por responsabilidad civil"). En el caso de que exista más de un operador todos ellos serán responsables solidarios.

En cuanto a los plazos de prescripción de las acciones, en el caso de que los daños afecten a la vida, salud o integridad física, el plazo de prescripción para el ejercicio de las acciones se establecía 30 años desde que se produjo el daño. En el caso de los daños materiales y morales se fijan unos plazos de prescripción especiales, siendo aplicable el que venza antes, y que consisten en: a) Diez años a partir de la fecha en que se produjo el menoscabo a los bienes o la pérdida económica comprobable resultante del daño moral significativo, respectivamente, o b) Treinta años a partir de la fecha en que tuvo lugar la operación del sistema de IA de alto riesgo que causó posteriormente el menoscabo a los bienes o el daño moral. Todo ello, sin perjuicio, de la interrupción de la prescripción conforme a las legislaciones de los Estados miembros.

4.2. La construcción bicéfala de la responsabilidad por daños de la IA

4.2.1. ¿Dos proyectos normativos para un solo fin?

Frustrada la anterior iniciativa, el 28 de septiembre de 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 relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial (Directiva sobre responsabilidad en materia de IA), 28 de septiembre de 2022[8] (en adelante, "PD-RCIA"), cuya tramitación se halla paralizada y la Propuesta de Directiva del Parlamento europeo y del Consejo sobre responsabilidad por los daños causados por productos defectuosos, 28 de septiembre de 2022[9] (en adelante, "PD-RPD"). Las instituciones europeas abandonan con estas dos últimas propuestas normativas la pretensión de regular de forma unitaria la responsabilidad civil derivada de daños causados por la IA a través de un reglamento.

Por lo que hace al PD-RPD, la proyección laboral es relativa. El mismo parte de que cuando un sistema de IA, por ser inseguro, produce daños que estén cubiertos por esta normativa, la víctima podrá ser indemnizada con arreglo a la misma. La legislación vigente de productos defectuosos, resultado de la trasposición de la Directiva 85/374/CEE del Consejo, de 25 de julio de 1985, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados Miembros en materia de responsabilidad por los daños causados por productos defectuosos, está llamada a regular los daños causados por aquellos bienes que adolecen de falta de seguridad, incluidos los sistemas de inteligencia artificial[10]. De acuerdo con la Directiva 85/374, La víctima podrá ser cualquier "perjudicado", lo que incluye tanto personas físicas como jurídicas, aunque la PD-RPD limita su aplicación a los "daños sufridos por personas físicas". Desde un punto de vista pasivo, el sujeto responsable de indemnizar estos daños será, con carácter general, el fabricante del producto final (o la persona que se presenta como tal al comercializar el producto) así como el fabricante de una materia prima o de cualquier elemento integrado en dicho producto

La PD-RCIA tiene por objeto establecer normas uniformes para facilitar el acceso a la información necesaria para probar los presupuestos de la responsabilidad extracontractual por culpa en los casos de daños causados por sistemas de inteligencia artificial y para facilitar la prueba, especialmente de la culpa y de la relación de causalidad, mediante el uso de presunciones y de otros elementos de facilitación probatoria (art. 1.1 a) y b) PD-RCIA).

Como resume Martín Casals, las particularidades que ofrece la PD-RCIA son las siguientes[11]: (i) solo es aplicable a los sistemas de IA y, por regla general, solo a los de alto

[8] COM/2022/496 final.

[9] COM/2022/495 final). Aprobación en primera lectura mediante «Resolución legislativa del Parlamento Europeo, de 12 de marzo de 2024, sobre la propuesta de Directiva del Parlamento 12 de octubre de 2023 - [COM (2022)0495 – C9-0322/2022 – 2022/0302(COD)].

[10] I. HERBOSA MARTÍNEZ, *Encaje de los sistemas de IA en la definición de producto en la legislación de productos defectuosos. Análisis de la legislación vigente con la vista puesta en la Propuesta de Directiva del Parlamento europeo y del Consejo de 28 de septiembre de 2022* (COM/2022/495), InDret: Revista para el Análisis del Derecho, 2024, nº 3, p. 69.

[11] 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. 69-71.

riesgo. (ii) Se aplicará en los casos que, de acuerdo con el Derecho del Estado miembro correspondiente, rija la responsabilidad por culpa y ante cualquier causante del daño. (iii) No altera las reglas nacionales de la distribución de la carga de la prueba ni del estándar probatorio. (iv) No se limita a un determinado tipo de daños, como los daños a las personas o a las cosas, sino que cubre todos los daños causados por ilícitos civiles que puedan dar lugar a responsabilidad de acuerdo con la legislación nacional aplicable. Así, por ejemplo, tales normas facilitarán el resarcimiento de daños causados por la intromisión en los derechos de la personalidad, como en caso de intromisión en la intimidad, o de derechos fundamentales, como, por ejemplo, la discriminación que pueda producirse en un proceso de contratación que use sistemas de IA para realizar la selección. (v) es una Directiva de armonización mínima, por lo que los Estados miembros pueden adoptar o mantener normas nacionales que sean más favorables para los demandantes, siempre que dichas normas sean compatibles con el Derecho de la Unión (cf. Art. 1.4 PD-RIA). Así, por ejemplo, las legislaciones nacionales podrían mantener la inversión de la carga de la prueba en el contexto de regímenes nacionales de responsabilidad por culpa o incluso establecer regímenes nacionales de responsabilidad objetiva.

4.2.2. Propuesta de Directiva del Parlamento europeo y del Consejo relativa a la adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial

Como señala la Exposición de Motivos de la PD-RCIA, "las características específicas de la IA, incluidas su complejidad, su autonomía y su opacidad (el denominado efecto de «caja negra»), pueden dificultar o hacer excesivamente costoso para las víctimas determinar cuál es la persona responsable y probar que se cumplen los requisitos para una demanda de responsabilidad civil admisible". Precisando su Considerando (3) que "*cuando la IA se interpone entre el acto u omisión de una persona y el daño, las características específicas de determinados sistemas de IA, como la opacidad, el comportamiento autónomo y la complejidad, pueden hacer excesivamente difícil, si no imposible, que el perjudicado satisfaga la carga de la prueba. En particular, puede resultar excesivamente difícil demostrar que un dato de entrada concreto del que es responsable la persona potencialmente responsable ha dado lugar a una información de salida específica de un sistema de IA que, a su vez, ha provocado el daño en cuestión*".

El art. 2.6 c) de la PD-RCIA establece que se puede interponer una demanda por daños y perjuicios por una persona que actúe en nombre de una o varias partes perjudicadas, de conformidad con el Derecho de la Unión o nacional. Esta disposición tendría por objeto "*brindar más posibilidades a las personas perjudicadas por un sistema de IA de que un tribunal conozca de su demanda, incluido en aquellos casos en interponer una demanda individual pueda parecer demasiado costoso o engorroso, o cuando una demanda conjunta*", como indica la Exposición de Motivos de la PD-RCIA

A) La exhibición y aseguramiento de pruebas en la PD-RCIA

Señala la Exposición de Motivos de la PD-RCIA que: "*la presente Directiva pretende proporcionar a las personas que soliciten una indemnización por los daños causados por sistemas de IA de alto riesgo medios eficaces para determinar las personas potencialmente responsables y las pruebas pertinentes de cara a una demanda. Al mismo tiempo, estos medios sirven para excluir a posibles demandados determinados erróneamente, ahorrando tiempo y costes a las partes implicadas y reduciendo la carga de trabajo de los tribunales*".

A este respecto, el art. 3.1 de la PD-RCIA establece que un órgano jurisdiccional puede ordenar la exhibición de "*pruebas pertinentes que obran en su poder [del demandado actual o potencial] sobre un determinado sistema de IA de alto riesgo del que se sospeche que ha causado daños*". Las solicitudes de pruebas se dirigen al proveedor de un sistema de IA, a una persona sujeta a las obligaciones del proveedor establecidas en el art. 24 o el art. 28.1 RIA, o a un usuario con arreglo a la RIA (un empleador a nuestros efectos). De acuerdo con el art. 3.2 "*en apoyo de su solicitud, el demandante potencial deberá presentar hechos y pruebas suficientes para sustentar la viabilidad de una demanda de indemnización por daños y perjuicios*".

De conformidad con el art. 3.4 de la PD-RCIA el órgano jurisdiccional únicamente puede ordenar dicha exhibición en la medida necesaria para sustentar la demanda, dado que la información podría constituir una prueba fundamental para la demanda de la persona perjudicada en caso de daños en los que hayan mediado sistemas de IA. De este modo, se ha dicho que "con buen criterio, el legislador europeo trata de poner los medios para conjurar uno de los mayores peligros inherentes a los mecanismos de obtención de prueba, a saber, los quebrantos a la confidencialidad y la necesidad de reserva de ciertas informaciones y, particularmente, la necesidad de preservar el secreto empresarial, evitando que litigantes maliciosos abusen de las diligencias para realizar denostables fishing expeditions"[12].

La negativa a secundar la orden de exhibición de pruebas dictada por el órgano judicial desencadenamiento de una presunción contra quien se muestra reacio a secundar la orden de exhibición. El art. 3.5 de la PD-RCIA dispone que "*cuando un demandado incumpla la orden de un órgano jurisdiccional nacional en una demanda por daños y perjuicios de exhibir o conservar las pruebas que obran en su poder con arreglo a los apartados 1 o 2, el órgano jurisdiccional nacional presumirá el incumplimiento por parte del demandado de un deber de diligencia pertinente, en particular en las circunstancias a que se refiere el artículo 4, apartados 2 o 3, que las pruebas solicitadas estaban destinadas a probar a efectos de la correspondiente demanda por daños y perjuicios. Al demandado le asistirá el derecho de refutar esa presunción*".

[12] G. ORMAZABAL SÁNCHEZ, *La prueba en los procesos de responsabilidad civil por daños causados por sistemas de inteligencia artificial. Análisis del Derecho vigente y de las propuestas normativas de la UE,* InDret, 2024, nº 3, p. 432.

B) Presunción de relación de causalidad en caso de culpa

En lo que respecta a los daños causados por sistemas de IA, la PD-RCIA pretende proporcionar un fundamento eficaz para reclamar una indemnización en relación con la culpa consistente en el incumplimiento de un deber de diligencia en virtud del Derecho de la Unión o nacional.

Puede resultar difícil para los demandantes probar que existe un nexo causal entre dicho incumplimiento y la información de salida producida por el sistema de IA o la no producción de una información de salida por parte del sistema de IA que haya dado lugar a los daños en cuestión. El art. 4.1 de la PD-RCIA "introduce una compleja presunción" que debe reunir tres condiciones[13]:

- Que el demandante haya demostrado o el órgano jurisdiccional haya supuesto, de conformidad con el art. 3.5 de la PD-RCIA, la culpa del demandado o de una persona de cuyo comportamiento sea responsable el demandado, consistente en el incumplimiento de un deber de diligencia establecido por el Derecho de la Unión o nacional destinado directamente a proteger frente a los daños que se hayan producido.
- Que pueda considerarse razonablemente probable, basándose en las circunstancias del caso, que la culpa ha influido en los resultados producidos por el sistema de IA o en la no producción de resultados por parte del sistema de IA.
- Que el demandante haya demostrado que la información de salida producida por el sistema de IA o la no producción de una información de salida por parte del sistema de IA causó los daños.

Por último, el art. 4.7 de la PD-RCIA, establece que el demandado tiene derecho a refutar la presunción de causalidad la anterior presunción.

En el caso de los sistemas de IA de alto riesgo, tal como se definen en la Ley de IA, el art. 4.4 del PD-RCIA, establece una excepción a la presunción de causalidad cuando el demandado demuestre que el demandante puede acceder razonablemente a pruebas y conocimientos especializados suficientes para demostrar el nexo causal. Esta posibilidad puede incentivar a los demandados a cumplir sus obligaciones de exhibición, las medidas establecidas por la Ley de IA para garantizar un alto nivel de transparencia de la IA o los requisitos de documentación y registro.

En el caso de los sistemas de IA de riesgo no elevado, el art. 4.5 de la PD-RCIA, establece una condición para la aplicabilidad de la presunción de causalidad en virtud de la cual esta última está sujeta a que el órgano jurisdiccional determine que es excesivamente difícil para el demandante demostrar el nexo causal. Tales dificultades deben evaluarse a la luz de las características de determinados sistemas de IA, como la autonomía

[13] G. ORMAZABAL SÁNCHEZ, *La prueba en los procesos de responsabilidad civil por daños causados por sistemas de inteligencia artificial*, cit, p. 413.

y la opacidad, que hacen muy difícil en la práctica la explicación del funcionamiento interno del sistema de IA, lo que afecta negativamente a la capacidad del demandante para demostrar el nexo causal entre la culpa del demandado y la información de salida de IA.

La función última de todas estas disposiciones, como recalca la Exposición de Motivos de la PD-RCIA es "ofrecer a todos los que participan en actividades relacionadas con sistemas de IA un incentivo adicional para cumplir sus obligaciones en relación con la conducta que se espera de ellos".